



UNIVERSITÀ DEGLI STUDI DELL'AQUILA

**Department of Information Engineering, Computer  
Science, and Mathematics**

Doctoral Program in Information and Communication Technology Emerging  
computing models, software architectures, and intelligent systems

# **Systems and Techniques for Post-Stroke Rehabilitation and Assessment**

Relatore

**Prof. Giuseppe Placidi**

Correlatore

**Prof. Filippo Mignosi**

Candidato

**Alessandro Di Matteo**

**288997**

**Anno Accademico 2024-2025**





# Contents

|           |                                                                                                                           |           |
|-----------|---------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>I</b>  | <b>Introduction</b>                                                                                                       | <b>4</b>  |
| <b>1</b>  | <b>Stroke</b>                                                                                                             | <b>5</b>  |
| 1.1       | Pathophysiology . . . . .                                                                                                 | 5         |
| 1.2       | Risk factors . . . . .                                                                                                    | 5         |
| 1.3       | Diagnosis of Stroke . . . . .                                                                                             | 6         |
| 1.4       | Stroke recovery stages . . . . .                                                                                          | 7         |
| 1.5       | AI for medical imaging segmentation for stroke . . . . .                                                                  | 7         |
| 1.5.1     | Model Training and Optimization . . . . .                                                                                 | 9         |
| 1.5.2     | From CNNs to Advanced Models . . . . .                                                                                    | 10        |
| 1.5.3     | Evaluation Metrics . . . . .                                                                                              | 12        |
| <b>2</b>  | <b>Post-Stroke Rehabilitation</b>                                                                                         | <b>16</b> |
| 2.1       | Qualitative and Quantitative Rehabilitation Approaches . . . . .                                                          | 16        |
| 2.2       | Technological Advances in Stroke Rehabilitation . . . . .                                                                 | 17        |
| 2.2.1     | Robotic Systems . . . . .                                                                                                 | 19        |
| 2.2.2     | Virtual, Augmented Reality and Digital Twins systems . . . . .                                                            | 20        |
| 2.2.3     | Gesture Recognition . . . . .                                                                                             | 23        |
| <b>II</b> | <b>Hand Telerehabilitation</b>                                                                                            | <b>28</b> |
| <b>3</b>  | <b>Patient–Therapist Cooperative Hand Telerehabilitation through a Novel Framework Involving the Virtual Glove System</b> | <b>29</b> |
| 3.1       | The Cooperative Telerehabilitation Framework . . . . .                                                                    | 30        |
| 3.1.1     | Hardware Architecture and Data Model . . . . .                                                                            | 31        |
| 3.1.2     | Use Cases and Software Architecture . . . . .                                                                             | 35        |
| 3.2       | Execution example . . . . .                                                                                               | 39        |
| 3.3       | Discussion and Conclusions . . . . .                                                                                      | 41        |
| <b>4</b>  | <b>Double Digital Twin for the Hand Rehabilitation through the Virtual Glove</b>                                          | <b>43</b> |
| 4.1       | Background . . . . .                                                                                                      | 44        |
| 4.2       | The Double Digital Twin System . . . . .                                                                                  | 44        |
| 4.3       | Methodology . . . . .                                                                                                     | 45        |
| 4.3.1     | Therapist Task Creation . . . . .                                                                                         | 46        |
| 4.3.2     | Task Execution with Healthy Hand . . . . .                                                                                | 46        |
| 4.3.3     | Task Execution with Injured Hand . . . . .                                                                                | 47        |
| 4.3.4     | Quantitative and Qualitative Evaluation . . . . .                                                                         | 48        |
| 4.4       | Results and Discussion . . . . .                                                                                          | 49        |
| 4.5       | Conclusion . . . . .                                                                                                      | 51        |

|            |                                                                                                       |           |
|------------|-------------------------------------------------------------------------------------------------------|-----------|
| <b>5</b>   | <b>Portable Head-Mounted System for Mobile Forearm Tracking</b>                                       | <b>53</b> |
| 5.1        | Related Works . . . . .                                                                               | 54        |
| 5.2        | System Design . . . . .                                                                               | 55        |
| 5.3        | Measurements . . . . .                                                                                | 58        |
| 5.3.1      | Power Consumption . . . . .                                                                           | 58        |
| 5.3.2      | CPU Usage and Temperature . . . . .                                                                   | 60        |
| 5.3.3      | Hand Tracking . . . . .                                                                               | 62        |
| 5.4        | Conclusion . . . . .                                                                                  | 65        |
| <b>6</b>   | <b>BNNAction-Net: Binary Neural Network on Hands Gesture Recognitions</b>                             | <b>66</b> |
| 6.1        | Method . . . . .                                                                                      | 66        |
| 6.1.1      | Knowledge Distillation . . . . .                                                                      | 67        |
| 6.1.2      | Implementation Details . . . . .                                                                      | 67        |
| 6.1.3      | Evaluation . . . . .                                                                                  | 67        |
| 6.2        | Conclusion . . . . .                                                                                  | 68        |
| <b>III</b> | <b>Physical Rehabilitation Data</b>                                                                   | <b>69</b> |
| <b>7</b>   | <b>A DICOM-based Standard for Quantitative Physical Rehabilitation</b>                                | <b>70</b> |
| 7.1        | Background . . . . .                                                                                  | 71        |
| 7.2        | The DICOM Architecture . . . . .                                                                      | 72        |
| 7.2.1      | Core Information Model and Methods . . . . .                                                          | 72        |
| 7.3        | DICOM IOD design for Physical Rehabilitation . . . . .                                                | 74        |
| 7.3.1      | Equipment . . . . .                                                                                   | 74        |
| 7.3.2      | Task-Based Numerical Tracking and Data Generation . . . . .                                           | 75        |
| 7.3.3      | PR-IOD Structure . . . . .                                                                            | 76        |
| 7.4        | PR-IOD for the QPR of the Hand: an example . . . . .                                                  | 79        |
| 7.4.1      | Proof of Concept . . . . .                                                                            | 80        |
| 7.5        | Conclusion . . . . .                                                                                  | 81        |
| <b>IV</b>  | <b>Assessment through MRI</b>                                                                         | <b>84</b> |
| <b>8</b>   | <b>Deep Learning and Multi-Modal MRI for the Segmentation of Sub-Acute and Chronic Stroke Lesions</b> | <b>85</b> |
| 8.1        | Introduction . . . . .                                                                                | 86        |
| 8.2        | Methods and Materials . . . . .                                                                       | 87        |
| 8.3        | Results . . . . .                                                                                     | 90        |
| 8.4        | Discussion . . . . .                                                                                  | 92        |
| 8.5        | Conclusion . . . . .                                                                                  | 93        |
| 8.6        | Data availability statement . . . . .                                                                 | 94        |
| <b>V</b>   | <b>Conclusions</b>                                                                                    | <b>97</b> |
| <b>9</b>   | <b>Conclusion and Future Directions</b>                                                               | <b>98</b> |
| 9.1        | Summary of Contributions . . . . .                                                                    | 98        |
| 9.2        | Synthesis and Significance . . . . .                                                                  | 99        |
| 9.3        | Limitations . . . . .                                                                                 | 100       |
| 9.4        | Future Research Directions . . . . .                                                                  | 100       |

|             |                                                                                                                                       |            |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>VI</b>   | <b>Lists</b>                                                                                                                          | <b>102</b> |
|             | List of Figures                                                                                                                       | 103        |
|             | List of Tables                                                                                                                        | 105        |
|             | Acronyms                                                                                                                              | 108        |
|             | Candidate's Publications                                                                                                              | 112        |
|             | Bibliography                                                                                                                          | 114        |
| <b>VII</b>  | <b>Appendix</b>                                                                                                                       | <b>130</b> |
|             | A Supplementary Material Chapter 8. Deep Learning and Multi-Modal MRI for<br>the Segmentation of Sub-Acute and Chronic Stroke Lesions | 131        |
| <b>VIII</b> | <b>Acknowledgement</b>                                                                                                                | <b>150</b> |

# Abstract

Stroke is one of the leading causes of death and long-term disability worldwide, representing a major global health and socio-economic challenge. Although advances in acute treatment have improved survival, many stroke survivors experience persistent motor deficits that compromise independence in daily life. Effective recovery, therefore, depends on extensive, personalized rehabilitation, particularly for upper-limb function. In this context, the convergence of artificial intelligence, medical imaging, and rehabilitation engineering is fostering a new generation of systems and techniques for quantitative, data-driven, and patient-centered care.

The first research direction introduces advanced technological systems for hand telerehabilitation. Initially, the Virtual Glove (VG) system, along with its framework, is presented, which integrates real-time computer vision with therapist-patient interaction to support remote and quantitative rehabilitation. In conjunction with the VG, the Double Digital Twin (DDT) concept is introduced, enabling the simultaneous modeling of both healthy and impaired limbs to allow quantitative comparison, adaptive therapy planning, and personalized feedback. Building on this foundation, a portable head-mounted hand-tracking system is proposed, combining the Leap Motion Controller 2 device with a single-board computer to achieve autonomous, lightweight motion capture in real-world scenarios. To further enhance real-time functional analysis, a Binary Neural Network architecture (BNNAction-Net) is developed for efficient hand-gesture recognition on resource-constrained platforms. Together, these systems form a comprehensive technological layer for quantitative rehabilitation, providing both continuous motion monitoring and objective functional assessment.

To ensure interoperability and clinical integration, in this thesis, a new DICOM-based standard is proposed: the Physical Rehabilitation Information Object Definition (PR-IOD). The PR-IOD is specifically designed to represent quantitative rehabilitation data. By extending DICOM beyond imaging, PR-IOD establishes a standard structure for storing, sharing, and visualizing multimodal rehabilitation metrics across devices and institutions, linking functional performance with standardized clinical assessment and paving the way for large-scale data integration. Complementing the functional dimension, the final research line focuses on neuroimaging-based assessment. Deep-learning models trained on the nnU-Net framework automatically segment subacute and chronic stroke lesions from multimodal MRI. This structural analysis provides objective biomarkers of brain damage that can inform and personalize rehabilitation strategies, connecting functional recovery with underlying neural reorganization.

Together, these contributions form a coherent technological ecosystem, from portable sensing and AI-driven motion analysis to standardized data management and MRI-based lesion assessment, thus advancing the vision of integrated, personalized, and quantitatively driven post-stroke rehabilitation and assessment, where objective measurement, clinical interpretation, and adaptive therapy are unified within a single data-centric continuum.

Keywords: Post-Stroke, Rehabilitation, Hand-Tracking, Virtual Glove, Hand-gesture Recognition, MRI, Lesion Segmentation, deep learning

---

## Contributions of this thesis

This doctoral thesis explores the intersection of rehabilitation engineering, medical imaging, and artificial intelligence to support and optimize post-stroke recovery, with a specific focus on upper limb rehabilitation. The research presented herein encompasses the development and evaluation of innovative systems for hand tracking, gesture recognition, rehabilitation data management, and lesion segmentation based on multi-modal magnetic resonance imaging (MRI). The Research Questions (RQs) that guide this thesis are:

- RQ1: How can portable and low-cost hand-tracking systems be made robust and usable for daily-life stroke rehabilitation?
- RQ2: Can Digital Twin approaches provide reliable and personalized measures of upper limb function during rehabilitation?
- RQ3: Can lightweight neural networks, such as BNNs, enable gesture recognition on resource-limited devices?
- RQ4: How can rehabilitation data be standardized (e.g., with DICOM) to improve interoperability, monitoring, and data sharing?
- RQ5: How effective are deep learning models (e.g., nnU-Net) in segmenting stroke lesions from multi-modal MRI to support personalized therapy?

A first research direction investigates telerehabilitation, where assistive technologies such as robotics, virtual reality, and computer vision tracking systems enable remote therapy, objective evaluation, and continuous monitoring of patient progress. Within this framework, the Virtual Glove (VG) is introduced, a computer-vision-based system for real-time (40 fps) hand tracking that leverages two Leap Motion Controller sensors in an orthogonal way. The VG provides therapists with precise, quantitative information on joint-level hand mobility and enables both patient–therapist interaction and therapist–therapist data sharing. Building upon this foundation, the thesis presents the Double Digital Twin (DDT) of the hand, a model that simultaneously represents both the healthy and impaired limbs to support personalized rehabilitation. Calibration procedures align the healthy hand model with the impaired one, enabling synchronized, repeatable task execution and comparison. This framework not only quantifies mobility deficits but also provides an advanced tool for comparative analysis and adaptive therapy planning.

Afterwards, a portable head-mounted hand-tracking device is proposed, combining the Leap Motion Controller 2 (LMC) with a single-board computer (SBC) powered by a compact power bank. This lightweight and autonomous solution enhances portability and enables users to perform rehabilitation tasks in everyday environments. The system’s hardware and software architectures are described in detail, and its robustness is evaluated under variable lighting conditions, power consumption, CPU temperature, and frame rate, demonstrating its suitability for mobile and outdoor rehabilitation contexts.

Another line of research focuses on efficient gesture recognition for immersive rehabilitation scenarios. In this context, BNNAction-Net is introduced, a Binary Neural Network (BNN)-based architecture designed for real-time hand gesture recognition in resource-constrained environments

---

such as untethered head-mounted displays (HMDs). By binarizing both weights and activations, the model drastically reduces computational complexity and memory usage while maintaining competitive accuracy. Evaluated on the EgoGesture dataset, BNNAction-Net demonstrates its effectiveness for wearable rehabilitation systems relying on low-power embedded hardware.

While these systems enable detailed data acquisition and computation, their potential remains limited by the absence of a unified standard for rehabilitation data. Current limitations include poor interoperability across devices, difficulties in tracking patient histories, limited temporal evaluation, and barriers to data sharing and statistical analysis. To address these issues, this thesis proposes a novel DICOM Information Object Definition (PR-IOD) tailored for Quantitative Physical Rehabilitation (QPR). By extending the well-established DICOM standard beyond medical imaging, PR-IOD ensures structured data representation, interoperability, and multi-device compatibility. Its implementation is demonstrated by integrating Virtual Glove data into PR-IOD objects, subsequently visualized through a dedicated viewer dashboard. This contribution paves the way toward standardized rehabilitation data management and large-scale clinical integration.

Finally, this thesis addresses the problem of lesion segmentation from multi-modal MRI to better characterize stroke-related brain damage and inform rehabilitation strategies. Leveraging the nnU-Net framework, deep learning models are designed and evaluated for automated segmentation of sub-acute and chronic stroke lesions using both T1-weighted and FLAIR sequences. This last component bridges medical imaging with functional rehabilitation, enabling a comprehensive pipeline where neural models support diagnosis, patient stratification, and therapy personalization. This work proposes an integrated framework that spans from portable hand-tracking systems to digital twins approaches, from standardized rehabilitation data management to efficient gesture recognition, and finally to lesion segmentation with deep learning. Together, these contributions advance the vision of personalized, data-driven, and technologically enhanced stroke rehabilitation.

## Overview of the Thesis

The thesis is organized into five main parts. Part I provides an overview of the clinical and technological background, outlining the pathophysiology of stroke, the stages of recovery, current rehabilitation strategies and assessment, with particular attention to immersive environments and AI techniques. Part II focuses on telerehabilitation systems, presenting studies that employ VR devices for hand tracking, such as the Leap Motion Controller (versions 1 and 2), and the integration of AI techniques for hand gesture recognition. Part III presents a novel DICOM-based standard for quantitative physical rehabilitation, designed to enhance data interoperability and facilitate clinical integration. Part IV explores the development and validation of advanced deep learning models for the segmentation of sub-acute and chronic stroke lesions from MRI data. Finally, Part V provides an overall discussion of the findings, highlighting their implications and limitations, outlining directions for future research.

## Part I

# Introduction

# Chapter 1

## Stroke

Stroke, also known as a Cerebrovascular accident, is the third leading cause of death globally, following cardiovascular diseases and all forms of cancer. It remains the primary cause of severe, long-term disability, thus posing a substantial socio-economic burden worldwide. Over recent decades, the impact of stroke has grown not only in terms of incidence but also in media coverage, public awareness, healthcare service planning, and scientific research [1].

According to the World Health Organization, 15 million people worldwide suffer a stroke annually. Of these, 5 million die, and another 5 million are left permanently disabled [2]. The overall incidence rate is estimated at 2-2.5 cases per 1,000 individuals per year, with a high (15–40%) risk of recurrence within 5 years of the initial event. Furthermore, estimates suggest that nearly one in four men and almost one in five women aged 45 is expected to experience a stroke by the age of 85 [3]. While stroke is uncommon in people under 40 years (where the leading cause is high blood pressure), it also occurs in about 8% of children with sickle cell disease.

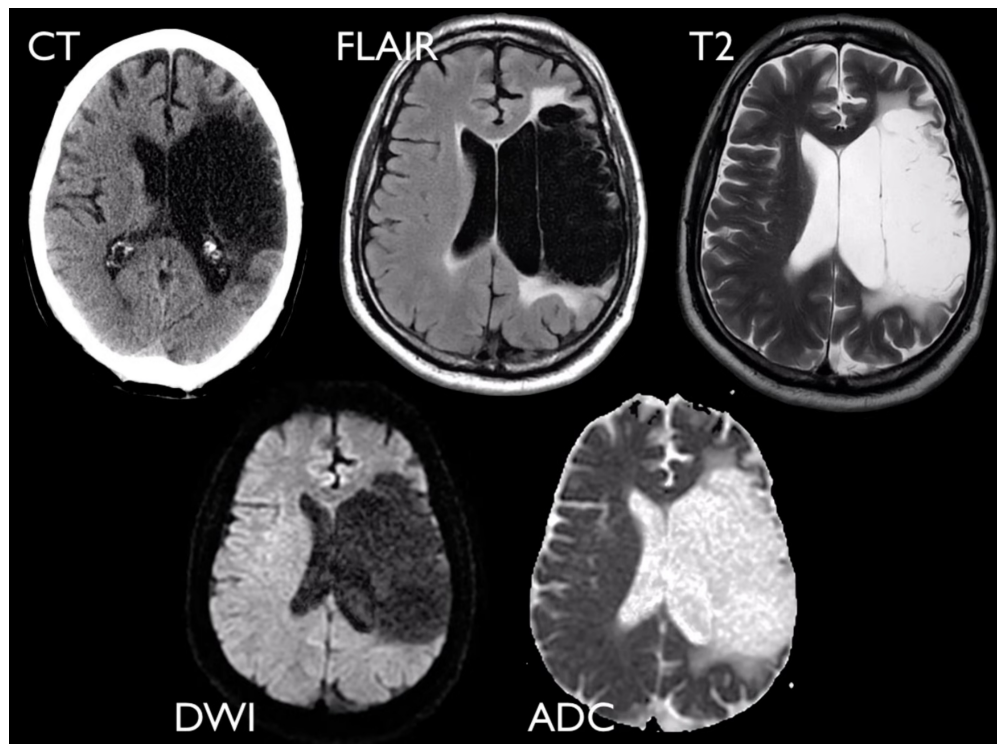
### 1.1 Pathophysiology

A stroke occurs when the blood supply to a part of the brain is interrupted or reduced, depriving brain tissue of essential oxygen and nutrients. This sudden disruption leads to the rapid death of brain cells in the affected area. There are two primary types of stroke: ischemic stroke, caused by an obstruction typically a blood clot that blocks blood flow to the brain, and hemorrhagic stroke, resulting from the rupture of a blood vessel, which leads to bleeding within or around the brain and impairs normal circulation. Both types constitute a medical emergency requiring immediate intervention to minimize brain damage and improve outcomes.

### 1.2 Risk factors

Although numerous factors contribute to the risk of stroke, such as age, family history, diabetes mellitus, chronic kidney disease, and sleep apnea, the most significant modifiable risk factors include hypertension, atrial fibrillation, smoking, symptomatic carotid artery disease, and sickle cell disease. Lifestyle-related factors, such as physical inactivity, frequent consumption of sugar-sweetened beverages, and low intake of fish, fruits, and vegetables, have also been associated with a higher risk of stroke. In women, additional risk is conferred by the use of oral contraceptives, migraine





**Figure 1.1:** Illustration of a chronic stroke lesion across different imaging modalities: CT, MRI-FLAIR, T2-weighted, DWI, and ADC; highlighting the complementary diagnostic information each provides; image taken from <https://radiopaedia.org/> ("Stroke: Evolution from acute to chronic infarction - radiology video tutorial (CT, MRI)).

with aura, the immediate postpartum period, and preeclampsia, although these factors generally contribute to a modest absolute increase in stroke risk [4].

### 1.3 Diagnosis of Stroke

When a patient is admitted after a stroke, one of the priorities is a thorough and timely clinical assessment. This evaluation provides an initial picture of the patient's neurological condition by examining motor function, sensation, coordination, and cognitive abilities. Such an approach allows clinicians to promptly identify deficits caused by stroke and plan appropriate interventions. Particular attention must also be given to speech, language, and swallowing functions, as communication difficulties and dysphagia are frequent consequences that require the early involvement of speech and language therapists.

Neuroimaging plays a central role in this initial phase [5]. All patients with suspected stroke should undergo urgent brain imaging, usually starting with a non-contrast computed tomography (CT) or, when available, magnetic resonance imaging (MRI) (Fig. 1.1).

The goal is not only to confirm the presence of a stroke but also to differentiate ischemic from hemorrhagic forms and to exclude other conditions that may mimic stroke, such as tumors or brain abscesses. CT remains the most accessible and rapid option, with good sensitivity for detecting acute hemorrhage and mass lesions. However, its capacity to reveal ischemic changes is limited in the very early hours. In contrast, multimodal MRI, and especially diffusion-weighted sequences, provides much higher sensitivity for acute ischemia and comparable accuracy to CT in identifying

intracerebral hemorrhage. Despite these advantages, CT remains the first-line tool in many settings, as it is faster, less expensive, and suitable even for patients with contraindications to MRI.

The information obtained from these imaging techniques is fundamental, not only for diagnosis and treatment but also for understanding the location and extent of brain damage, key factors in shaping an effective rehabilitation plan.

## 1.4 Stroke recovery stages

Stroke recovery is typically categorized into three stages: acute, subacute, and chronic.

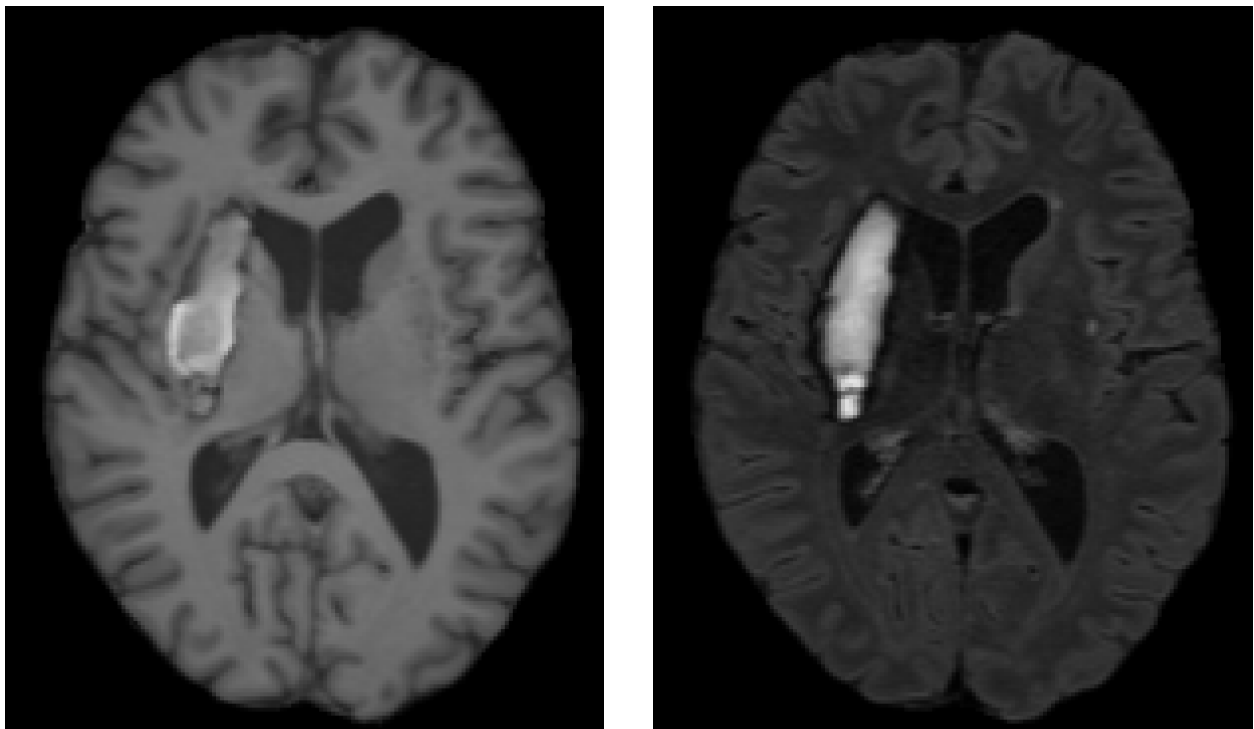
- **The acute stage:** encompasses the first 24h to 1 week;
- **The subacute stage:** occurs from 1 week to 3 to 6 months post stroke;
- **The chronic stage of stroke:** begins 6 months after the initial event and extends indefinitely.

Whilst in the acute phase, rapid identification of brain lesions is vital for determining optimal treatment strategies and mitigating potential long-term effects, often using techniques such as diffusion-weighted and perfusion MRI. In later stages, including the sub-acute and chronic phases, conventional structural MRI is used to examine the consequences and underlying causes of stroke. T1-weighted (T1-w) images provide high-resolution anatomical detail, essential for evaluating structural damage. Although less frequently used, FLAIR sequences are more sensitive to stroke-related abnormalities, such as hyperintensities and edema, and can offer complementary information to T1-w scans. Accurate segmentation of lesions in the sub-acute and chronic phases is essential for understanding brain recovery, predicting neurological outcomes, and personalizing rehabilitation strategies. By identifying the location and extent of affected brain regions, segmentation supports targeted therapy planning and helps monitor disease progression over time. However, segmenting lesions in these later stages is particularly challenging: compared to acute lesions, chronic lesions tend to be more diffuse, less well-defined, and highly variable across individuals [6, 7].

## 1.5 AI for medical imaging segmentation for stroke

While manual segmentation remains the gold standard for lesion identification, it is inherently labor-intensive, time-consuming, and prone to inter-rater variability. These limitations highlight the pressing need for automated approaches. In this context, deep learning has driven remarkable progress, particularly for post-stroke lesion segmentation [8–11]. However, robust methods for subacute and chronic lesion segmentation are still under development.

The growing need for automation in medical imaging and specifically for post-stroke lesion segmentation has been paralleled by the evolution of artificial intelligence (AI). This field seeks to emulate human cognitive functions such as perception, learning, and decision-making. Deep learning, as a subset of AI, is grounded in the concept of the artificial neuron, which models how information is processed and transmitted in the brain. The origins of modern neural networks can be traced back to the **Perceptron**, introduced by Rosenblatt in 1958 [12], one of the earliest models of artificial neurons. The perceptron is a simple computational unit that receives multiple inputs, applies a set of weights to them, and produces an output through a step function. Despite



**Figure 1.2:** Post-Stroke patient images across multiple imaging modalities: T1-W and Flair [6].

its simplicity, the perceptron laid the groundwork for more advanced architectures, though it was initially limited in its ability to solve problems that were not linearly separable.

Building upon this foundation, the concept of the **artificial neuron** was generalized. A neuron can be seen as a basic computational unit that processes weighted inputs and generates an output through a nonlinear activation function. Formally, the output  $y$  of a neuron is expressed as:

$$y = f \left( \sum_{i=1}^n w_i x_i + b \right)$$

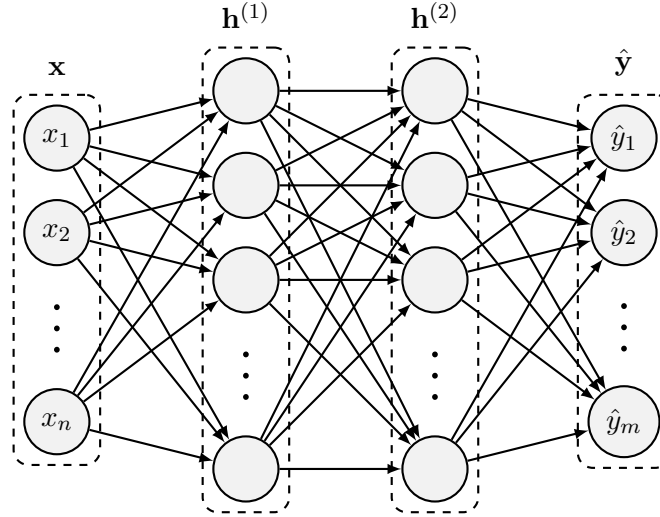
where  $x_i$  are the inputs,  $w_i$  their associated weights,  $b$  the bias term, and  $f$  a nonlinear activation function (e.g., sigmoid, ReLU, tanh).

By combining many neurons into **layers**, and stacking multiple layers sequentially, one obtains a **multi-layer neural network**.

Each layer learns increasingly abstract representations of the input data, enabling the network to extract hierarchical features and capture complex relationships automatically. Unlike traditional algorithms that rely heavily on handcrafted features, deep learning models can represent and approximate highly nonlinear mappings directly from raw data.

The theoretical foundation of this capability is provided by the **Universal Approximation Theorem** [13–15]. It states that a feedforward neural network with at least one hidden layer can approximate any continuous function on a compact subset of  $\mathbb{R}^n$  with arbitrary accuracy, under the following conditions:

1. The hidden layer must contain a sufficient number of neurons.
2. The activation functions must be nonlinear.



**Figure 1.3:** Schematic representation of a multilayer perceptron (MLP). The network is composed of an input layer  $\mathbf{x}$ , two hidden layers  $\mathbf{h}^{(1)}$  and  $\mathbf{h}^{(2)}$ , and an output layer  $\hat{\mathbf{y}}$ . Each neuron computes a weighted sum of its inputs, adds a bias term, and applies a nonlinear activation function. The dense connectivity between layers enables the model to learn hierarchical feature representations and approximate complex nonlinear mappings.

3. The target function must be continuous and defined on a bounded domain.

In other words, any continuous function defined over a finite interval can be approximated by a suitable linear combination of nonlinear transformations. This result formally established that multilayer feedforward networks are universal function approximators.

### 1.5.1 Model Training and Optimization

The optimization of the weights ( $w_i$ ) and the biases ( $b$ ) is generally performed through the **back-propagation algorithm** [16], which minimizes a cost function  $J(w, b)$  using the gradient-based methods. At the core of this process lies the **loss function** (also called **cost function**), which quantifies the discrepancy between the model's prediction  $\hat{y}$  and the ground truth  $y$ . In practice, the loss provides a scalar measure of “how wrong” the network is: a higher loss indicates worse performance, while a lower loss corresponds to more accurate predictions. Minimizing the loss thus *guides* learning by penalizing errors. The choice of the cost function, also called **loss function**, depends on the type of task (Regression, Classification, Segmentation). In this work we focus on segmentation losses appropriate for highly imbalanced medical images.

Some common segmentation losses are presented below:

- **Cross-Entropy (CE):** measures the dissimilarity between the predicted probability distribution and the ground truth; it stabilizes training and works well with calibrated probabilistic outputs;
- **Dice loss:** region-based surrogate of the Dice coefficient, particularly effective under severe foreground–background imbalance (e.g., small lesions);
- **Hybrid CE+Dice:** combines CE’s optimization stability with Dice’s imbalance robustness and is widely adopted in medical image segmentation.

The training process of a neural network can be divided into two main phases:

1. **Forward propagation:** the input data are successively transformed as they pass through the layers of the network. For a single neuron, the output is computed as:

$$z = \sum_{i=1}^n w_i x_i + b, \quad a = \sigma(z)$$

where  $x_i$  are the inputs,  $w_i$  the corresponding weights,  $b$  the bias,  $z$  the net input, and  $\sigma(\cdot)$  an activation function. At the end of the forward pass, the network produces the prediction  $\hat{y}$ , which is compared with the ground truth  $y$  through a loss function  $J(y, \hat{y})$ . This scalar  $J$  quantifies the prediction error and is the quantity to be minimized.

2. **Backward propagation:** the error computed in the forward step is propagated backwards through the network. By applying the chain rule, the gradients of the loss function with respect to each parameter are calculated:

$$\frac{\partial J}{\partial w_i} = \frac{\partial J}{\partial z} \cdot \frac{\partial z}{\partial w_i}$$

These gradients quantify how much each parameter influences the error. They are then used to update the parameters in the direction that most rapidly reduces  $J$ .

Finally, the weights are updated according to the gradient descent rule:

$$w_i \leftarrow w_i - \eta \cdot \frac{\partial J}{\partial w_i}$$

with  $\eta$  denoting the learning rate, which controls the step size of the update.

The basic version of this procedure is known as **Stochastic Gradient Descent (SGD)**. This algorithms aim to reach an optimal (or near-optimal) solution of the minimization problem, reducing the risk of getting trapped in poor local minima or experiencing very slow convergence. SGD works by iteratively updating the model parameters in the direction of the negative gradient of the loss function. The gradient of the loss function is a vector that points in the direction of the steepest ascent of the loss function. By moving in the direction of the negative gradient, SGD is able to find the minimum of the loss function. In medical image segmentation, **adaptive** optimizers such as Adam or Root Mean Square Propagation (RMSProp) are frequently employed, as they adjust effective learning rates per-parameter and often yield faster, more stable convergence.

### 1.5.2 From CNNs to Advanced Models

Within deep learning, **Convolutional Neural Networks (CNNs)** have emerged as a key architecture for data with a grid-like structure, such as images or EEG signals. In CNNs, each convolutional layer applies a kernel  $W$  over the input  $X$ , producing an output map  $Y$ :

$$Y[i, j] = \sum_{m=0}^{K_h-1} \sum_{n=0}^{K_w-1} X[i+m, j+n] \cdot W[m, n] + b$$

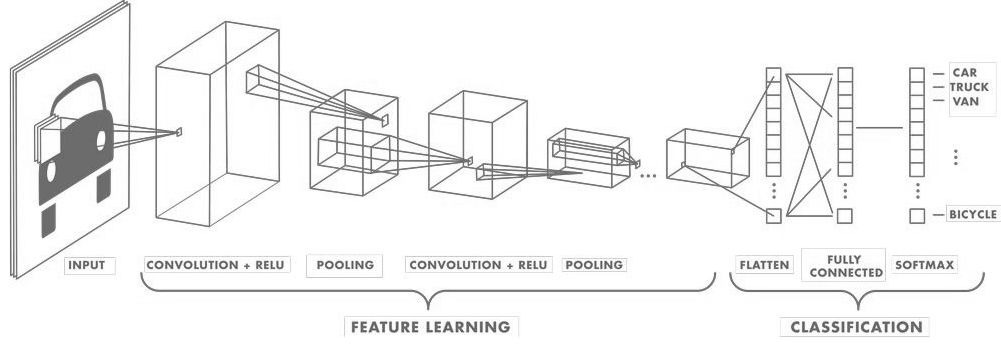


Figure 1.4: Convolutional Neural Network schema [18]

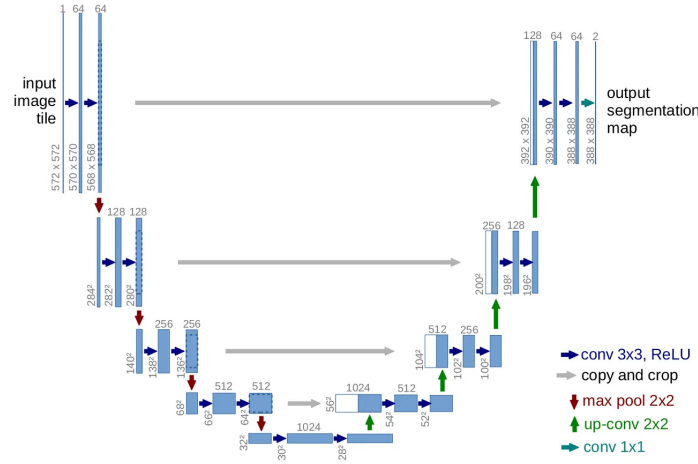
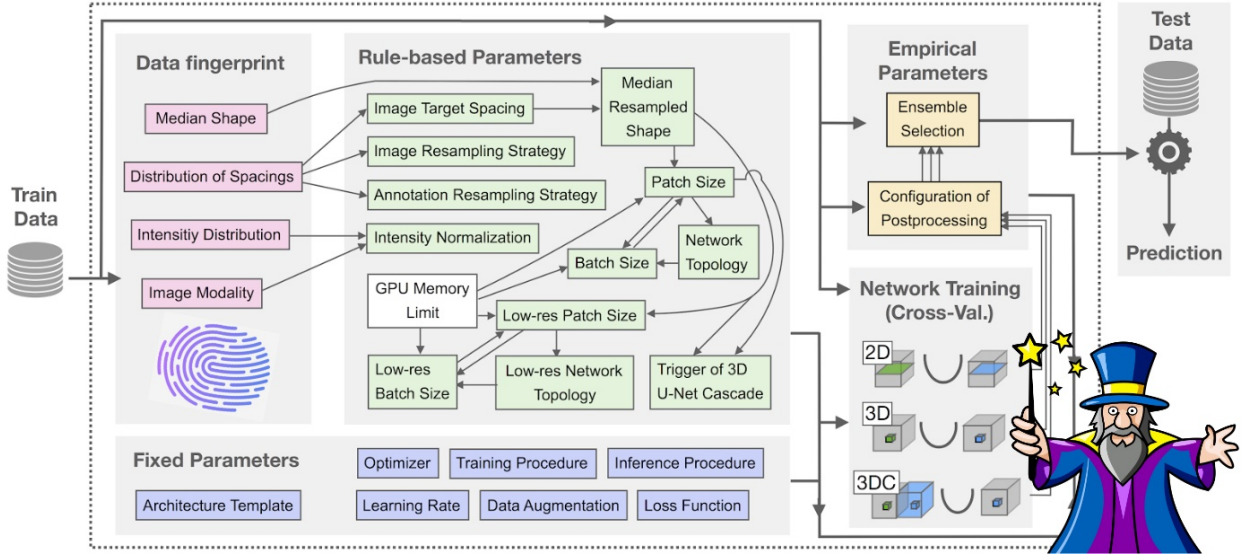


Figure 1.5: Schematic representation of the 3D U-Net architecture. image taken from the original paper [19]

where  $K_h$  and  $K_w$  are the kernel dimensions. By hierarchically stacking such layers, CNNs learn increasingly abstract features, from edges and textures to high-level structures. The organization of the Human Visual Cortex inspired their architecture and has proven highly effective in computer vision and speech recognition [17].

A widely used architecture in this context is the **U-Net** [19], which is specifically designed for biomedical segmentation. Its **encoder-decoder** design follows three main principles:

- The **encoder path** compresses the spatial resolution while extracting semantic context. This is achieved through repeated applications of 3D convolutions, followed by nonlinear activations and **down-convolutions** (strided convolutions) or **max-pooling layers**, which progressively reduce the spatial dimensions while increasing the number of feature channels.
- The **decoder path** restores the original resolution of the input volume. This is performed through **up-convolutions** (transposed convolutions) or interpolation-based upsampling, which expand the spatial resolution while gradually reducing the number of channels.
- **Skip connections** directly link encoder and decoder layers at the same resolution, allowing the network to combine high-level semantic information with low-level spatial details. This mechanism is introduced to cut the vanishing gradient problem.



**Figure 1.6:** Schematic overview of the **nnU-Net** pipeline [20]. The framework automatically configures preprocessing, network architecture, training, and post-processing steps based on dataset characteristics. This adaptive design allows nnU-Net to achieve state-of-the-art performance across diverse biomedical segmentation tasks without manual tuning.

This formulation enables the network to jointly learn spatial dependencies across all three dimensions. By combining down-convolutions, pooling, and up-convolutions with skip connections, the U-Net effectively captures both global context and fine structural details.

Building upon the success of U-Net, several extensions and frameworks have been proposed to improve generalization and simplify model design for diverse biomedical applications. Among these, one of the most influential is the **nnU-Net** [20], which serves as the foundation for the models developed in this thesis.

The experiments conducted in this thesis were based on the **nnU-Net framework** [20], a self-configuring deep learning system that has become a reference standard for biomedical image segmentation. Unlike manually designed pipelines, nnU-Net automatically adapts its architecture, preprocessing, training schedule, and post-processing parameters to the characteristics of a given dataset. It builds upon the U-Net architecture, extending it with a robust set of heuristics that infer task-specific configurations such as patch size, batch size, network depth, and normalization directly from dataset fingerprints.

Through this automated configuration, nnU-Net minimizes the need for manual tuning while ensuring competitive performance across a wide variety of medical imaging modalities and segmentation targets. Its modular design also facilitates custom extensions, making it particularly suitable for research applications.

### 1.5.3 Evaluation Metrics

The performance of classification models is usually assessed using **accuracy**, **precision**, **recall**, and **F1-score**, which are derived from the four basic outcomes of a confusion matrix:

- **True Positives (TP):** correctly predicted positive cases (e.g., lesion correctly detected).
- **False Positives (FP):** non-lesion regions wrongly classified as lesions.

- **True Negatives (TN):** correctly predicted negative cases (e.g., healthy tissue identified as such).
- **False Negatives (FN):** missed positive cases (e.g., undetected lesion).

In segmentation tasks, however, evaluation must consider both voxel-level accuracy and region-level correctness. Unlike simple classification, the output is a structured 2D/3D mask. Commonly used metrics include:

- **Dice Similarity Coefficient (DSC):**

$$DSC = \frac{2|P \cap G|}{|P| + |G|}$$

Measures the overlap between the predicted mask ( $P$ ) and the ground truth ( $G$ ). Particularly suited for medical imaging, where class imbalance is common. DSC is mathematically equivalent to the **F1-score**:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}},$$

with

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}.$$

- **Intersection over Union (IoU):**

$$IoU = \frac{|P \cap G|}{|P \cup G|}$$

Also known as the Jaccard Index, it penalizes over-segmentation more strongly than Dice.

- **Hausdorff Distance (HD):**

$$HD(P, G) = \max \left\{ \sup_{p \in P} \inf_{g \in G} d(p, g), \sup_{g \in G} \inf_{p \in P} d(g, p) \right\}$$

Captures the maximum boundary deviation, highlighting worst-case segmentation errors.

- **Average Surface Distance (ASD):** Computes the *mean distance* from all points on one surface to the closest point on the other. It is a directional metric; for instance, the ASD from  $P$  to  $G$  is:

$$ASD(P, G) = \frac{1}{|S_P|} \sum_{p \in S_P} \inf_{g \in S_G} d(p, g)$$

providing a measure of the average deviation of the predicted surface from the ground truth.

- **Average Symmetric Surface Distance (ASSD):**

$$ASSD(P, G) = \frac{1}{|S_P| + |S_G|} \left( \sum_{p \in S_P} \inf_{g \in S_G} d(p, g) + \sum_{g \in S_G} \inf_{p \in S_P} d(g, p) \right)$$



Represents the bidirectional mean distance between predicted and reference surfaces. By symmetrizing the ASD, it provides a more stable and balanced estimate of boundary accuracy, less affected by local outliers than the HD.

- **Voxel-wise metrics (Accuracy, Specificity):** Applied voxel by voxel, these capture under- and over-segmentation:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad \text{Specificity} = \frac{TN}{TN + FP}.$$

While useful, they may be biased by large background regions.

### Lesion-wise metrics

Lesion-wise evaluation is performed on the **connected components** of the binary masks, where each connected component is considered as a single lesion [21]. Let  $N_{GT}$  be the number of connected components in the ground-truth mask and  $N_{Pred}$  the number of connected components in the predicted mask.

We denote:

- $g_i$ : the  $i$ -th ground-truth lesion, i.e. the  $i$ -th connected component of the ground-truth mask;
- $p_j$ : the  $j$ -th predicted lesion, i.e. the  $j$ -th connected component of the predicted mask;
- $G = \bigcup_{i=1}^{N_{GT}} g_i$ : the union of all ground-truth lesions;
- $P = \bigcup_{j=1}^{N_{Pred}} p_j$ : the union of all predicted lesions.

We define the following lesion-wise counts:

- $TP_L$ : number of correctly detected lesions (true positives lesion-wise);
- $FP_L$ : number of predicted lesions not corresponding to any ground-truth lesion (false positives lesion-wise);
- $FN_L$ : number of ground-truth lesions not detected (false negatives lesion-wise).

**1.5.3.0.1 Detection rule.** A ground-truth lesion  $g_i$  is counted as  $TP_L$  if:

1. its lesion sensitivity is greater than or equal to a threshold  $\alpha$ :

$$s_i = \frac{|g_i \cap P|}{|g_i \cap P| + |g_i \setminus P|} \geq \alpha,$$

where  $|g_i \cap P|$  are the voxels of  $g_i$  correctly detected and  $|g_i \setminus P|$  are the missed voxels;

2. ordering the predicted lesions  $p_j$  overlapping  $g_i$  by decreasing intersection size, the contributors required to cover at least  $\gamma$  of the detected voxels satisfy

$$r_j = \frac{|p_j \setminus G|}{|p_j|} \leq \beta,$$

where  $|p_j \setminus G|$  are the voxels of  $p_j$  not overlapping any ground-truth lesion.

If  $g_i$  does not satisfy these conditions, it is counted as  $FN_L$ . Similarly, each predicted lesion  $p_j$  that does not match any ground-truth lesion according to the same rules is counted as  $FP_L$ .

#### 1.5.3.0.2 Role of the parameters $\alpha$ , $\beta$ , $\gamma$ .

- $\alpha$  (sensitivity threshold): defines the minimum fraction of voxels of a ground-truth lesion that must be covered by predicted voxels for the lesion to be considered detected.
- $\beta$  (overflow threshold): controls how much of a predicted lesion is allowed to extend outside the ground-truth regions. A predicted lesion  $p_j$  is accepted as a valid contributor only if the fraction of voxels outside  $G$  is at most  $\beta$ .
- $\gamma$  (coverage threshold): defines how much of the detected part of a ground-truth lesion must be explained by valid predicted lesions (those that satisfy  $\beta$ ). This ensures that the majority of the detected voxels come from sufficiently accurate predictions.

**1.5.3.0.3 Lesion-wise metrics.** Based on  $TP_L$ ,  $FP_L$  and  $FN_L$  the lesion-wise metrics are computed as:

$$\text{Precision (PPV)}_L = \frac{TP_L}{TP_L + FP_L} \quad (1.1)$$

$$\text{Recall (Sensitivity)}_L = \frac{TP_L}{TP_L + FN_L} \quad (1.2)$$

$$\text{F1-score}_L = \frac{2 \cdot \text{Precision}_L \cdot \text{Recall}_L}{\text{Precision}_L + \text{Recall}_L} \quad (1.3)$$

$$\text{False Discovery Rate (FDR)}_L = \frac{FP_L}{TP_L + FP_L} \quad (1.4)$$

$$\text{False Negative Rate (FNR)}_L = \frac{FN_L}{TP_L + FN_L} \quad (1.5)$$

## Chapter 2

# Post-Stroke Rehabilitation

Rehabilitation is a cornerstone of post-stroke care, aimed at restoring functional abilities and improving patients' quality of life. It encompasses a broad range of therapeutic approaches addressing motor, cognitive, and daily-life functions. Physical therapy focuses on mobility, balance, and coordination through structured exercises and task-specific training. Occupational therapy supports independence in daily living by adapting activities and environments, while speech and language therapy targets communication and swallowing difficulties. Cognitive rehabilitation, instead, addresses impairments in memory, attention, executive functions, and problem-solving that frequently follow brain injury [5, 22].

Early and intensive intervention is strongly recommended, since the first three months after a cerebrovascular event represent a critical window of heightened neuroplasticity, during which the brain is particularly capable of reorganizing and recovering lost functions [23]. Moreover, timely rehabilitation can prevent further deterioration, foster residual capacities, and promote reintegration into social and daily life. Nevertheless, functional improvements may still occur in the chronic stage, where recovery relies mainly on cortical reorganization and adaptive neuroplastic processes [5]. These mechanisms are the foundation of neurophysiological approaches such as mental practice and motor imagery.

Within this wide spectrum, the present thesis focuses on upper-limb physical rehabilitation, a specialized domain that targets the recovery of arm and hand functionality. Interventions include active-assisted exercises, robotic-assisted therapy, mirror therapy, and repetitive task practice, all designed to enhance range of motion, reduce spasticity, and improve motor control. Importantly, post-stroke rehabilitation also contributes to secondary prevention by addressing modifiable risk factors (e.g., hypertension, diabetes, hypercholesterolemia), thereby reducing the likelihood of recurrent events [5].

### 2.1 Qualitative and Quantitative Rehabilitation Approaches

Physical rehabilitation after stroke can be delivered through either **qualitative** or **quantitative** approaches, each with distinct strengths and limitations. Rehabilitation may occur in conventional face-to-face settings or remotely via *telerehabilitation* [24, 25] (Figure 2.1). In traditional practice, therapists largely rely on qualitative evaluations derived from clinical observation and subjective judgment. By contrast, telerehabilitation increasingly incorporates assistive technologies that enable

standardized and objective *quantitative assessments*.

The paradigm of *Quantitative Physical Rehabilitation (QPR)* leverages objective metrics, such as kinematic and kinetic data, to track patient progress, personalize therapy, and facilitate interoperable data sharing across clinical teams [26]. This data-driven framework aligns closely with the principles of *precision medicine*, promoting individualized interventions based on measurable outcomes. Evidence suggests that QPR, particularly when combined with robotic systems or virtual environments, can deliver superior functional recovery compared to conventional methods [27]. Moreover, it provides a valuable solution for patients with limited mobility or those living in underserved areas with restricted access to specialized care [28].

A major challenge in quantitative physical rehabilitation is the lack of a unified standard for data representation. Currently, each device manufacturer or research group relies on proprietary formats, leading to fragmented and heterogeneous datasets. This absence of standardization has several consequences: Interoperability issues: data generated by different motion tracking systems cannot be easily exchanged or integrated. Incomplete clinical records: rehabilitation data are rarely linked to electronic health records (EHRs), preventing the creation of longitudinal patient histories. Lack of comparability: results from different devices, institutions, or studies are difficult to align, limiting cross-patient evaluations and multicenter research. Difficulty in defining common metrics: without shared data structures, therapists and researchers cannot establish reproducible scales or quantitative benchmarks. Fragmented ecosystem: the lack of harmonization slows down the development of large-scale datasets and hinders the clinical adoption of innovative rehabilitation technologies. To overcome these barriers, the rehabilitation domain can draw inspiration from other areas of medicine where data standardization has already been successfully achieved. One of the most relevant examples is the Digital Imaging and Communications in Medicine (DICOM) standard. Originally developed for radiological imaging, DICOM has progressively expanded into a versatile and extensible framework capable of supporting diverse biomedical data, including time series such as ECG and EEG. Its widespread adoption, integration with hospital infrastructures (e.g., PACS), and ability to store both raw measurements and rich metadata make it a promising candidate for structuring rehabilitation data. Leveraging DICOM for QPR would enable interoperability, facilitate integration with clinical workflows, and foster the development of standardized evaluation metrics. In this thesis in order to overcome this issue we presented a DICOM-like standard to process and elaborate these data.

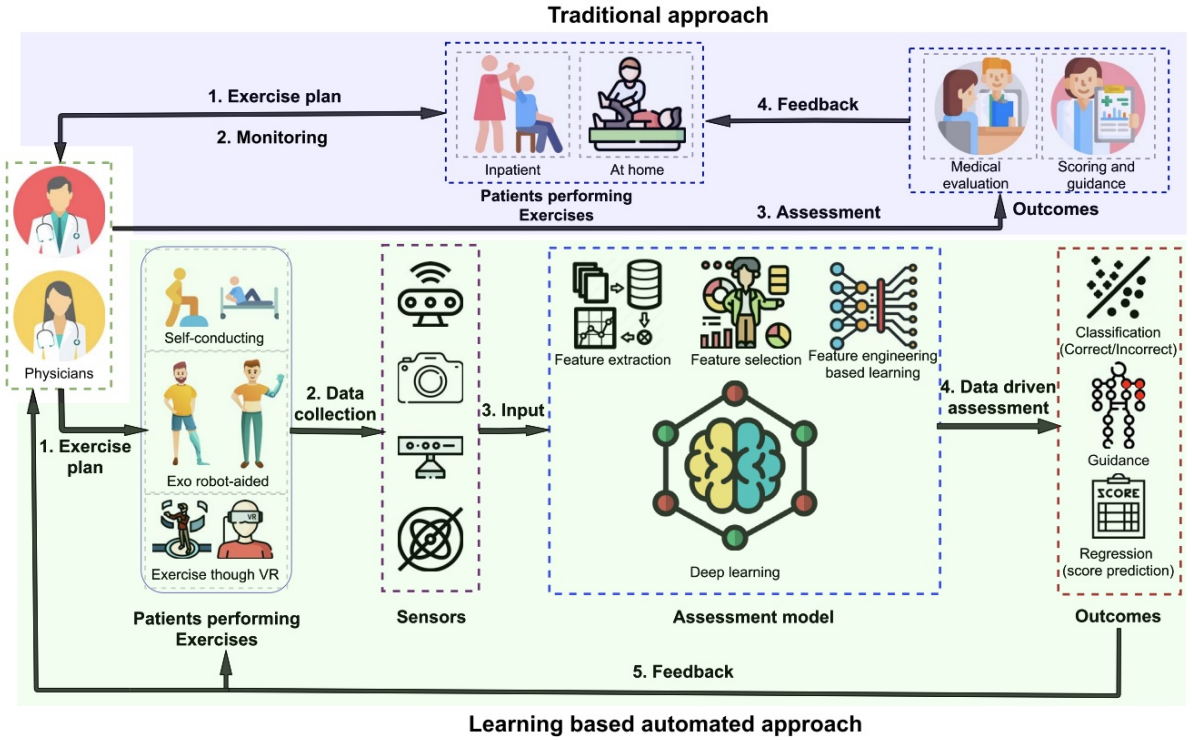
Nonetheless, **qualitative rehabilitation** continues to play an essential role in capturing patient-specific aspects that are not easily quantifiable, such as emotional state, fatigue, pain, and compensatory behaviors. Yet, its reliance on subjective observation introduces **variability and limited reproducibility**, which may compromise consistency across practitioners [29, 30]. For this reason, while qualitative insights remain clinically relevant, they are increasingly complemented, and in some cases even outperformed, by standardized, technology-driven quantitative approaches better suited to precision-based rehabilitation models.

## 2.2 Technological Advances in Stroke Rehabilitation

In recent decades, technological progress has profoundly reshaped stroke rehabilitation. One of the most prominent innovations is the adoption of **robotics**, which offers controlled, repetitive, and

adaptive training tailored to each patient's functional abilities. Robotic-assisted therapy has been shown to enhance motor recovery, functional outcomes, and patient engagement [31].

Alongside robotics, the expansion of immersive technologies has brought **Virtual Reality (VR)** and **Augmented Reality (AR)** systems into rehabilitation practice. These platforms allow patients to perform task-oriented and motivating exercises in interactive environments that encourage neuroplasticity and adherence. VR systems may rely on *contactless solutions* (e.g., depth cameras, motion-tracking sensors) or *wearable devices* (e.g., data gloves, inertial measurement units), each presenting distinct trade-offs in usability, accuracy, and patient comfort.



**Figure 2.1:** Traditional qualitative approaches versus technology-driven quantitative approaches in post-stroke rehabilitation [31]

- **Robot-assisted rehabilitation** provides mechanical support and guidance during movement execution, making it suitable across a wide spectrum of patients, including those in the *acute* and *sub-acute* phases with limited voluntary control. Its adaptability ensures feasibility even in severe impairment conditions.
- **VR-based rehabilitation**, on the other hand, often requires a baseline level of cognitive and motor function, as patients must interact with virtual tasks and process visual-spatial feedback. This makes VR particularly suitable in the *sub-acute* and *chronic* stages, where residual motor abilities and cognitive engagement are preserved.

A common feature of both paradigms is the closed-loop interaction between patient and system. The patient performs specific tasks, which are captured via sensors and processed by the system. The assessment module then analyzes the data to generate an evaluation of performance. This may include classification of movement correctness, score prediction, or real-time feedback to optimize motor recovery (Figure 2.1).

Overall, both paradigms present complementary strengths: while robotics ensures feasibility even in severely impaired patients, VR/AR systems maximize engagement and functional specificity when adequate cognitive-motor function is available. Therefore, tailoring the rehabilitation modality to the patient’s clinical stage (acute, sub-acute, chronic) and functional status is essential.

### 2.2.1 Robotic Systems

Robotics has become an important component of modern stroke rehabilitation, offering new approaches to support motor recovery and enhance patients’ functional outcomes [31]. In particular, robotic systems for upper limb rehabilitation, such as exoskeletons and end-effector devices, enable highly precise and intensive therapy. By guiding patients through repetitive, task-specific movements, these systems enhance motor control, strength, and range of motion in the affected arms and hands. Their adaptability allows clinicians to tailor exercises to individual impairments, thereby promoting targeted and effective interventions [31].

Robotics also plays a central role in gait rehabilitation. Exoskeletons and treadmill-based devices provide mechanical assistance and structured guidance, helping patients relearn proper gait patterns, correct asymmetries, and improve balance and stride length [32]. Crucially, real-time feedback facilitates motor learning and accelerates the retraining of walking abilities [33].

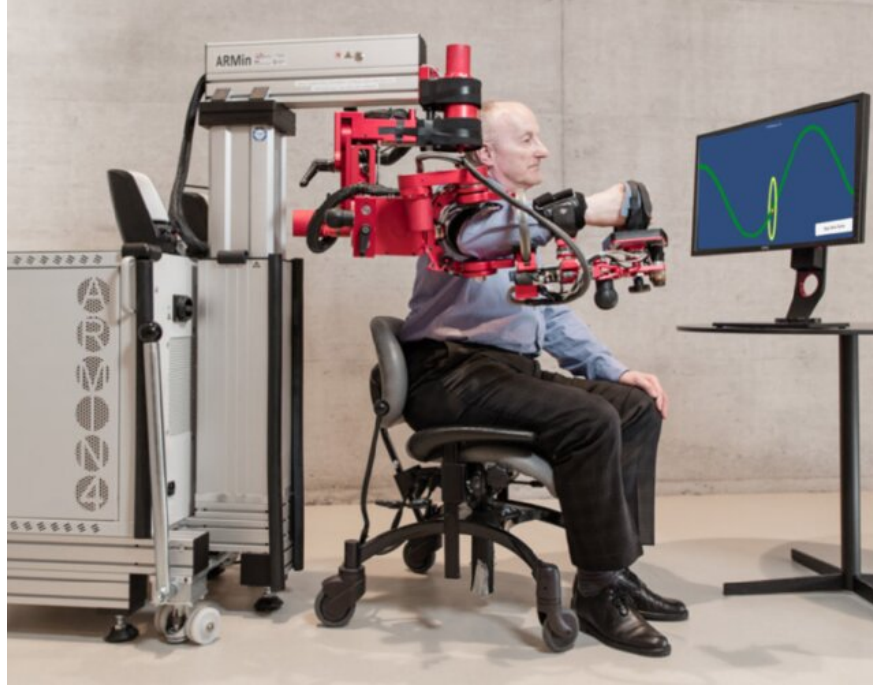
One of the major strengths of robotic-assisted rehabilitation lies in its capacity for *task-specific training*. By replicating real-life activities, these systems encourage neuroplasticity, supporting the reorganization of neural pathways and the recovery of functional movement [34]. Furthermore, robotic devices collect detailed performance data, including range of motion, applied force, and movement accuracy, allowing therapists to monitor progress, adjust therapy intensity, and optimize patient-specific treatment plans.

The development of portable and user-friendly robotic devices has further expanded rehabilitation beyond hospital walls. Home-based robotic training provides continuity of therapy, improves accessibility, and supports long-term adherence, key factors in sustaining recovery [35]. Additionally, the integration of gamified interfaces and AI-driven adaptation enhances patient motivation and engagement, transforming therapy into a more interactive and enjoyable experience [36].

Their clinical relevance is supported by landmark studies and large-scale trials, including:

- **MIT-MANUS:** Demonstrated safe and effective upper limb therapy over more than 500 hours of training [37, 38].
- **RATULS Trial:** Compared robotic therapy with conventional care, showing that high-intensity therapy is beneficial regardless of modality [39].
- **ARMin, ReoGo, Bi-Manu-Track, T-WREX, VA-ROBOTICS:** Reported varying improvements in motor function, frequently assessed with different scores [40–45].

Despite these advantages, several limitations remain. High acquisition and maintenance costs restrict large-scale implementation, while specific usage protocols, particularly assistive or passive modes, can reduce patient effort and engagement, potentially limiting therapeutic benefits. For this reason, therapist supervision and careful monitoring of participation remain essential to maximize outcomes.



**Figure 2.2:** ARMin-IV, a 7-DoF robotic device designed for upper-limb rehabilitation [40].

Overall, while challenges persist, robotic systems represent a promising frontier in stroke rehabilitation. By combining precision, adaptability, and engagement, they offer powerful tools to support neuroplasticity and functional recovery.

### 2.2.2 Virtual, Augmented Reality and Digital Twins systems

Virtual Reality (VR) and Augmented Reality (AR) technologies are increasingly applied in stroke rehabilitation, providing immersive and interactive tools that facilitate both motor and cognitive training [46, 47]. In VR, patients are immersed in computer-generated environments that replicate real-life scenarios and therapeutic tasks, ranging from upper limb exercises to balance-related activities. These simulated settings stimulate neuroplasticity and motor learning by encouraging functional practice in safe and controlled contexts [48]. AR, conversely, overlays digital elements onto the real world, delivering real-time visual cues and corrective feedback to guide movement execution and enhance spatial awareness. Beyond motor recovery, both VR and AR have proven valuable in cognitive rehabilitation, where gamified tasks challenge memory, attention, and executive functions, thereby promoting neural adaptation and cognitive recovery [49]. Importantly, these technologies allow patients to rehearse everyday tasks, such as cooking, shopping, or navigating public spaces, helping restore autonomy and self-confidence before re-engagement in real-life environments [49].

According to the degree of sensory involvement they provide, virtual reality (VR) systems can generally be classified into three categories [50]: non-immersive, semi-immersive, and fully immersive. Non-immersive VR environments are typically presented in two dimensions through conventional computer monitors or gaming consoles, allowing users to interact with the virtual scenario via external devices such as a mouse, joysticks, data gloves, or force sensors. In this configuration, the perspective is usually allocentric, with the user observing the scene through a third-person

viewpoint often represented by an avatar on the screen.

Semi-immersive VR solutions employ three-dimensional visualizations, frequently generated through stereoscopic projections or fixed-perspective displays. These systems enable a more engaging and interactive experience than non-immersive setups, largely because users' physical movements are tracked through dedicated sensors that influence their interaction with the virtual world.

In contrast, fully immersive VR offers an egocentric perspective, placing the individual directly within a computer-generated environment [51]. By employing dedicated hardware, such as head-mounted displays (HMDs), large projection screens (LPS), or cave automatic virtual environments (CAVE), users can experience real-time correspondence between visual feedback and their head or body movements, thereby simulating interactions that closely resemble those of the physical world. An HMD typically integrates two small displays positioned in front of the eyes, combined with head-tracking sensors to adapt the rendered imagery, and headphones to provide synchronized auditory input. The effectiveness of VR/AR-based rehabilitation depends, however, on appropriate patient selection. Cognitive impairments, balance disturbances, or visual deficits may reduce their applicability, making comprehensive clinical assessment essential before prescription [5]. Nevertheless, the inherently engaging and enjoyable nature of VR/AR has been shown to improve patient motivation, adherence, and consistency in rehabilitation programs, ultimately enhancing outcomes [48, 49].

A critical enabling technology for these immersive interventions is **motion tracking**. Accurate tracking of body parts provides quantitative insights into motor performance during rehabilitation. Different solutions exist:

**Marker-based optical systems** (e.g., Vicon) ensure high accuracy but are costly and limited to controlled environments. **Markerless vision-based systems**, such as Microsoft Kinect or the Leap Motion Controller (LMC), provide affordable, non-intrusive tracking. The LMC, for instance, uses stereo vision to reconstruct hand and finger motion with sub-millimetric precision across 25 joints, while Kinect combines RGB, depth, and IR data to track full-body movements. **Wearable devices** (e.g., data gloves) capture joint angles and motion trajectories through accelerometers and gyroscopes, enabling continuous monitoring both in clinical and home-based settings.

These systems generate rich time-series data describing spatial position, velocity, acceleration, and joint angles, which are essential for quantifying motor function, identifying abnormal movement patterns, and monitoring recovery trajectories. Their comparative characteristics are summarized in Table 2.1.

Wearable devices, such as data gloves or inertial motion sensors, offer high precision and are less affected by occlusion. Nonetheless, their use can limit the naturalness of movements, reduce comfort, and in some cases be unsuitable for patients with specific medical conditions [53]. In contrast, contactless technologies, such as the Leap Motion Controller (LMC) and Microsoft Kinect, allow unobtrusive tracking of motor functions and promote more natural interactions. Their main limitation lies in the sensitivity to occlusions, particularly during tasks involving complex gestures or object manipulation. To address these challenges, **multisensor contactless systems** have been developed. By combining multiple viewpoints and integrating the acquired data through fusion algorithms, these systems achieve greater spatial accuracy and robustness against occlusion. In the A2VI-LAB laboratory, the Virtual Glove was designed, obtained by combining two Leap Motion Controllers V1 (LMC) arranged orthogonally [54]. This configuration, adopted also in the present work, was conceived to reduce occlusion issues and to increase the accuracy of hand tracking, which





**Figure 2.3:** Wearable device manus xsense metaglove [52].

makes the device especially suitable for rehabilitation scenarios. The development of the VG began with a preliminary evaluation of the two LMC sensors considered separately. Successively, they were merged into a single unit, allowing simultaneous observation of the hand from two perpendicular viewpoints. An illustration of the device during data acquisition is reported in Figure 1.4. Each LMC operates within its own three-dimensional reference system, offering a field of view of approximately  $150^\circ$  along the X-axis and  $120^\circ$  along the Z-axis. Because the infrared LEDs are powered through the USB interface, the effective detection range is limited to about 0.6 m. Within this range, the hands are illuminated and their reflections are captured by two infrared cameras, producing grayscale stereo images. These images are then processed to reconstruct a 3D model of the hand, including the position of joints and the orientation of fingers, with the possibility of tracking up to two hands at once. To integrate the signals from both devices, a calibration phase was carried out in order to compute the roto-translational transformation linking the two local reference systems. The small residual mismatch observed between the vertical and horizontal frames of reference was not large enough to impair the functioning of the VG, which thus achieved more precise hand tracking than a single LMC could provide. Data fusion algorithms were subsequently implemented to merge the two input streams into a single enriched dataset [55]. Data contains kinematic information of the reconstructed hand, such as joint positions, velocities, and angular values.

Augmented and Virtual Reality (AR/VR), when integrated with Digital Twins (DTs), allow the gamification of rehabilitation tasks, thereby enhancing patient motivation and adherence to therapy. DTs provide a safe environment for simulating high-risk interventions, reducing potential harm during recovery [56]. Furthermore, they can contribute to cost-effective care by improving diagnostic accuracy and minimizing redundant procedures. Beyond patient care, DTs act as educational and planning resources for clinicians, enabling the visualization of complex cases and the co-design of rehabilitation strategies. When combined with AI, DTs gain enhanced decision-support capabilities, leveraging large datasets to generate tailored insights and adaptive therapeutic recommendations. Continuous advances in computational power and machine learning are expected to yield increas-

**Table 2.1:** Comparison between Contactless, Wearable, and Marker-based Optical Motion Tracking Systems

| Feature                      | Contactless Systems             | Wearable Systems                           | Marker-based Optical Systems                          |
|------------------------------|---------------------------------|--------------------------------------------|-------------------------------------------------------|
| <b>Examples</b>              | Leap Motion, Kinect             | IMUs, Data Gloves                          | Vicon, OptiTrack                                      |
| <b>Intrusiveness</b>         | Low (non-invasive)              | Medium–High (physical wear required)       | High (markers must be attached to body)               |
| <b>Mobility impact</b>       | No constraint                   | May limit natural movement                 | Minimal, but markers can detach during motion         |
| <b>Occlusion sensitivity</b> | High (self/object occlusion)    | Not affected                               | High (markers easily occluded from cameras)           |
| <b>Signal noise</b>          | Sensitive to clutter/background | Sensor drift, calibration issues           | Very low (gold standard accuracy, mm range)           |
| <b>Setup requirements</b>    | Minimal (plug-and-play)         | Requires correct placement and calibration | Complex, requires multi-camera calibration            |
| <b>Natural movement</b>      | Preserved                       | Potentially hindered by sensors            | Preserved (markers are lightweight, but visible)      |
| <b>Cost</b>                  | Generally lower                 | Variable, often higher                     | Very high (specialized equipment and labs)            |
| <b>Medical constraints</b>   | Broad applicability             | Limited in skin sensitivity or hemiplegia  | Limited use in clinical routine due to cost and setup |

ingly precise and dynamic DTs, extending their application from motor recovery to cognitive and emotional rehabilitation as well [56]. DTs represent dynamic, data-driven replicas of patients that are continuously updated through multimodal inputs, including sensor data, clinical records, and behavioral information. By coupling real-time monitoring with predictive computational modeling, DTs enable bi-directional interaction between the physical and virtual domains.

Within rehabilitation, DTs can simulate therapy scenarios *in silico*, design personalized intervention plans, and refine strategies by evaluating patient performance and forecasting treatment outcomes. Reported applications include upper-limb rehabilitation, robot-assisted training, and the development of virtual prosthetics [57, 58]. By shifting the focus from reactive to predictive paradigms, DTs contribute to precision rehabilitation, enabling patient-specific and continuously adaptive care pathways.

### 2.2.3 Gesture Recognition

Gesture recognition moves from the mere acquisition of motion data to the meaningful interpretation of functional movements. Rather than simply tracking joint coordinates or hand trajectories, the goal is to identify and classify specific motor actions, such as shoulder abduction, finger flexion, hand grasp, or reach-to-grasp movements, which are highly relevant in daily activities and therapeutic



**Figure 2.4:** Assembly of the Virtual Glove system, consisting of two Leap Motion Controller (V1) sensors mounted orthogonally on an aluminum base. The image shows a potential user positioning the hand above the device, demonstrating the correct placement during task execution.

tic exercises. By identifying these maladaptive patterns, the system can provide targeted feedback to both patients and therapists, fostering more effective motor relearning. Ultimately, this capability lays the foundation for adaptive and personalized rehabilitation programs, where exercises are continuously tailored to the patient’s progress and needs.

From a technical perspective, gesture recognition goes beyond the mere acquisition of motion data, moving toward the meaningful interpretation of functional movements. Instead of simply tracking joint coordinates or hand trajectories, the goal is to identify and classify clinically relevant motor actions such as shoulder abduction, finger flexion, hand grasp, or reach-to-grasp gestures, which are fundamental for daily life activities and rehabilitation exercises.

Several methodological strategies have been proposed for this purpose. Early solutions relied on template matching, where the patient’s trajectory is directly compared with a reference or pre-recorded one. Among these, Dynamic Time Warping (DTW) gained wide adoption, as it aligns two temporal sequences by stretching or compressing the time axis, ensuring robustness to speed variations [59, 60]. Simpler implementations employ direct Euclidean Distance as a similarity measure, which is computationally efficient but less flexible.

Subsequent classical machine learning approaches improved recognition by extracting kinematic descriptors (e.g., velocity, acceleration, trajectory curvature) and using them as features for classifiers such as Support Vector Machines (SVMs), k-Nearest Neighbors (kNN), or Random Forests. These methods are relatively data-efficient and interpretable, although their performance depends heavily on the quality of handcrafted features and may deteriorate in high-dimensional or noisy contexts [59, 60].

Hidden Markov Models (HMMs) further advanced gesture recognition by modeling gestures as sequences of latent states with transition probabilities, naturally handling variable-length sequences and repetitive motions.

With the advent of deep learning, more sophisticated techniques emerged. Recurrent Neural Networks (RNNs), including LSTM and GRU variants, capture long-term temporal dependencies and can differentiate subtle variations between intended gestures and compensatory movements. CNNs have also been applied to gesture recognition, both on 1D sensor time-series and 2D/3D spatiotemporal data, automatically learning hierarchical features from raw inputs.

The latest progress has been driven by transformer-based architectures, which leverage self-attention mechanisms to model long sequences without recurrence. These models have demonstrated superior robustness and scalability, particularly in the presence of noisy or heterogeneous rehabilitation data, making them a promising foundation for intelligent, adaptive rehabilitation systems [61–63].

Lightweight variants such as quantized or binarized networks are particularly relevant when targeting **home-based or embedded rehabilitation systems**, where computational efficiency and low latency are crucial. Clinically, recognition verifies task execution and detects compensatory strategies (e.g., trunk inclination), enabling targeted feedback and adaptive, personalized therapy progression. A promising frontier to overcome the computational burden of deep models is represented by **Binarized Neural Networks (BNNs)**. In this thesis, BNNs were used to develop a lightweight hand gesture recognition model specifically designed for real-time execution on resource-constrained platforms, such as portable or head-mounted rehabilitation devices. By drastically reducing both the number of parameters and the computational complexity, BNNs enable efficient inference while maintaining competitive accuracy compared to their full-precision counterparts. Unlike conventional networks, which use high-precision (typically 32-bit floating point) weights and activations, BNNs constrain both parameters and intermediate representations to binary values (usually  $\{-1, +1\}$ ). This constraint drastically reduces the memory footprint and replaces costly arithmetic operations (multiplications and additions) with bit-wise operations (e.g., XNOR and bitcount), which can be executed extremely efficiently on embedded or low-power devices.

Formally, given an input tensor  $X$  and a kernel  $W$ , the output of a standard convolutional layer is:

$$Y[i, j] = \sum_{m=0}^{K_h-1} \sum_{n=0}^{K_w-1} X[i+m, j+n] \cdot W[m, n] + b$$

where  $K_h, K_w$  are the kernel dimensions and  $b$  is the bias term.

In a **BNN**, both  $X$  and  $W$  are binarized:

$$X_b = \text{sign}(X), \quad W_b = \text{sign}(W)$$

where  $\text{sign}(x) = +1$  if  $x \geq 0$  and  $-1$  otherwise. The convolution then reduces to:

$$Y_b[i, j] = \alpha \cdot (X_b \text{ XNOR } W_b)$$

where  $\alpha$  is a scaling factor (often computed as the mean of the absolute values of the original real-valued weights before binarization), and the XNOR operation can be efficiently implemented at the bit level, followed by a population count to obtain the number of matching bits. This replacement dramatically reduces the computational cost, since bitwise operations are orders of magnitude faster

than floating-point multiplications.

**2.2.3.0.1 Training and Backpropagation.** The main challenge in BNNs lies in training, because the  $\text{sign}(\cdot)$  function used for binarization is non-differentiable, and thus its gradient is zero almost everywhere. To address this, BNNs typically employ the **Straight-Through Estimator (STE)**.

- In the forward pass, real-valued weights  $W$  are binarized into  $W_b = \text{sign}(W)$ .
- In the backward pass, instead of propagating a zero gradient through the sign function, the STE approximates the derivative as:

$$\frac{\partial \text{sign}(x)}{\partial x} \approx \mathbb{I}_{|x| \leq 1}$$

meaning that the gradient is passed unchanged for inputs within  $[-1, 1]$  and set to zero outside this interval.

During training, the update is applied to the underlying real-valued weights  $W$  (kept in memory for optimization), but the binarized version  $W_b$  is used for the forward propagation. This way, optimization proceeds in the continuous domain while inference relies on the binarized weights.

**2.2.3.0.2 Advantages.**

- **Memory efficiency:** weights require only 1 bit instead of 32, achieving up to  $32\times$  memory savings.
- **Computational efficiency:** convolutions reduce to XNOR + bitcount operations, which are hardware-friendly and energy-efficient.
- **Deployment suitability:** models can be executed in real time on embedded platforms such as Raspberry Pi, FPGAs, or standalone VR devices.

**2.2.3.0.3 Limitations.** BNNs usually suffer from a drop in accuracy compared to full-precision models, especially for complex tasks. However, hybrid approaches (e.g., keeping the first and last layers in full precision, or using quantized instead of strictly binary activations) can mitigate this problem and achieve a balance between efficiency and performance.

A major challenge in quantitative physical rehabilitation is the lack of a unified standard for data representation. Currently, each device manufacturer or research group relies on proprietary formats, leading to fragmented and heterogeneous datasets. This absence of standardization has several consequences: Interoperability issues: data generated by different motion tracking systems cannot be easily exchanged or integrated. Incomplete clinical records: rehabilitation data are rarely linked to electronic health records (EHRs), preventing the creation of longitudinal patient histories. Lack of comparability: results from different devices, institutions, or studies are difficult to align, limiting cross-patient evaluations and multicenter research. Difficulty in defining common metrics: without shared data structures, therapists and researchers cannot establish reproducible scales or quantitative benchmarks. Fragmented ecosystem: the lack of harmonization slows down

the development of large-scale datasets and hinders the clinical adoption of innovative rehabilitation technologies. To overcome these barriers, the rehabilitation domain can draw inspiration from other areas of medicine where data standardization has already been successfully achieved. One of the most relevant examples is the Digital Imaging and Communications in Medicine (DICOM) standard. Originally developed for radiological imaging, DICOM has progressively expanded into a versatile and extensible framework capable of supporting diverse biomedical data, including time series such as ECG and EEG. Its widespread adoption, integration with hospital infrastructures (e.g., PACS), and ability to store both raw measurements and rich metadata make it a promising candidate for structuring rehabilitation data. Leveraging DICOM for QPR would enable interoperability, facilitate integration with clinical workflows, and foster the development of standardized evaluation metrics. In this thesis in order to overcome this issue we presented a DICOM-like standard to process and elaborate these data.

## Part II

# Hand Telerehabilitation

## Chapter 3

# Patient–Therapist Cooperative Hand Telerehabilitation through a Novel Framework Involving the Virtual Glove System

---

The results of this chapter are published in:

G. Placidi, A. Di Matteo, D. Lozzi, M. Polsinelli, E. Theodoridou, “Patient–therapist cooperative hand telerehabilitation through a novel framework involving the virtual glove system”, *Sensors*, vol. 23, no. 7, pp. 3463, 2023. DOI: 10.3390/s23073463

E. Theodoridou, A. Cacchio, A. Di Matteo, M. Polsinelli, and G. Placidi, “Materials and techniques for effective at-home rehabilitation for hand mobility restoration based on the Virtual Glove,” in *Proceedings of the 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2023, pp. 579–584. DOI: 10.1109/CBMS58004.2023.00282.

---

The increasing demand for accessible, home-based rehabilitation solutions, particularly for patients recovering from stroke or with impaired hand mobility, has driven the creation of flexible, low-cost, and high-precision systems that allow remote assessment and guidance by therapists. This chapter presents the design and development of a novel patient–therapist cooperative framework for hand telerehabilitation, centered on the use of the Virtual Glove (VG) system. While Chapter 3 introduced a portable head-mounted hand-tracking setup designed for mobile and outdoor applications, the present chapter focuses on a static, laboratory-oriented configuration. This shift enables higher tracking accuracy, greater system stability, and closer integration with supervised rehabilitation protocols, providing a complementary perspective within the same research framework. The VG, developed at the A2VI Lab, is a touchless multisensor device designed to capture high-resolution kinematic data of hand movements in real time (40 fps). The VG system is lightweight, cost-effective, and fully self-contained, making it suitable for real-world deployment in non-clinical environments. Leveraging its capabilities, a comprehensive telerehabilitation architecture is proposed, enabling patients to carry out therapeutic exercises at home while being continuously monitored and assisted by therapists, even under limited network connectivity conditions.

The chapter is structured in two main parts. The first introduces the Virtual Glove system, detailing its hardware architecture, calibration procedures, data acquisition strategy, and software components. Particu-



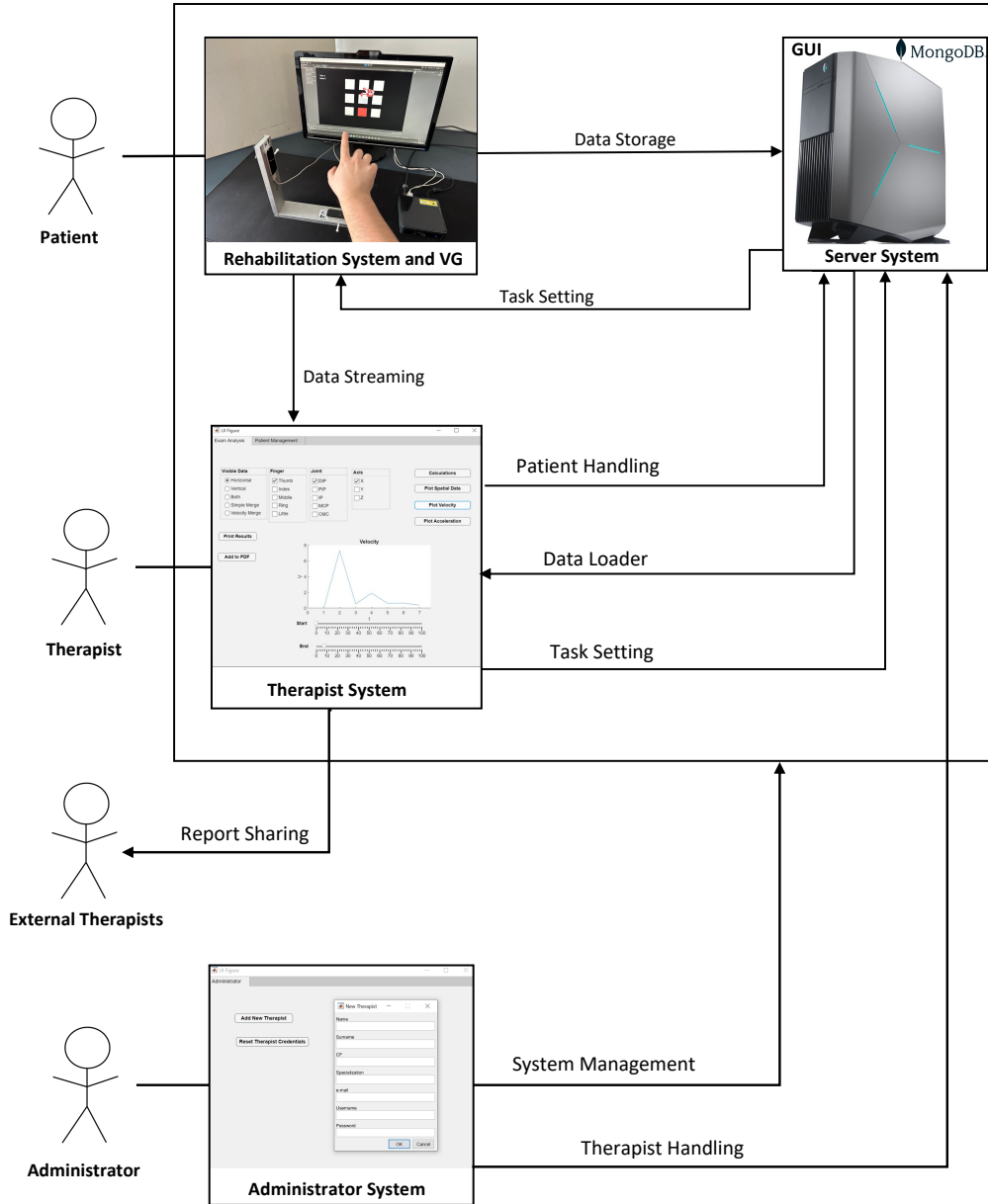
lar emphasis is placed on the multisensor setup, designed to minimize occlusions and ensure accurate hand tracking across different orientations.

The second part presents the integration of the Digital Twin paradigm into the telerehabilitation process, introducing the concept of a Double Digital Twin (DDT). In this approach, both the healthy and the impaired hands of the patient are modeled digitally, allowing the healthy hand to act as a reference for guiding the impaired one. This calibration-driven alignment enhances exercise accuracy, task repeatability, and rehabilitation efficacy.

Finally, the chapter details the software architecture, the cooperative roles of the patient, therapist, and administrator, and provides real-world examples of the system in use. The framework not only supports quantitative evaluation of hand mobility but also enables long-term data storage, multi-session monitoring, and therapist-to-therapist data sharing. These features make the system a promising tool both for clinical rehabilitation and for research purposes in motor function recovery.

### **3.1 The Cooperative Telerehabilitation Framework**

In the following, a description of the proposed telerehabilitation framework whose ensemble is reported in Figure 3.1. The persons involved in the process, the actors, are the Therapist, the Patient, and the system Administrator. From now on, these actors will be mentioned with a capital letter as a reference to their role in the system. The hardware consists of a Virtual Glove (VG) and a server for long-term data storage and for hosting the software for data analysis and presentation. This last software consists of a GUI for assisting the Therapist in real-time interaction with the Patient, data extraction, processing, and presentation. A particular view of the same software will assist the system Administrator to manage the whole system and to check system security. The VG is located in the Patient’s house, while the server is located in a clinical environment. The Patient uses the VG directly, and the VG transmits the data to the server through an Internet connection. The Therapist and the Administrator can access the server remotely. In what follows, the hardware and the software architecture are described in detail.



**Figure 3.1:** Hardware and software architecture of the framework. The hardware consists of the rehabilitation system with the VG and a server for data storage and for hosting the software. The software consists of a GUI for the Therapy and one for the Administrator. The persons involved in the process are the Therapist (or Therapists), the Patient and the system Administrator.

### 3.1.1 Hardware Architecture and Data Model

Hand tracking is ensured by VG [64], a multisensor touchless device. The advantages of touchless systems are that they do not interfere with the hand's free movement, and are user-independent, meaning that they do not demand personalized equipment or, in the other case, equipment that many people should touch [65]. Since the system is driven by hand movements, patients with minimal finger mobility can also utilize it. The software is designed to be handled both by full-hand and single finger movements, so patients can

navigate through the procedure and practice exercises designed according to their capabilities. In the case of closed hands (for example, as a result of a stroke) the exercises can be modified to include minimal finger movements, something that is possible thanks to the spatial resolution of a few millimeters that the VG provides. The important advantage of using multiple sensors is overcoming occlusion issues, which have always complicated hand tracking. In the case where no object is handled, one single sensor cannot track the hand and finger locations with detail when the palm is turned perpendicularly to the sensor.

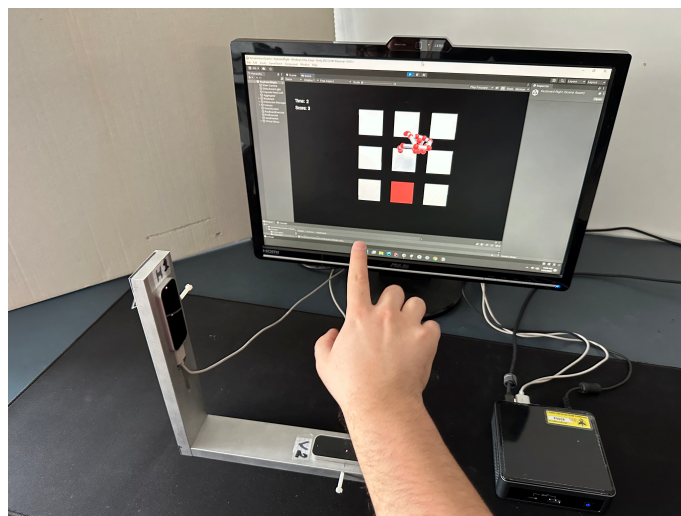
### 3.1.1.1 VG Architecture

The VG was first presented in [66, 67] using, then, a set of RGB sensors for hand tracking. In a new prototype [54, 68], the RGB sensors were replaced by two infrared Leap Motion Controller (LMC) sensors [69–71] that, among other advantages, provide higher spatial resolution than visual light systems with a reduced effect of lighting conditions ([72–76]), and their own hand model that offers direct positioning of each hand joint, without the use of markers. The LMC is a small stereo-vision-based device for 3D visualisation of the human hand movement. It includes three Infrared (IR) LED sources and two IR cameras. It offers a spatial resolution of a few millimeters and frequencies that make it suitable for finger movement tracking [77].

Here, the system is composed of two LMC sensors mounted on rigid aluminum support used to fix the sensors in an orthogonal set-up. The proposed stereotactic multisensor hand tracking system needs calibration to align different views in a single-hand model. Calibration requires a precise two-step procedure to ensure precision and stability. In the first step, an external cloud of points was acquired by both sensors and, using Singular Value Decomposition (SVD) analysis, rigid transformation matrices were acquired. For this step, the ability of the LMCs to detect not only fingers but also thin, pencil-like structures were used. A wooden stick was fixed to a numerically controlled mill, allowing it micrometric, three-axial, movements. By moving the stick in the three axes along the region of interest, measurements of its position were acquired by both sensors. This way, a cloud of points was produced and used for the calculation of the transformation matrices.

In a second step, to ensure that internal differences between the two sensors were taken into account, measurements of the stick position were taken first by one and then by the other sensor, each put in the exact same positions. This resulted in calculating small axial translations that were integrated into the initial transformation matrices.

The resulting transformation is automatically performed in real-time on the acquired data, during the data collection from the two sensors.



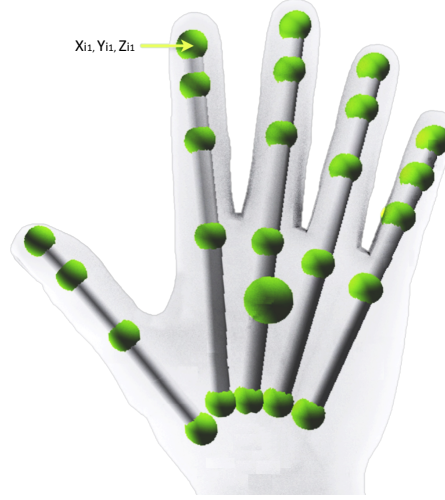
**Figure 3.2:** VG final implementation, including sensor assembly (left), NUC-based driving unit (right), and video screen for the active interaction (center).

The significant advantage of a multisensor design is the ability to measure data from different orientations and to reduce occlusions. The disadvantage of using multiple LMC sensors is that they demand an equal number of operating systems, as one operating system cannot control more than one LMC, and, even if recently it seems that this limitation is being overcome [78], it remains for the oldest LMC devices. For this reason, the original architecture of the VG used a primary-secondary system configuration where each sensor was controlled by one of two virtual machines (Secondaries) hosted on a physical machine (Primary), where the data fusion took place. In a second embodiment, leveraging advances in microprocessor technology, the PC hosting three servers was replaced by three Raspberry Pi (RP) machines [64], providing an economical, space-efficient solution. In this scheme, the Raspberry Pi operating system is installed on every RP, and each LMC is associated with a single machine. The third RP is where the data collection and fusion are performed, while it also hosts a virtual environment. In the RP version of this system, there was the constraint that the driver of the LMC needed to be run by an emulator of an X86 Machine with Windows OS on the Ubuntu OS RP. However, due to the progress of the LMC drivers, allowing those several LMC sensors to be managed by a single OS, the final deployment of the VG has been done by using a single MINI PC Intel NUC with Intel Core i5, 16 GB of RAM, and 256 GB of internal memory (Figure 3.2). The small size of the NUC gives us the advantage of portability and good performance.

By the described changes, a system has been created that, for the first time, does not require the use of one or more personal computers. Instead, it integrates all the necessary devices, except for the screen, into a single set, which can also be contained (and sold) in one box as a complete product.

#### **3.1.1.2 Storage System**

As mentioned before, it is essential that the Patient can perform rehabilitation exercises at home or in a laboratory under the supervision of the Therapist. To ensure the performance of telerehabilitation, the Server System is deployed on a PC with an Intel i7 processor, 32 GB of RAM, and an NVIDIA GeForce GTX 1080 to handle the large volume of data from the RS and TS. For the Therapist System and the Administrator System, the choice of PC is not crucial, as the software does not require high-performance specifications. The connection between the macro-components was made via the WAN network. The location of the server is indicated in the upper-right part of Figure 3.1. For data storage, MongoDB was chosen as a tool to design the database. Unlike SQL databases, MongoDB does not use a fixed scheme for its documents. Since the data from a LMC is a JSON object, MongoDB is the best choice for this project because no preprocessing is necessary before inserting the data into the database; the collecting system maintains the data structure to facilitate data sharing among therapists or with third-party software. MongoDB is one of the fastest databases for read and write operations. This also allows streaming the Patient's Task in real time to the Therapist or extending the number of LMC used for the VG with reduced effort.

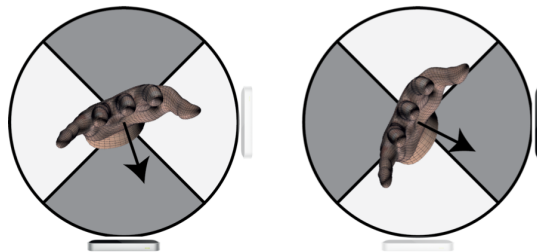


**Figure 3.3:** Skeletal hand model (skin contour is reproduced for reference) in which each joint is represented by a green sphere.

### 3.1.1.3 The Data Model

LMC provides a skeletal hand model where all joints are represented. Part of this model is used to obtain the hand skeletal model represented in Figure 3.3. In total, for each frame, the  $x$ ,  $y$ , and  $z$  coordinates of 25 points are given, together with the respective time stamps (the temporal position of the current frame): 4 joints for the thumb; 5 from all other fingers; 1 is to represent the center of the palm. For the specificity of VG, information containing three hand models is saved on the dataset: one from the vertical LMC, one from the horizontal LMC, and one produced by their fusion. This way, for each time instant, a JSON object is saved containing the hand models. Additional information provided by the LMC, such as the orientation of the palm normal or a reliability score for each joint collected by a sensor, are also saved in the header of the streaming file, together with person-specific and task-specific information, for future exploitation.

In the VG, data fusion is performed in two possible ways. The first one is a binary switch approach. Here the palm's normal direction (a vector orthogonal to the palm) defines what sensor contributes to the final result. More specifically, palm's normal direction is utilised to compute the angle between the X-axis of the horizontal LEAP reference system and the projection of  $p$  on the X-Y plane. If the angle is between 225 and 315 (palm facing down) or 45 and 135 (palm facing up), the horizontal LEAP is activated, but the vertical sensor data is ignored. Outside this range, the sensors' functions are reversed: vertical LEAP data becomes active, while horizontal LEAP data is disregarded (Figure 3.4).



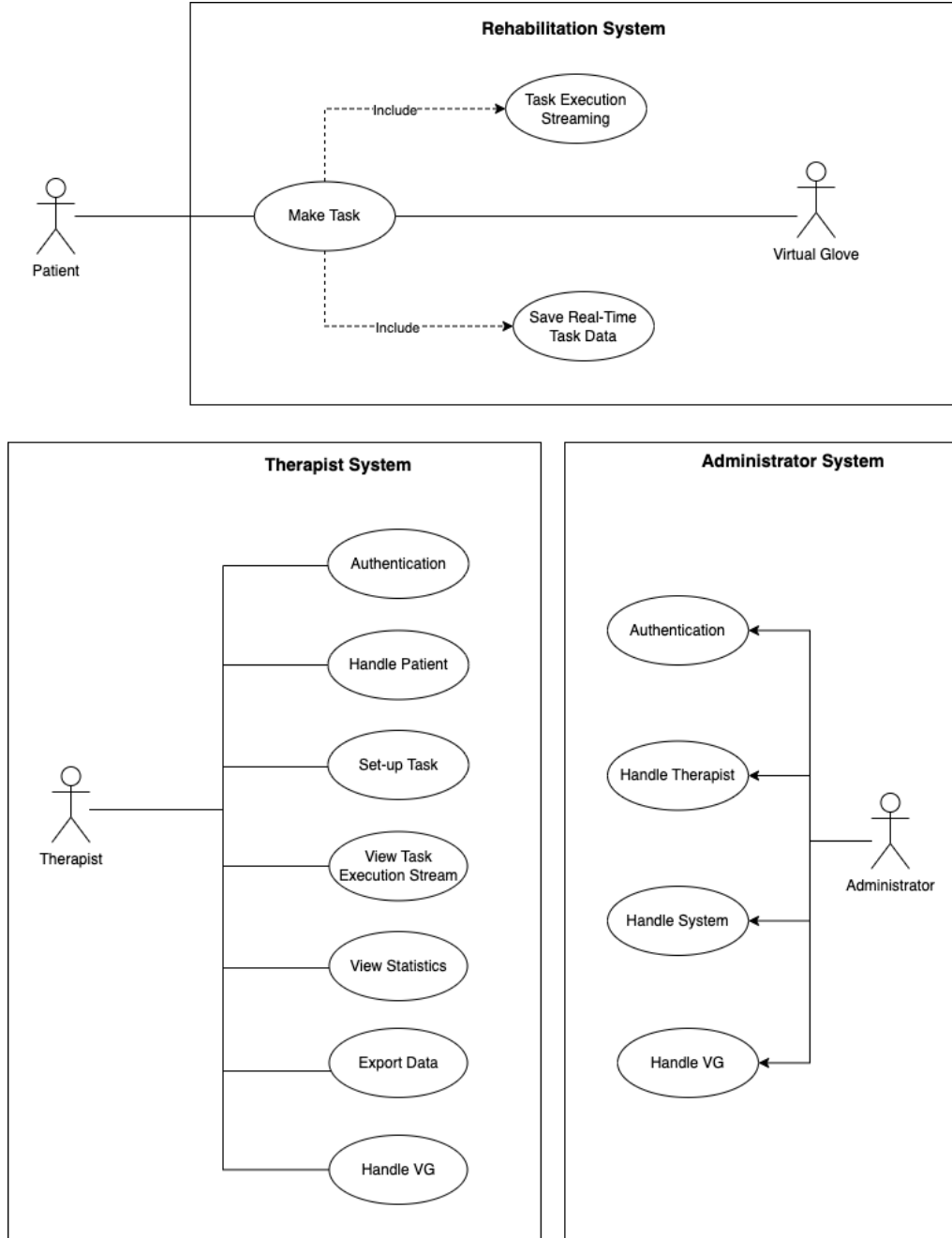
**Figure 3.4:** Binary switch approach: depending the orientation of the hand, one sensor data are used, while the others are discarded.

A second method is a joint-wise switch approach where instead of taking data from the whole hand, from the same sensor in each frame, some joint is collected by a sensor and some are collected by the other. Here, the selection is done by comparing the inter-sensor with the intra-sensor velocity of each joint, based on the

consideration that slower-tracked velocity probably signifies more robust data. Specifically, the metrics of intra- and inter-sensor joint speed are defined. If the inter-sensor spatial difference of the active sensor in consequent frames is higher than the intra-sensor one, this indicates tracking loss, and taking data from the other sensor is starting. This produces a final dataset where both sensors contribute to each frame [55].

For the sake of simplicity, in the embodiment described therein, just the first fusion method is used.

### 3.1.2 Use Cases and Software Architecture



**Figure 3.5:** Use Case diagram presenting the functionalities and actors involved to the framework usage.

The actors involved in the proposed framework are four: the Patient, the VG, the Therapist, and the Administrator (see Figure 3.5). The most important of those is the Patient, a person with hand mobility problems whose role is to perform rehabilitation tasks, as part of the rehabilitation procedure, using Rehabilitation

System (RS) combined with the VG. The Patient should only focus on carefully completing his assigned tasks to the best of his ability, without having to take care of anything else.

When the RS turns on, the task personalized for the Patient is automatically loaded. When the task is completed, the RS saves the data externally in the Server System. This data-storage procedure is very important since it allows the Patient to either run the task at home or in the presence of the Therapist, for a face-to-face rehabilitation session. The task has a double effect: motivation, needing a "competition" to reach a score; rehabilitation, driving the hand toward specific movements. The Therapist, once authenticated, can create, read, update and delete (CRUD) regarding the Patient's information, tasks, and session data.

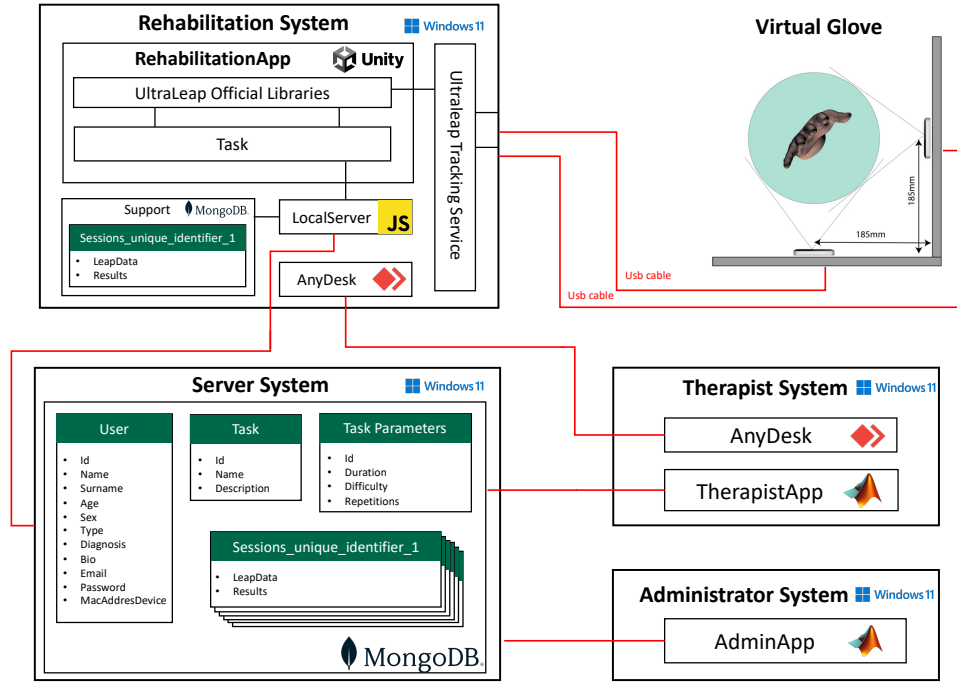
Specifically, for the "create" operation: the Therapist inserts Patient Information into the system and, additionally, assigns to the Patient a physical machine (the VG). Once the "insert" operation is completed, the Therapist assigns a task. Additionally, the Therapist defines the duration, difficulty, and repetition frequency. The Therapist is able to monitor a remotely performed session in real-time, using streaming between the Therapist System and the RS. This is particularly useful the first time a treated person performs a new task or tries new settings of the same task (mostly in the case of a new difficulty level). A Patient-Machine-Therapist connection is generated and the data files coming from that specific VG will contain, in its header, information about the Patient, the VG, the Therapist, and the executed task (type, difficulty level, score, etc.). A direct Therapist-Patient connection is also ensured by conventional channels, such as email and phone.

Once the task is complete and the data is stored from the RS to the Server System, the Therapist can load data and perform calculations and statistical analysis provided by the System through a specific GUI (detailed below). The Therapist can export a report on the resulting analysis and share it with colleagues for consultation and comparison. In the report, the Therapist can find the Patient's information, together with selected plots.

The Administrator is responsible for the smooth operation of the whole system. He first appoints the Therapist and assigns to him the personal credentials to access the system. The Administrator is also responsible for standard data security, system maintenance, crash recovery, and software upgrade.

### **3.1.2.1 Software Architecture**

The system is divided into four macro-components (Figure 3.6): the RS, the Server System (SS), the Therapist System (TS), and the Administrator System (AS). RS executes the rehabilitation task. SS saves rehabilitation files. TS allows the Therapist to perform data analysis and visualization. Finally, AS allows the Administrator to supervise and manage the whole activity. In the following, these components are described in detail.



**Figure 3.6:** Software architecture diagram representing all the system involved: Rehabilitation System, Server System, VG, Therapist System, and Administrator System.

**3.1.2.1.1 Rehabilitation System** The RS was created to handle the rehabilitation process. The goal of this component is to let the Patient perform the tasks in an Virtual Reality (VR) environment by using the VG. It also saves locally the session data related to the task and, once the session is complete, it transfers the data to the server, inside the MongoDB database.

To collect data from both LMC of the VG, the LMC GEMINI 5.7.2 driver was used: this is the first that can manage more than one LMC per operating system. This procedure allows the use of the UltraLeap Tracking Service (UTS) to collect the data from each LMC, build the numeric hand model, and share the corresponding model with the other components.

To use these data, the RehabilitationAPP component in the UNITY3D Game Engine was created. This application, using the data coming from the UTS, creates an VR environment with rehabilitation tasks. In detail, the UltraLeap Official Libraries (UOL) was integrated with the numeric hand models generated by the UTS. It is important to note that the hand data coming from the UTS are raw. The UOL applies a mirror transformation with respect to the X-axis, to present the virtual hand as seen from the point of view of the user (otherwise a right hand would be seen as a left hand, and vice versa). Then, translation, rotation, and scaling are applied to the LMCs to locate them inside the Unity world space. Similarly, the two LMCs composing the VG were placed inside the Unity World in the same positions as in the real world. The rigid transformation matrices obtained from the calibration were applied to the LMCs within the Unity environment: the translation and scales vectors as they were, whilst the rotation matrices in Euler Angles. Afterward, since the UOL allows fusing the hand models by using the procedures reported above, the switching algorithm in which only one hand data model is chosen for each frame was implemented for testing purposes .

Finally, the hand model was integrated into an example rehabilitation task consisting of a Keyboard in which some buttons enlighten randomly, and have to be turned off with parts of the hand that also enlighten.

Thus, when a rehabilitation session starts, the RehabilitationApp loads on the screen task-specific summary information, used to inform the Patient about the task he is approaching to execute. Then, by placing the hand on the Virtual Glove, the RehabilitationApp loads the task with the settled task parameters. A ten-second counter starts, and when it finishes, the task starts; this allows the sensors to better track the



hand for the execution. When the task time expires the RehabilitationApp checks if repetitions are necessary. If yes, RehabilitationApp loads a new screen in which the Patient can take a rest before putting his hand in the view of the VG to start the new repetition. If not, the system loads an end screen with the task execution summary.

All the real-time data generated by the VG during the rehabilitation session are saved in a Support instance of MongoDB. Since Unity does not have an official integration of MongoDB, a LocalServer application was created to save the data. The RehabilitationApp and LocalServer are connected by Websocket protocol. The Support instance is integrated to avoid that Internet connection issues could degrade the system performance. Once the Patient finishes a task, the local server transfers the data from the Support instance into the Server instance, from where it can be loaded and analyzed by the Therapist. Finally, the AnyDesk application was embodied to the RS to connect via streaming RS and TS.

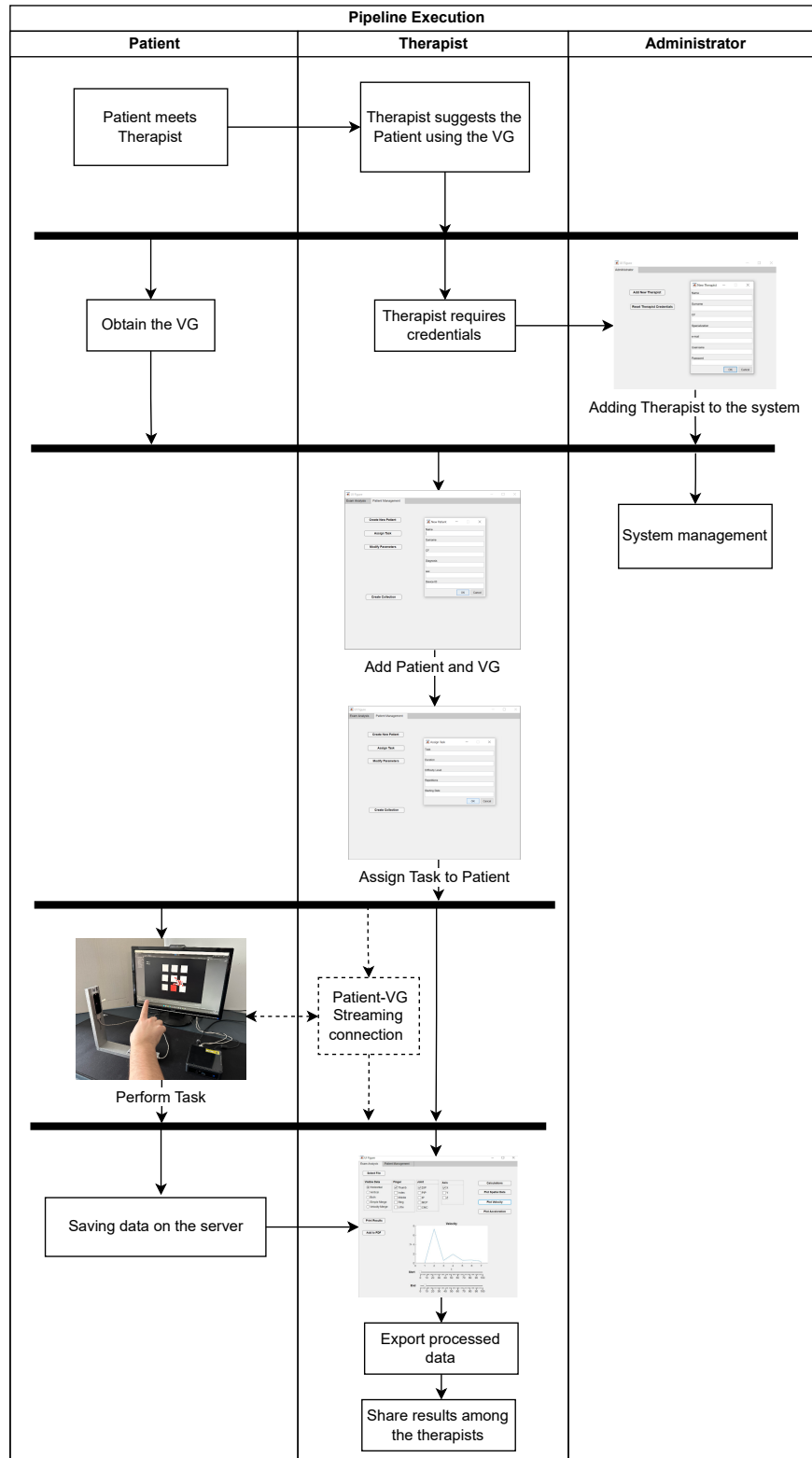
**3.1.2.1.2 The Server System** For this project, three MongoDB collections were created named “User”, “Task”, “Task\_Parameters”. Also,  $n$  collections called “Session\_unique\_identifier”. The “User” collection contains personal and identification information for each Patient or Therapist. The collection “Task” stores objects regarding the 3D interactive tasks (name, description). The collection “Task\_Parameters” contains information regarding configuration parameters of a specific task (duration, difficulty, repetitions and starting date). Session\_unique\_identifier is the collection where the data from each rehabilitation session are stored. The unique\_identifier is a string composed of id\_userp (Patient), id\_usert (Therapist), id\_task\_parameters, and current date. This way, each task session is saved in a unique collection with all necessary retrieval information on its name, and the execution time can be optimized for the queries by using a solution that only requires one collection for all the task sessions.

**3.1.2.1.3 The Therapist System** The TS was created aiming to allow therapists to administer their patients. To this aim, the TherapistAPP was developed. Its goal is to control in a simple and user-friendly mode, collections that contain Actors and task data. It includes finding tasks by using as query criteria the Patient’s name, id, task type, or execution time, analyzing and visualizing the movements, and saving the results in an easy-to-read pdf form. In addition, the interface gives the possibility of comparing results from different sessions, so that the Therapist can monitor a therapy plan’s progress and the effect of the therapy plan on the mobility of the hand. At the same time, through the app, it is possible to create new structures and practice programs for new or already existing patients. Throughout the procedure, the original data remain intact, available for replicating the analysis with different criteria. The app is organized into two different tabs. One is used for the analysis of preexisting task data, while the second one gives the user the possibility to insert new patients into the database and organize their task plan.

Although the app was created on MATLAB, its code has been converted into C++ using MATLAB Coder. In this macro-component, the AnyDesk application was integrated in order to ensure a streaming connection between the RS and the TS to let the Therapist monitor remotely the Patient during task execution.

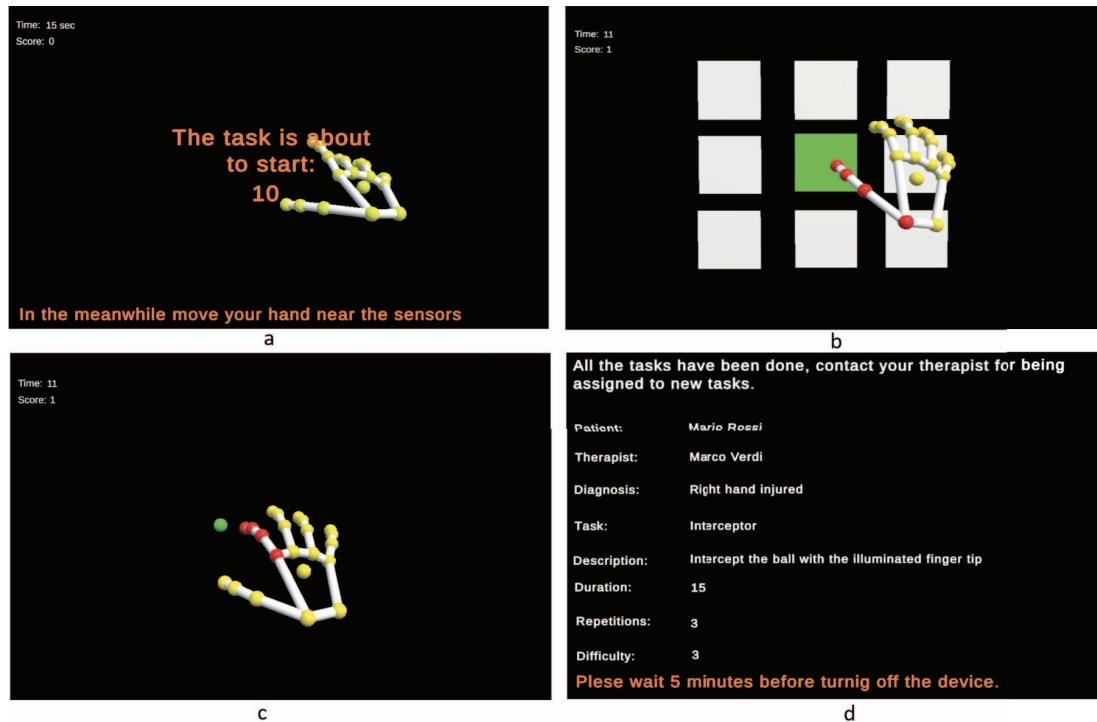
**3.1.2.1.4 Administrator System** The AS is a MATLAB application, the AdministratorApp, through which the Administrator is able to “create” a Therapist and allow him to manage data. Through the app, besides creating new Therapists, the Administrator can define/revoke Therapist’s credentials. Furthermore, the AdministratorApp allows the Administrator to edit information on Patient-Therapist links, allowing the Therapist assigned to a Patient to be changed if required. Also in this case MATLAB was used to define the AdministratorApp, then MATLAB Coder was used to generate the C++ code.

### 3.2 Execution example



**Figure 3.7:** Swim Lane Diagram showing the pipeline execution. The pipeline describes the rehabilitation process from the beginning when the Patient is suggested to use the VG until data are processed and shared among therapists. Obviously, the process can be continued by the Therapist, assigning a new task or updating the task’s difficulty, or restarting with a new Patient.





**Figure 3.9:** Summary screen loaded before Patient starts the task b) Keyboard task execution c) Interceptor task execution c) Summary screen loaded when all repetitions are completed.

### 3.3 Discussion and Conclusions

In this chapter, the structure with both hardware and software, of a novel system designed for at-home telerehabilitation sessions is presented. The novelty of this study lies in the fact that it is designed to give to the Therapist quantitative and not only qualitative information about the Patient’s mobility, for every hand joint separately, while at the same time providing the possibility to control the result of the rehabilitation plan by monitoring also quantitatively the progress of the hand mobility. The exercise plan has the form of simple, easy-to-understand, and easy-to-perform tasks, creating less fatigue in comparison with exercise plans composed of repetitive movements, thus making the task completion easier and more probable.

Although the main goal of this system is to be used for real rehabilitation purposes, it can also be used for research, since it provides information in an easy-access form, and it contains a database that can be used for sharing information between authorized people without risking the Patient’s privacy. For this, it helps also the fact that the hand movement is represented by a hand model, thus not even visual information needs to be shared.

The computational systems used are compact and relatively inexpensive, having replaced bulky personal computers with an Intel NUC MINI PC featuring an Intel Core i5 processor, 16 GB of RAM, and 256 GB of internal storage, resulting in a setup that fits within the space of a shoebox. It is economical enough to be a viable solution for patients that are in need of at-home rehabilitation but cannot afford the cost of multiple visits from a therapist. At the same time, being a personal device, it is safe for use by immunocompromised individuals, and suitable even in situations where self-isolation is needed.

A limitation of this study is that for now the system has been preliminary tested only by healthy individuals with ages from 25 to 30 years and with no diagnosed anatomical or functional hand issues. In these trials the system has been found to be effective. More precisely, people that took the “Patient” role have mentioned that the interfaces are easy to use, the guidelines of the tasks clear and that the exercise games work without issues. However, the system has not yet been tested by people with reduced hand mobility. In a future work, trials from patients will be performed and the collected data will be analyzed

with advanced multi-variable strategies.

The goal of this system is also to provide the therapists with a useful tool for performing and evaluating telerehabilitation practices. In a future application, the therapists can use the quantitative data the system provides for validating the hand mobility in a standard goal scale. This is possible because VG can be used with both hands: a standard goal scale can be calculated by analyzing data from the healthy hand of the same patient, thus making it possible to create a patient-specific goal scale. Furthermore, in a future work, the functionality of the system and especially of the GUIs needs to be evaluated by rehabilitation therapists on real or similar to reality therapy conditions, in order to identify possible points for improvement through all the parts of the architecture, including the possibility of constructing high-level interactive graphic user interfaces which allow the therapists to implement patient-specific or, better, infirmity-specific rehabilitation tasks, as in [79, 80].

Right now work, is in progress in two directions: to provide a more effective strategy for hand model fusion among different LMCs, and to build a standardized data format as a new tool for data-sharing among therapists. This can also contain other-source multi-modal information (besides task-related information, it could also include, for example, simultaneous EEG recordings).

## Chapter 4

# Double Digital Twin for the Hand Rehabilitation through the Virtual Glove

---

The results of this chapter are published in:

A. Di Matteo, D. Lozzi, E. Mattei, F. Mignosi, S. Montagna, M. Polsinelli, G. Placidi, “Calibration of the Double Digital Twin for the Hand Rehabilitation by the Virtual Glove”, *2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 521–526, 2024.

DOI: 10.1109/CBMS61543.2024.00092

A. Di Matteo, D. Lozzi, E. Mattei, S. Montagna, A. Cipriani, F. Mignosi, M. Polsinelli, G. Placidi, “Proactive Hand Rehabilitation Through a Double Digital Twin Provided by the Virtual Glove”, in *IEEE MetroX-RAINE 2024 - Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering*, 2024. DOI: 10.1109/MetroXRAINE62247.2024.10796497.

---

Hand rehabilitation is a complex and prolonged process that demands continuous engagement and precision in monitoring motor performance over time. To address this challenge, integrating Digital Twin (DT) technology with Virtual Reality (VR) environments has emerged as a promising strategy for delivering personalized, interactive, and adaptive rehabilitation programs. The DT paradigm enables the creation of a dynamic digital replica of a physical system, continuously updated through real-time data acquisition. Applied to rehabilitation, this approach allows for modeling of patient-specific biomechanics and supports the delivery of tailored exercises guided by quantitative feedback.

In this context, the Virtual Glove (VG) enables the acquisition of synchronized three-dimensional data from both the impaired and healthy hands, allowing the construction of detailed digital representations of each. This capability supports the implementation of an innovative framework known as the Double Digital Twin (DDT). The DDT involves generating two parallel digital models, one for the healthy hand and the other for the impaired hand, allowing real-time comparison, accurate motion analysis, and continuous progress monitoring throughout the rehabilitation process.

This chapter introduces a methodology based on the DDT paradigm that enables the execution of rehabilitation tasks in a VR environment while capturing and analyzing patient performance at a fine-grained level. Using mirrored tasks and alignment techniques, the movement patterns of the healthy hand are transformed into a reference model (ground truth), against which the impaired hand’s movements are continuously compared. The system provides both visual guidance to the patient and quantitative metrics to the therapist, fostering a proactive rehabilitation process. The remainder of this chapter is organized as follows. Section 4.1 introduces the background and main concepts of the Digital Twin paradigm in

healthcare. Section 4.2 describes the proposed Double Digital Twin (DDT) system for hand rehabilitation, detailing its architecture and data acquisition process. Section 4.3 outlines the methodology, including task design, calibration, and comparative analysis between healthy and impaired hands. Section 4.4 presents the experimental results and discussion. Finally, Section 4.5 draws the conclusions and highlights future developments.

## 4.1 Background

In the last decade, the DT paradigm has been explored across different domains as an approach to virtualize entities in the real world, creating software counterparts that provide smart services on them [81, 82]. They facilitate data acquisition and storage by definition, as they remain continuously linked with the physical system, exchanging information and constructing a digital thread that tracks the evolution of the asset's condition. Additionally, they offer a range of tools, including simulation and cognitive services, to model data and knowledge, predict future states, and analyze them to support real-world decision-making.

In recent years, they have attracted significant attention also in the medical sector, where their potential to improve diagnosis, therapy, and patient management has been recognized [83, 84]. The potential applications of DTs in healthcare are numerous and varied, spanning beyond the more traditional applications in manufacturing, including medical devices and hospital processes, to simulating human organs and individuals themselves.

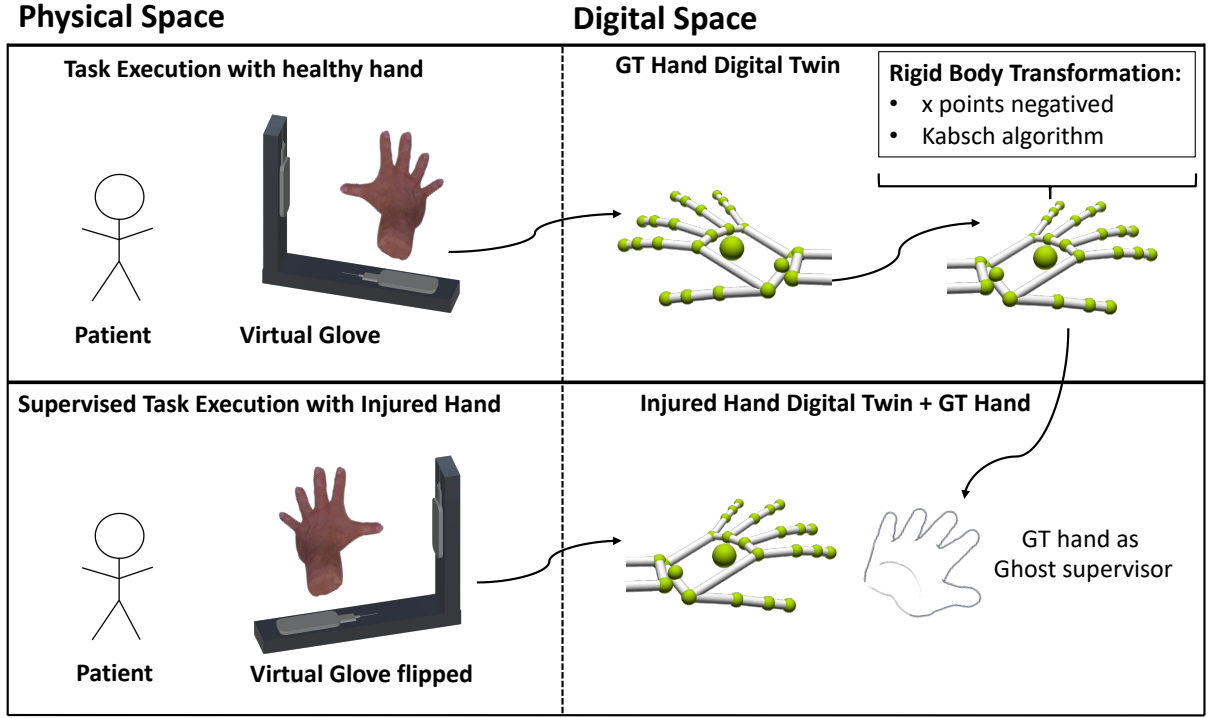
Without the claim to be exhaustive, example applications in literature are, for instance, those that adopt this approach to build a human body or human organ (i.e., heart [85]) DT, being applied across various medical conditions, like cardiovascular diseases or diabetes [86]. In this context, potential benefits of DT adoption have been envisioned for precision [87] and personalized medicine [88], where the integration and processing of vast amounts of data are mandatory to define highly customized treatments and interventions. Their objective is to collect data and predict the dynamic evolution of patients' health or organs and to evaluate the efficacy of personalized therapies [89].

Another research trend proposes the application of DTs to healthcare management, with the goal of optimizing resource allocation efficiency while minimizing resource shortages and defining appropriate clinical care plans. In this context, DTs typically serve to capture, aggregate, and analyse operational data in order to evaluate the system performance and, accordingly, design optimization strategies [90]. Several studies have explored digital twin (DT) approaches as part of rehabilitation frameworks for the hand, as highlighted by works such as [58, 91–93]. Despite these advancements, none have integrated the DT approach with the concept of utilizing data from a healthy hand for comparative analysis with an injured hand. This gap in research includes the transformation of the coordinate system from a healthy to an injured hand, which could potentially enhance rehabilitation strategies. This paper focuses on the organ DT, creating a 4D virtual digital counterpart of each hand to support individuals with hand impairments during rehabilitation sessions. Furthermore, this represents the first endeavor within the literature to employ DTs for such a purpose.

## 4.2 The Double Digital Twin System

In this chapter, a model for a DTs-based system aimed at replicating hand movements is presented. The system is composed of two DTs, as shown in Fig. 4.1, modeling the left and right hand, the so-called Double Digital Twin. The DT model includes measures performed on the physical world over time, enabling the computation of parameters for reconstructing and modeling hand movements. This facilitates the definition of rehabilitation tasks tailored to the patient. The advantage of employing DTs to model hand movements lies in their ability to track and process these movements accurately over time and provide Therapists with complete and real-time information.

To acquire data on the physical world, each DT is coupled with a VG. The VG incorporates two LMCs, each generating a dynamic 4D numerical model of the hand. This model is structured into frames, where each frame represents a temporal snapshot of the hand at a specific moment (Fig. 4.2). These frames capture detailed information about any detected hands, including the precise positioning of 25 distinct joints with a timestamp. Each joint represents the articulation between two bones, providing precise coordinates in a three-dimensional space ( $x, y, z$ ) along with calculated metrics for velocity and acceleration. In addition



**Figure 4.1:** DDT formation. First, the DT of the healthy hand (DTH) is formed by a single execution of the task, generating the Ground Truth. Then, a repeated rehabilitation sessions is performed with the impaired hand, generating the DT for the impaired hand (DTI).

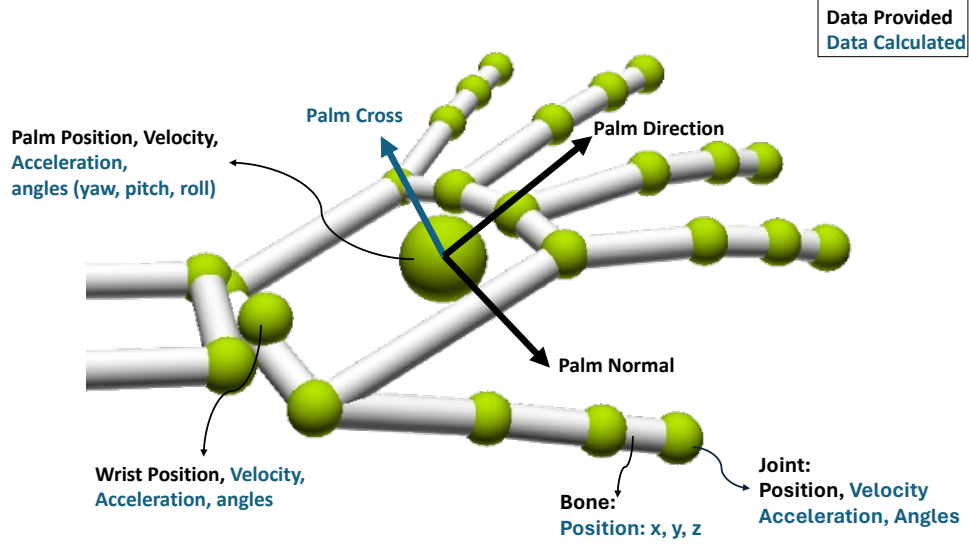
to joint data, the system also captures other hand-related information, such as the position and velocity of the palm. Furthermore, it provides two unit vectors: the palm normal and the palm direction. The former points from the palm towards the fingers, and the latter indicates the direction the palm faces relative to the controller. These vectors are exploited to compute the inner product and subsequently derive a rotation matrix for the hand. This matrix facilitates the computation of the hand's orientation angles, yaw, pitch, and roll, as well as the angle between any two bones. Velocity and acceleration are also calculated for each joint, exploiting the two adjacent frames. Finally the two LMCs have been calibrated to ensure that the views from both LMCs are aligned, enhancing the integration of data, and then using the roll angle the data provided from the optimal viewer are used. This algorithm ensures that the data from the optimal viewer is used at any instant [54].

### 4.3 Methodology

The framework around the DDT unfolds as follows:

1. creation of a rehabilitation task using Unity3D, in a low code manner
2. execution of the task with the healthy hand as Ground Truth (GT)
3. execution of the task with the injured hand in a supervised fashion with GT as supervisor
4. quantitative and qualitative comparison of the two executions





**Figure 4.2:** DT Internal Model: Data provided are highlighted in black; Data processed and provided are highlighted in skyblue

### 4.3.1 Therapist Task Creation

The framework in Unity3D [94] enables therapists to design rehabilitation tasks in a low-code style, enhancing customization and adaptability for patient-specific needs. The system offers two predefined tasks, Interceptor and Keyboard [95], with parameters that can be adjusted according to the therapist's specifications. In the Interceptor task, patients are required to touch balls that spawn and move towards their hand. The therapist can configure the number and position of the balls. In the Keyboard task, patients must touch various illuminated keys; the therapist can set the number of keys, their positions, the pattern of illumination, and the duration each key remains lit. A third task revolves around gestures. Here, the therapist constructs an ordered "dictionary" where each gesture is described along with its required execution time. During therapy sessions, patients follow instructions displayed on a label for a specified duration. To support the execution for both hands, a mirrored version of the Interceptor and Keyboard is automatically generated. For the Keyboard task, the entire layout is rotated 180 degrees around a central pivot point, effectively creating a mirror version of the original keyboard configuration. In the Interceptor task, mirroring the position of the spawned balls is simpler. The mirrored position is calculated by subtracting the ball's position from the sum of the VR environment's width ( $\text{Min} + \text{Max}$ ).

### 4.3.2 Task Execution with Healthy Hand

To acquire the DT of the healthy hand (DTH), it is necessary to acquire it as the injured patient's hand type. However, to make it possible, some elaborations are necessary. In fact, it is not possible to compare the left hand with the right hand (and vice versa) due to the chirality problem. Chirality means that a hand cannot be superimposed onto its mirror hand, regardless of their similar shapes. For this reason, the hand data will be transformed with the following two steps: firstly, the x-axis of the horizontal LMC is multiplied by -1, and secondly, the points of the vertical LMC are aligned to the horizontal one using the Kabsch algorithm [96, 97]. Given Two sets of corresponding points,  $P$  and  $Q$ , each containing  $n$  points the Kabsch algorithm involves the following steps:

- Calculate the centroids of  $P$  and  $Q$ :

$$C_P = \frac{1}{n} \sum_{i=1}^n P_i,$$

$$C_Q = \frac{1}{n} \sum_{i=1}^n Q_i.$$

- Center the points around the origin:

$$P'_i = P_i - C_P,$$

$$Q'_i = Q_i - C_Q.$$

- Compute the covariance matrix  $A$ :

$$A = \sum_{i=1}^n P'_i Q'^T_i.$$

- Compute the singular value decomposition (SVD) of  $A$ :

$$A = U \Sigma V^T.$$

- Calculate the rotation matrix  $R$ :

$$R = V U^T.$$

- Check for a proper rotation:

$$\text{If } \det(R) < 0, \text{ then } R = V \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} U^T.$$

- Output the rotation matrix  $R$  and calculate the RMSD:

$$\text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n \|Q_i - (R P_i + C_Q)\|^2}.$$

This process is done until a constant epsilon (0.02) is less than the Root Mean Squared Deviation. Thus, there is this alignment before the task is executed. It is important to emphasize that this procedure is carried out without the patient's awareness. Therefore, from the patient's perspective, the execution appears entirely ordinary, with his real hand reconstructed as it is and visualised during the exercise.

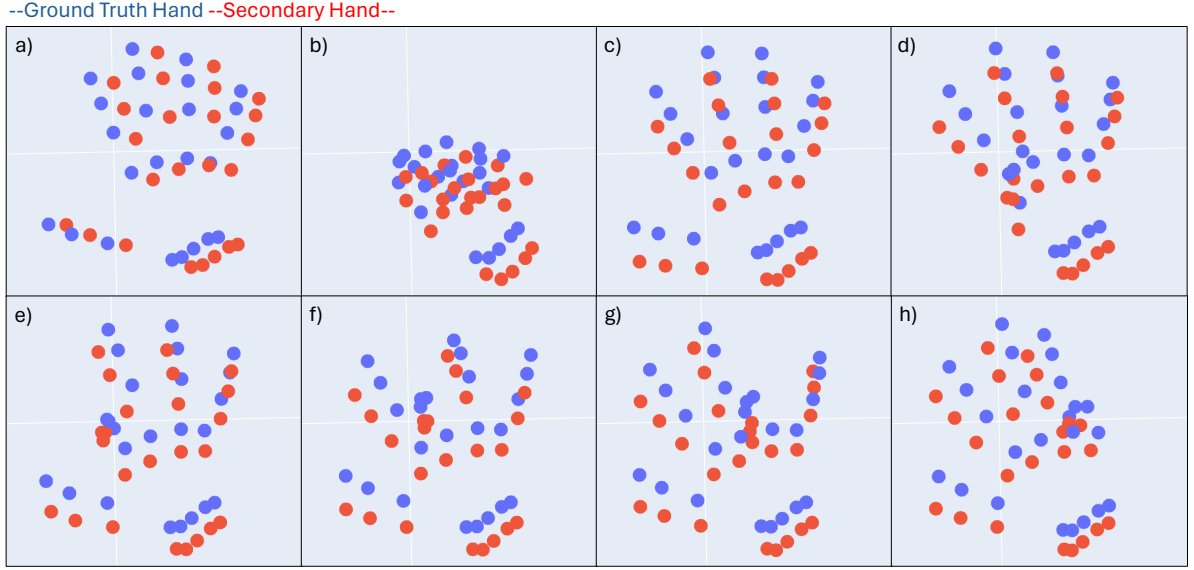
### 4.3.3 Task Execution with Injured Hand

To construct the DT of the Injured Model (DTI), the same task previously done with the healthy hand has to be completed with the injured hand. To do it, the VG is flipped, to maintain the condition that the palm is visible from the vertical LMC. Afterwards, the mirrored task is loaded, and the patient will start it. During the task, the hand model of the healthy hand previously acquired is played back alongside the hand model of the injured hand that is currently executing the exercise. The patient has to do his best to match the same movement done with the healthy hand with the injured hand, trying also to match the

execution time as closely as possible. To facilitate this task, the healthy hand is rendered on the screen like a semi-transparent glove that the injured hand has to "wear" and follow during the duration of the tasks.

#### 4.3.4 Quantitative and Qualitative Evaluation

Data from DTH and DTI are saved in MongoDB [98], and exploited using Python [99]. To enable qualitative analysis of the data, a custom software was developed to render and replay hand motion data using the Plotly library [100]. The therapist can move forward and backward in time and pause the playback to focus on a particular moment. Moreover, it is possible to change the camera perspective to focus on a particular point of view, that for instance could be a particular finger or hand's joint (Fig. 4.3).



**Figure 4.3:** Qualitative evaluation of the discrepancies during the execution of the task, displaying various frames: a) Opened Hand; b) Closed Hand; c) Opened Hand ; d) Thumb Flexion; e) Index Flexion; f) Middle Flexion; g) Ring Flexion; h) Little Flexion.

In terms of quantitative evaluation, the motion performance between the two hands may vary significantly, reflecting differences in individual skill levels. Additionally, when attempting to perform the same task with the opposite hand, variations and inconsistencies are likely to occur. These can manifest as delays or differences in movement. Such variations primarily arise from the inherent difference in dexterity between a person's dominant and non-dominant hand. Even if the skill levels between the hands are similar, discrepancies in speed and precision are expected due to the dominance of one hand over the other. To overcome this issue as a comparison algorithm, the Dynamic Time Warping (DTW) algorithm[101] was used, using the the TSlearn [102] library. Thus, in this case, repeating the same task with a different hand ("impaired") can have lag or velocity displacements. The algorithm aims to find an optimal match between sequences by non-linearly warping them in the time dimension to measure similarity independently of time variations, and then it outputs the distance between sequences, indicating their similarity. The algorithm involves the following steps:

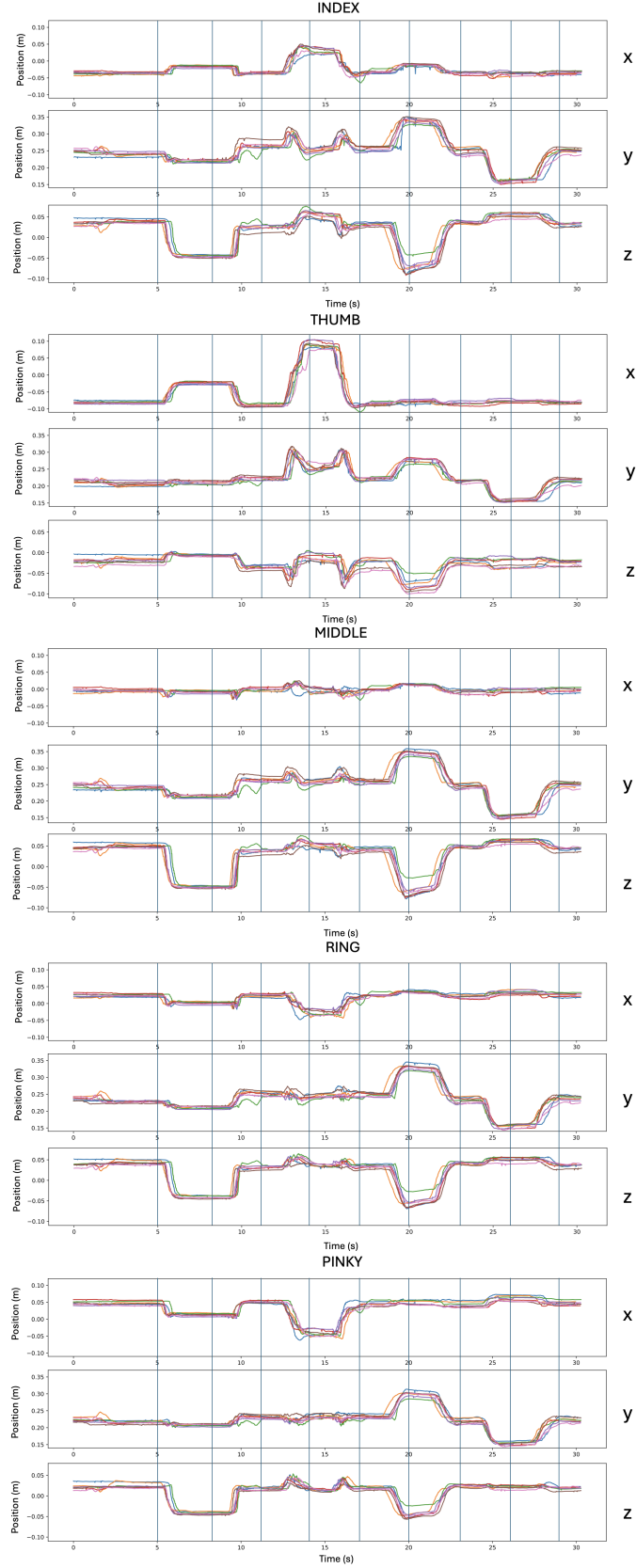
- **Distance Matrix Calculation:** Construct a matrix where each element represents the distance between points in the first and second sequences (typically using Euclidean distance).
- **Accumulated Distance Matrix:** Initialize the matrix's first row and column, then populate the rest using a recurrence relation that accumulates alignment costs.
- **Warping Path:** Identify the path that minimizes the total distance between the sequences by backtracking from the matrix's bottom right to the top left. This path shows corresponding points in

optimal alignment.

- **Normalization:** Normalize the total distance by the number of steps in the warping path to make it independent of path length, allowing comparisons between sequence pairs.
- **Output:** The optimal distance measure reflects sequence similarity, with lower distances indicating higher similarity.

## 4.4 Results and Discussion

In total, seven acquisitions were conducted with a user who has both hands in healthy condition. The task lasted approximately 30 seconds, capturing a total of approximately 1700 4D positions (x, y, z, and t), and included the following gestures: Start (5 seconds); Fist (3 seconds); Relax (3 seconds); Wrist supination (3 seconds); Relax (3 seconds); Move hand upward, maintaining the finger positions open (3 seconds); Relax (3 seconds); Move hand downward, maintaining the finger positions open; Relax (3 seconds). The first acquisition was performed to acquire the DTGT using the right hand, while the subsequent acquisitions were conducted using the left hand. The hand models were saved in MongoDB [98], along with the starting point for each trigger. Figure 4.4 represents the trajectories of the five fingertips, organized by axes. The figure shows seven lines: data from the DTH (colored line) and data from the DTI. As observed, the time series are quite similar across all plots. From a quantitative perspective, the DTW algorithm was used between the fingertips of the right and left hands according to x, y, and z positions for each time window (Table 4.1, 4.2). The results indicate that the executions with the left hand (DTI) closely mirrored those of the right hand, suggesting that the difficulty of future gestures should be increased. Furthermore, this is evident also from the DTW tables, the first execution is slightly minor compared to the other ones, or also the first executions, except for some rare cases. Thus, there are no constant improvements with the acquisition of the subsequent DTI, leading to the fact that the task is too simple for the user. It is important to note that this approach has been tested only on a healthy person; further experiments will be conducted with impaired subjects.



**Figure 4.4:** Fingertip trajectories on the x, y, and z axes for all five fingers during the 29-second task. Blue = DTH (right hand); other colors = DTIs (left hand).

**Table 4.1:** DTW distance (m) for each fingertip in X, Y, and Z position between DTH and each DTI for the "fist" (left) and "wrist supination" (right) gestures (**Less is better**)

| Fist   |   |        |        |        |        |        |        | Wrist Supination |   |        |        |        |        |        |        |
|--------|---|--------|--------|--------|--------|--------|--------|------------------|---|--------|--------|--------|--------|--------|--------|
|        |   | 1      | 2      | 3      | 4      | 5      | 6      |                  |   | 1      | 2      | 3      | 4      | 5      | 6      |
| THUMB  | x | 0.0519 | 0.0923 | 0.0810 | 0.0596 | 0.0441 | 0.0430 | THUMB            | x | 0.0474 | 0.0625 | 0.1844 | 0.2071 | 0.0842 | 0.0469 |
|        | y | 0.0241 | 0.1011 | 0.0389 | 0.1115 | 0.0816 | 0.0501 |                  | y | 0.0511 | 0.0405 | 0.0699 | 0.0310 | 0.1281 | 0.1936 |
|        | z | 0.0556 | 0.1106 | 0.0618 | 0.0684 | 0.1054 | 0.1534 |                  | z | 0.0621 | 0.1935 | 0.1482 | 0.2018 | 0.1160 | 0.0761 |
| INDEX  | x | 0.0309 | 0.0298 | 0.0315 | 0.0891 | 0.0367 | 0.0399 | INDEX            | x | 0.1179 | 0.1333 | 0.2211 | 0.2142 | 0.2107 | 0.0546 |
|        | y | 0.0524 | 0.0453 | 0.0934 | 0.0766 | 0.1063 | 0.0846 |                  | y | 0.0394 | 0.1823 | 0.0809 | 0.1315 | 0.1634 | 0.0358 |
|        | z | 0.0350 | 0.0447 | 0.0637 | 0.0668 | 0.1022 | 0.0921 |                  | z | 0.0992 | 0.2071 | 0.1497 | 0.1759 | 0.0817 | 0.0858 |
| MIDDLE | x | 0.0120 | 0.0267 | 0.0417 | 0.0982 | 0.0456 | 0.0476 | MIDDLE           | x | 0.0780 | 0.0841 | 0.1415 | 0.1719 | 0.1632 | 0.0810 |
|        | y | 0.0477 | 0.0187 | 0.0829 | 0.0720 | 0.0915 | 0.0697 |                  | y | 0.0382 | 0.1606 | 0.0763 | 0.1305 | 0.1071 | 0.0547 |
|        | z | 0.0272 | 0.0432 | 0.0644 | 0.0640 | 0.0931 | 0.0927 |                  | z | 0.0755 | 0.1435 | 0.1210 | 0.1624 | 0.0452 | 0.1065 |
| RING   | x | 0.0200 | 0.0565 | 0.0513 | 0.1023 | 0.0451 | 0.0529 | RING             | x | 0.0326 | 0.1015 | 0.1440 | 0.1974 | 0.1021 | 0.0798 |
|        | y | 0.0283 | 0.0246 | 0.0616 | 0.0513 | 0.0873 | 0.0541 |                  | y | 0.0409 | 0.1101 | 0.0772 | 0.1112 | 0.0924 | 0.0884 |
|        | z | 0.0184 | 0.0373 | 0.0775 | 0.0779 | 0.1069 | 0.0928 |                  | z | 0.0189 | 0.0229 | 0.0429 | 0.0453 | 0.0550 | 0.0375 |
| PINKY  | x | 0.0382 | 0.0611 | 0.0565 | 0.1005 | 0.0415 | 0.0640 | PINKY            | x | 0.0413 | 0.1046 | 0.1680 | 0.2229 | 0.0712 | 0.0588 |
|        | y | 0.0213 | 0.0306 | 0.0605 | 0.0686 | 0.0916 | 0.0509 |                  | y | 0.0465 | 0.0662 | 0.0806 | 0.0904 | 0.0836 | 0.0874 |
|        | z | 0.0140 | 0.0656 | 0.0914 | 0.0926 | 0.1184 | 0.1028 |                  | z | 0.0151 | 0.0235 | 0.0419 | 0.0156 | 0.0709 | 0.0412 |

**Table 4.2:** DTW distance (m) for each fingertip in X, Y, and Z position between DTH and each DTI for the "move hand upward" (left) and "move hand downward" (right) gestures (**Less is better**)

| Upward |   |        |        |        |        |        | Downward |        |   |        |        |        |        |        |        |
|--------|---|--------|--------|--------|--------|--------|----------|--------|---|--------|--------|--------|--------|--------|--------|
|        |   | 1      | 2      | 3      | 4      | 5      | 6        |        |   | 1      | 2      | 3      | 4      | 5      | 6      |
| THUMB  | x | 0.0399 | 0.1146 | 0.0878 | 0.1479 | 0.0463 | 0.1006   | THUMB  | x | 0.0165 | 0.0058 | 0.0782 | 0.0760 | 0.0046 | 0.0181 |
|        | y | 0.0149 | 0.1044 | 0.1119 | 0.0651 | 0.0853 | 0.0875   |        | y | 0.0121 | 0.0261 | 0.0109 | 0.0256 | 0.0810 | 0.0425 |
|        | z | 0.0629 | 0.1538 | 0.1742 | 0.1292 | 0.2479 | 0.3200   |        | z | 0.0284 | 0.2457 | 0.2034 | 0.2726 | 0.0793 | 0.0564 |
| INDEX  | x | 0.0428 | 0.1158 | 0.0664 | 0.0421 | 0.0844 | 0.0520   | INDEX  | x | 0.0997 | 0.0154 | 0.0930 | 0.0356 | 0.0172 | 0.0774 |
|        | y | 0.1096 | 0.1647 | 0.0847 | 0.0999 | 0.1077 | 0.0569   |        | y | 0.0243 | 0.0274 | 0.0682 | 0.0559 | 0.0348 | 0.0721 |
|        | z | 0.0500 | 0.3250 | 0.0561 | 0.0865 | 0.0787 | 0.0708   |        | z | 0.0438 | 0.0393 | 0.0365 | 0.0096 | 0.0638 | 0.1347 |
| MIDDLE | x | 0.0239 | 0.0863 | 0.0364 | 0.0183 | 0.0270 | 0.0509   | MIDDLE | x | 0.0269 | 0.0136 | 0.1445 | 0.0468 | 0.0828 | 0.0912 |
|        | y | 0.0865 | 0.2062 | 0.0859 | 0.1639 | 0.0782 | 0.1050   |        | y | 0.0227 | 0.0316 | 0.0772 | 0.0516 | 0.0258 | 0.0359 |
|        | z | 0.0940 | 0.3798 | 0.0632 | 0.1147 | 0.0904 | 0.0892   |        | z | 0.0222 | 0.0263 | 0.0165 | 0.0059 | 0.0544 | 0.1062 |
| RING   | x | 0.0444 | 0.0822 | 0.0577 | 0.0551 | 0.0497 | 0.0811   | RING   | x | 0.0635 | 0.0332 | 0.1592 | 0.0593 | 0.0767 | 0.0925 |
|        | y | 0.0984 | 0.2216 | 0.1028 | 0.1950 | 0.0921 | 0.1393   |        | y | 0.0304 | 0.0217 | 0.1075 | 0.0916 | 0.0388 | 0.1175 |
|        | z | 0.0686 | 0.3175 | 0.0558 | 0.1031 | 0.0870 | 0.0836   |        | z | 0.0484 | 0.0144 | 0.0066 | 0.0443 | 0.0611 | 0.1582 |
| PINKY  | x | 0.0675 | 0.0334 | 0.1192 | 0.1207 | 0.1679 | 0.1594   | PINKY  | x | 0.0358 | 0.0855 | 0.2169 | 0.1197 | 0.2419 | 0.2491 |
|        | y | 0.0952 | 0.2561 | 0.1038 | 0.1865 | 0.0817 | 0.1129   |        | y | 0.0612 | 0.0581 | 0.1254 | 0.1182 | 0.0868 | 0.1630 |
|        | z | 0.0304 | 0.2273 | 0.0344 | 0.0523 | 0.0680 | 0.0543   |        | z | 0.0487 | 0.0146 | 0.0102 | 0.0512 | 0.0682 | 0.1182 |

## 4.5 Conclusion

In this chapter, a new method for quantitative hand rehabilitation using Digital Twin (DT) technology, the VG, and VR is presented. the concept of a Double Digital Twin (DDT) for hands, enabling analytics, real-time comparisons, and assessments is presented. The incorporation of VG is essential as it facilitates data collection from both hands in a touchless manner, without the need for additional equipment. These models interact with a VR environment to synchronize and ensure the consistency of exercises. This approach allows us to create a personalized DT for the healthy hand, known as the Ground Truth (GT) DT, which serves as the benchmark for recovery. Simultaneously, multiple DTs for the impaired hand are generated and compared to the GT DT. The GT DT and the impaired DTs inform the design of the DDT, which calculates discrepancy trends to guide therapists in monitoring and adjusting the rehabilitation process predictively.

This proposed DDT supports a personalized, proactive, and optimized rehabilitation strategy. Furthermore, this study highlights two main points: Following the main hand could be beneficial for the patient, offering a clear point of reference for guidance. From the therapist's perspective, this approach could provide a more focused and robust method for enhancing rehabilitation comparison and study. Methodologically, it can offer a concise indicator of progress. In conclusion, this method helps therapists to increase/decrease the complexity of tasks based on the evaluation approach. However, this work was tested on only one healthy person. Thus, further research on impaired subjects will be essential to validate its efficacy.

The integration of VG and DT allowed us to develop the Double Digital Twin (DDT) to model both injured and healthy hands for comparison. The purpose of this chapter is to introduce a new methodology to support rehabilitation through DDT that allows the patient to be stimulated and makes the exercises interactive.

## Chapter 5

# Portable Head-Mounted System for Mobile Forearm Tracking

---

The results of this chapter are published in:

M. Polsinelli, A. Di Matteo, D. Lozzi, E. Mattei, F. Mignosi, L. Nazzicone, V. Stornelli, G. Placidi, “Portable Head-Mounted System for Mobile Forearm Tracking”, *Sensors*, vol. 24, no. 7, pp. 2227, 2024. DOI: 10.3390/s24072227

---

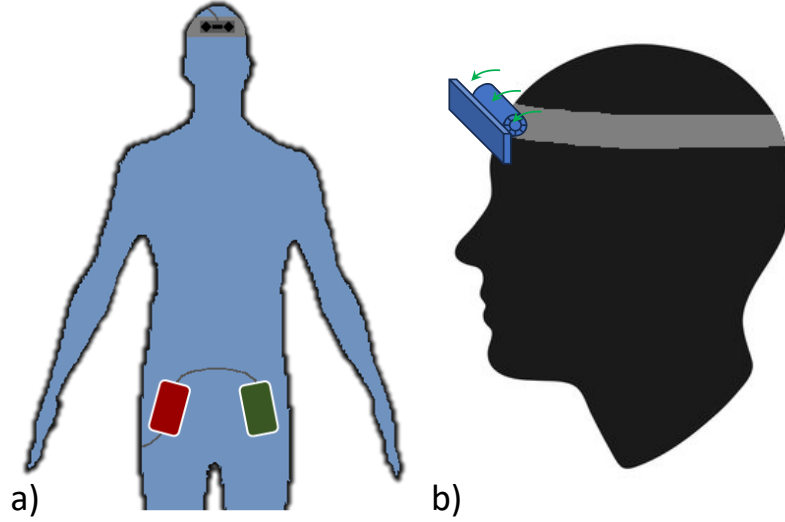
In recent years, there has been growing interest in developing lightweight, low-cost, and portable systems capable of capturing hand motion data in unconstrained environments. This is particularly relevant in healthcare applications such as motor rehabilitation, where continuous monitoring and assessment of hand movements, both at home and in outdoor settings, can enhance patient outcomes and promote autonomy. the need for a personal computer or external infrastructure.

In this chapter, a portable hand-tracking system based on an LMC (version 2 released in 2023) controlled by a single-board computer (SBC) powered by a power bank is presented. Figure 5.1a illustrates an embodiment of the proposed system: the LMC is mounted on the head and connected to the SBC powered by a power bank. Both the SBC and the power bank are compact and can be carried in a pocket. Head-mounted support has been adapted to accommodate the LMC, allowing for the adjustment of the orientation of the sensor (as shown in Figure 5.1b) according to the user’s requirements.

In this embodiment, the proposed system can be adapted for several applications, leaving both hands completely free. Moreover, the system is independent of an external power supply, and its weight is minimal, making it suitable for outdoor applications.

The remainder of this chapter is organized as follows. Section 5.1 reports the related works; Section 5.2 describes the proposed system in detail; Section 5.3 presents a use case and some experimental tests, assessing the robustness of the proposed device under variable lighting conditions, as well as its power consumption, CPU usage, temperature, and frame rate; and finally, Section 5.4 presents the conclusions and outlines future work.





**Figure 5.1:** System embodiment: (a) pocket location of the SBC and power bank with respect to the head-mounted LMC; (b) details of the LMC support that enables changing the vertical orientation of the sensor.

## 5.1 Related Works

One of the first works to utilize a head-mounted sensor for 2D hand-gesture recognition was [103], which employed a single RGB camera to minimize weight.

In [104], a head-mounted gaze-directed camera was used for 3D hand-tracking of everyday manipulation tasks. However, the system needed a powerful computational machine to achieve tracking at 12 frames per second (fps).

In [105], two RGB head-mounted synchronized cameras were proposed (the two cameras were calibrated offline) for 3D hand-gesture recognition, similar to [104]. The captured videos, with frames sized at  $320 \times 240$ , were processed at about 30 fps using a dedicated personal computer. The RGB system was lightweight (65 g), but its bottleneck was the personal computer, which was heavy and bulky.

In 2013, the first version of the LMC was presented for 3D hand tracking. It was a revolutionary device, being very lightweight (only 32 g), compact ( $80 \times 30 \times 11.30$  mm), fast (between 40 fps to 120 fps, depending on the managing hardware), and precise (about 1 mm) [54, 106]. In the beginning, the LMC was designed to be placed and used on a desk and connected to a personal computer, and several works appeared using it in disparate applications [107–112].

However, the sensor's features made it very suitable for head-mounted hand tracking, and in 2014, the software of the LMC was updated to support VR tracking mode, designed to provide hand tracking while the device was mounted on virtual reality headsets [113].

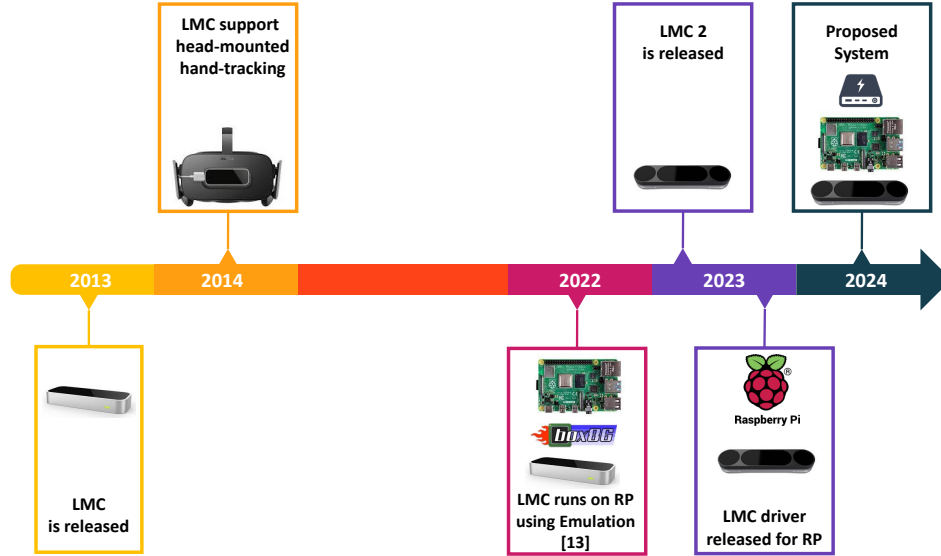
In the following years, several works were presented using the LMC in a head-mounted assembly, also in combination with VR/AR headsets [114–120], confirming the sensor's suitability for this type of task.

Nowadays, several VR/AR headsets have built-in hand-tracking capabilities, like Microsoft HoloLens 2 [121–123] and Meta Quest [124, 125].

However, in general, VR/AR headsets either continue to be connected to a personal computer (which limits their usable space) or have expensive built-in CPUs that mostly hinder access to raw data (such as Meta Quest [126]).

Recently, in [127], the authors proposed a system based on an LMC connected to a mini PC powered by a very powerful power bank. The LMC is head-mounted, whereas the mini PC and power bank are placed inside a backpack that the user has to wear. Indeed, this design allows for complete mobility by the user, but it is uncomfortable. As the authors pointed out, the LMC software requires an X<sub>86</sub> machine, limiting

the possibility of using a lightweight SBC. However, in [64], the authors demonstrated that it is possible to run the LMC software on an SBC equipped with an ARM CPU architecture through virtualization, reaching almost 30 fps. The SBC used was a Raspberry Pi 4 Model B (<https://www.raspberrypi.org/>), (abbreviated as RP). Finally, in December 2023, the LMC manufacturer released a new version that supports the RP [128, 129]. Figure 5.2 depicts the timeline of the LMC.



**Figure 5.2:** Timeline of the LMC: The sensor was first designed to be driven by a personal computer. In recent embodiments over the last two years, it has started to be driven by an SBC. Further, in its latest release, the ROI was increased and back-hand tracking was enabled, allowing for complete head-mounted embodiments.

The design and development of a portable forearm-tracking system requires satisfying several critical requirements: reduced costs and energy consumption, increased battery life, decreased temperature, reduced CPU usage, and maximized frame rate, among others. Without meeting these requirements, a forearm-tracking system would be incomplete. The proposed system herein satisfies the requirements for complete forearm tracking with an extensible ROI. Moreover, the overall cost of the system was kept low without compromising precision.

## 5.2 System Design

The LMC is a compact device designed for precise hand and finger tracking in three-dimensional space. Measuring just  $84 \text{ mm} \times 20 \text{ mm} \times 12 \text{ mm}$  and weighing only 29 grams, it offers portability and ease of use. With low power consumption requirements of 5 volts and 500 milliamps, it efficiently integrates into various setups without compromising performance [130]. The controller uses a combination of infrared sensors and cameras to track the movements of hands and fingers with high accuracy [106]. The drivers for the LMC are compatible with a wide range of desktop CPUs, operating systems, and VR/AR headsets. Recently, they have also been made available for the RP, resulting in minimized energy consumption and optimized performance for the LMC itself.

The RP used is equipped with a Broadcom BCM2711 CPU, Quad-core Cortex-A72 (ARM v8) 64-bit SoC at 1.8 GHz, and 4 GB of RAM, with the RP enclosed in its case. This is a powerful and low-cost RP with low dimensions and weight ( $88 \text{ mm} \times 58 \text{ mm} \times 19.5 \text{ mm}$  and 46 g) and low energy consumption (max 15 W). As a result, the RP stands out among the SOCs that best meet the requirements of the proposed system.

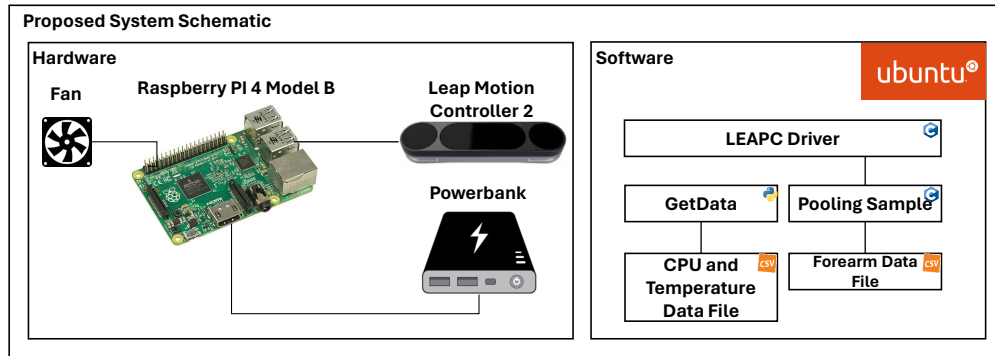
Given that the LMC now supports the RP, the natural synergy between them makes their combination an obvious choice. Together, they provide precise hand-tracking capabilities, enhancing the overall functionality and usability of the system. Moreover, the proposed system is conveniently powered by an external portable power supply, ensuring flexibility and ease of use in various settings.

In fact, through a USB-C port, the RP is connected to a power bank (with dimensions of 105 mm × 68 mm × 20 mm, weight of 190 g, and capacity of 10,000 mAh) capable of providing a maximum power of 20 W. The RP is connected to the LMC through a USB 3.0 port.

The design of the proposed system is shown in Figure 5.3. The RP is placed into a protective case (not shown in Figure 5.3 but indicated as 1 in Figure 5.4) that prevents it from being exposed to dust, humidity, and impacts that could damage the board. The RP is equipped with passive heat dissipation (heatsinks) and a fan for active dissipation to avoid overheating due to the high-intensity operations required for tracking, both hosted in the protective case.

The RP runs on the Ubuntu 20.04 operating system optimized for ARM architectures. The official LMC software for the RP was installed, and the LeapC API was used for the development of the scripts for data acquisition.

The “Pooling Sample” script, provided with the LMC software, was modified to enable data saving. The initial modification was to set the acquisition mode to head-mounted, running a specific function defined in the LeapC API. After the conclusion of this phase, data frame streaming was initialized. For each frame, data regarding the frame ID, the frame rate, the hand IDs, and all the coordinates of the joints of each hand are extracted and saved in a CSV file. Figure 5.3 also reports the “GetData” script, developed in Python 3.9, which, although not part of the final system embodiment, is used to collect data on CPU usage and temperature behavior.



**Figure 5.3:** System architecture.

A detailed illustration of how the proposed system operates, including the workflows and the interactions between the hardware and software components, is shown in Figure 5.5.

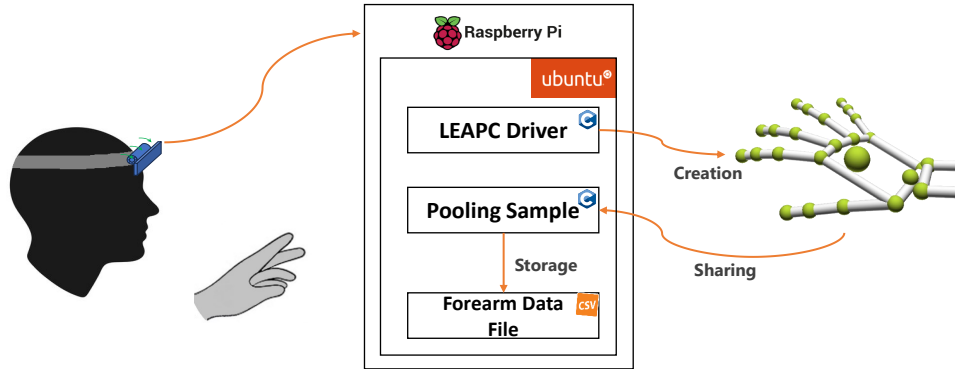
The LMC captures hand movements within a three-dimensional coordinate system that is centered around the device itself. The LeapC driver is hosted on the RP and acquires the raw data from the LMC. From the raw data, the LMC software creates a numerical model of the hand represented as a stick model, reflecting the real anatomy of the human hand, as shown in the right part of Figure 5.5.

In particular, the LMC data are organized into frames. Each frame represents a distinct snapshot in time and contains all the data captured by the controller at that moment. Within each frame, information about detected hands is provided. This includes data such as hand ID, palm position (x, y, z coordinates), palm velocity, palm normal (the direction the palm is facing), and palm direction (the direction the palm is moving). For each detected hand, detailed information about individual fingers is provided. This includes data such as finger ID, finger type (thumb, index finger, middle finger, ring finger, pinky finger), fingertip position (x, y, z coordinates), finger direction, and finger velocity. Additionally, the LMC can capture raw image data through its infrared cameras. However, these data are primarily used for internal processing and

calibration and can potentially be accessed by developers for specialized applications. For this reason, in this work, hand data provided by the API of the LMC are used because they are precise and reliable [106], and the algorithm used by the controller is designed to be efficient for use with applications that require a high frame rate. Finally, the Pooling Sample script establishes the connection with the LeapC driver and receives the LMC data frames. The script was modified to concatenate the data frames, convert each into comma-separated values (CSV), and save the data in a structured table format.



**Figure 5.4:** System embodiment: (1) the RP enclosed in its case, (2) the power bank, (3) the LMC inside the head-mounted support. The presence of the monitor, not used in the final embodiment, is necessary during system calibration to set the values of the parameters.



**Figure 5.5:** Workflow and interactions between the hardware and software components of the proposed system.

Throughout all the tests conducted, the Graphical User Interface (GUI) was intentionally disabled. This measure was implemented with the dual purpose of conserving energy and minimizing CPU usage. It is crucial to note that this particular characteristic is not a limitation; rather, it aligns with the system's design intention in scenarios where user–system interaction does not require visual interface elements. For this reason, the absence of a GUI and the option to do away with a monitor altogether serve to optimize the system for its intended purpose.

Figure 5.4 shows the final embodiment of the proposed system, arranged on a table. As stated above, the monitor is only used for calibration purposes, and for this reason, it is powered by an external power source.

In the scenario depicted in Figure 5.4, the RP is running the 'Pooling Sample' script, which prints the frame number and some specific numerical information regarding the forearms. In these conditions, where the GUI of the OS and the print of the strings in the system terminal are slowing down the system's performance, the frame rate is about 60 fps, providing near real-time capabilities.

## 5.3 Measurements

The proposed system is characterized by its portability, for which power consumption estimation is very important because it directly affects battery duration. Additionally, CPU usage and its temperature are other fundamental parameters to consider in portable applications, for which the RP could be placed in areas with little airflow, such as pockets.

For the above reasons, some measurements, tests, and evaluations are required for the complete characterization of the system. The tests were conducted considering two different scenarios: the RP without the case and the RP inside the case with a fan.

The first scenario is necessary to establish baseline measurements to be used for reference. The second scenario is the operative one and represents the real operating conditions.

For each scenario, four different configurations were tested:

1. **RP is on but without a connected LMC (RP)**: The RP is not actively involved in intensive operations or processes, consuming its lower-bound power and generating a lower-bound heat compared to when it is engaged in demanding tasks.
2. **LMC is connected to RP but not tracking (RP + LMC)**: Compared to just the RP, the LeapD daemon is running but data streaming is not requested by any software.
3. **LMC is connected to RP and data from forearm tracking is requested (RP + LMC + Data)**: Compared to the RP + LMC task, a script for data acquisition, provided as an example by the LMC manufacturer, is running. The script requests data from the LeapD.
4. **LMC is connected to the RP, forearm tracking is running, and data are printed on the console (RP + LMC + Data + Print)**: Compared to the RP + LMC + Data task, the script is also printing data on the console. This configuration represents the worst-case (upper-bound) scenario for power consumption since, for the considered portable applications, the monitor is not used.

For all the above configurations, the OS GUI was disabled, leaving only the system terminal rendered on the screen, the latter being powered by an independent power source. The only energy drawn from the power bank pertains to video data transmission via the HDMI port, which is considered negligible.

For these experiments, the proposed system was arranged on a table, and the LMC was placed on the head of the user. For configurations 3 and 4, the user typed on a keyboard and interacted with the smartphone to simulate different hand poses and occlusions. The methodology and results are reported in Sections 5.3.1 and 5.3.2.

Another fundamental aspect to consider is the frame rate of the proposed system. The frame rate of the LMC is variable and dependent on several factors, such as CPU availability, lighting conditions, the presence of hand occlusions, etc. For this reason, the proposed system was tested in real operational scenarios, in which the user interacted with different mechanical tools and under different lighting conditions. The methodology and results are reported in Section 5.3.3.

### 5.3.1 Power Consumption

For each scenario and configuration, electric voltage, current, and total power consumption are reported. Measurements were obtained using a USB tester placed between the USB-C power input of the RP and the

USB-C power output of the power bank. The sensitivities of the USB tester used are 0.01 V, 0.01 A, and 0.01 W.

The data were collected every 10 s for a total of 170 s. Table 5.1 reports the averaged results for both the baseline scenario (when the RP was not inside the case and without a fan) and the operating conditions (when the RP was inside the case and in the presence of a switched-on fan).

|          | Baseline Scenario (no case, no fan) |           |                 |                         | Operating Conditions (inside a case, fan on) |           |                 |                         |
|----------|-------------------------------------|-----------|-----------------|-------------------------|----------------------------------------------|-----------|-----------------|-------------------------|
|          | RP                                  | RP + LMC  | RP + LMC + Data | RP + LMC + Data + Print | RP                                           | RP + LMC  | RP + LMC + Data | RP + LMC + Data + Print |
| <b>V</b> | Avg: 5,05                           | Avg: 5,01 | Avg: 5,00       | Avg: 5,06               | Avg: 5,05                                    | Avg: 5,03 | Avg: 4,97       | Avg: 4,98               |
|          | Min: 5,04                           | Min: 5,00 | Min: 4,99       | Min: 5,05               | Min: 5,03                                    | Min: 5,02 | Min: 4,96       | Min: 4,97               |
|          | Max: 5,06                           | Max: 5,02 | Max: 5,01       | Max: 5,06               | Max: 5,06                                    | Max: 5,04 | Max: 4,98       | Max: 5,01               |
| <b>A</b> | Avg: 0,53                           | Avg: 0,84 | Avg: 1,19       | Avg: 1,30               | Avg: 0,58                                    | Avg: 1,03 | Avg: 1,37       | Avg: 1,38               |
|          | Min: 0,50                           | Min: 0,79 | Min: 1,15       | Min: 1,26               | Min: 0,55                                    | Min: 1,00 | Min: 1,30       | Min: 1,32               |
|          | Max: 0,62                           | Max: 0,95 | Max: 1,29       | Max: 1,36               | Max: 0,65                                    | Max: 1,08 | Max: 1,45       | Max: 1,48               |
| <b>W</b> | Avg: 2,68                           | Avg: 4,19 | Avg: 5,97       | Avg: 6,55               | Avg: 2,91                                    | Avg: 5,20 | Avg: 6,81       | Avg: 6,85               |
|          | Min: 2,55                           | Min: 3,95 | Min: 5,75       | Min: 6,34               | Min: 2,80                                    | Min: 5,02 | Min: 6,49       | Min: 6,62               |
|          | Max: 3,15                           | Max: 4,75 | Max: 6,47       | Max: 6,85               | Max: 3,29                                    | Max: 5,42 | Max: 7,20       | Max: 7,39               |

**Table 5.1:** Voltage, current, and power consumption in two different scenarios: Baseline scenario (left) and Operating conditions (right). In the operating conditions, the use of a fan is also considered to maintain the operating temperature low. Averaged measurements are reported.

Starting from the RP without the case (baseline scenario), the electric voltage was almost the same across all tasks, ranging between 4.99 V and 5.06 V.

Instead, the absorbed current varied depending on the task. In the first configuration (RP), the current consumption ranged between 0.50 A and 0.62 A (average 0.53 A), and in the RP + LMC configuration, it ranged between 0.79 A and 0.95 A (average 0.84 A). Consequentially, based on the measurements, 0.33 A was absorbed by the LMC, confirming the datasheet of the manufacturer, which stated that the LMC can require up to 0.5 A [130]. In fact, in this task, the LeapD was running, and the IR sensors were on, but there was no data request from any application. In the RP + LMC + Data configuration, the absorbed current ranged between 1.15 A and 1.29 A (average 1.19 A), and finally, in the RP + LMC + Data + Print configuration, it ranged between 1.26 A and 1.36 A (average 1.30 A). The increased current can mainly be attributed to the increased CPU usage.

Finally, active power consumption [W] was measured, which can also be calculated as  $P = V \times I$ , where  $V$  is the electric voltage in volts [V] and  $I$  is the current in amperes [A]. The measured values of  $P$  were compared with the calculated ones, yielding identical results. When the first configuration was considered (RP), the average power consumption ranged between 2.55 W and 3.15 W (average 2.68). In the RP + LMC configuration, it ranged between 3.95 W and 4.75 W (average 4.19 W); in the RP + LMC + Data configuration, it ranged between 5.75 W and 6.47 W; and in the RP + LMC + Data + Print configuration, it ranged between 5.75 W and 6.47 W (average 5.97 W). Since, as previously stated, the electric voltage remained constant across all tasks, the differences in power consumption can be attributed to the increment

in current consumption, and the same considerations hold.

The same measurements were repeated with the RP using the operating configuration (fan on). In this case, the electric voltage was almost the same for the RP and the RP + LMC configurations, but a slight drop was measured for the RP + LMC + Data and the RP + LMC + Data + Print configurations. Regarding current consumption, in the RP configuration, the values ranged between 0.55 A and 0.65 A (average 0.58 A). Compared with the RP without a fan, the current consumption increased by 0.05 A. After the connection of the LMC, the average current consumption increased to 1.03 A, ranging between 1.00 and 1.08. Compared with the RP without a fan, the current consumption increased by 0.19 A. In the RP + LMC + Data configuration, the average current consumption increased to 1.37 A, ranging between 1.30 and 1.45, with an increase of 0.18 A compared to the same configuration without the fan. Finally, in the RP + LMC + Data + Print configuration, the average current consumption increased to 1.38 A, ranging between 1.32 and 1.48, with an increase of 0.08 A. On average, compared with the 'fanless' configuration, the proposed configuration required 0.13 A, in line with the current consumption stated by the manufacturer of the fan. The increments in power consumption compared to the fanless configuration were 0.23 W, 1.01 W, 0.84 W, and 0.30 W, which equates to an average of 0.60 W. The increased power consumption cannot be totally attributed to the fan. The introduction of the fan has an acceptable impact on power consumption, but as discussed in the next section, it is fundamental to keep the RP temperature under control.

Finally, Figure 5.6 reports the plots for the baseline scenario and the operating conditions: the relevant result is that all measurements exhibit stability over time.

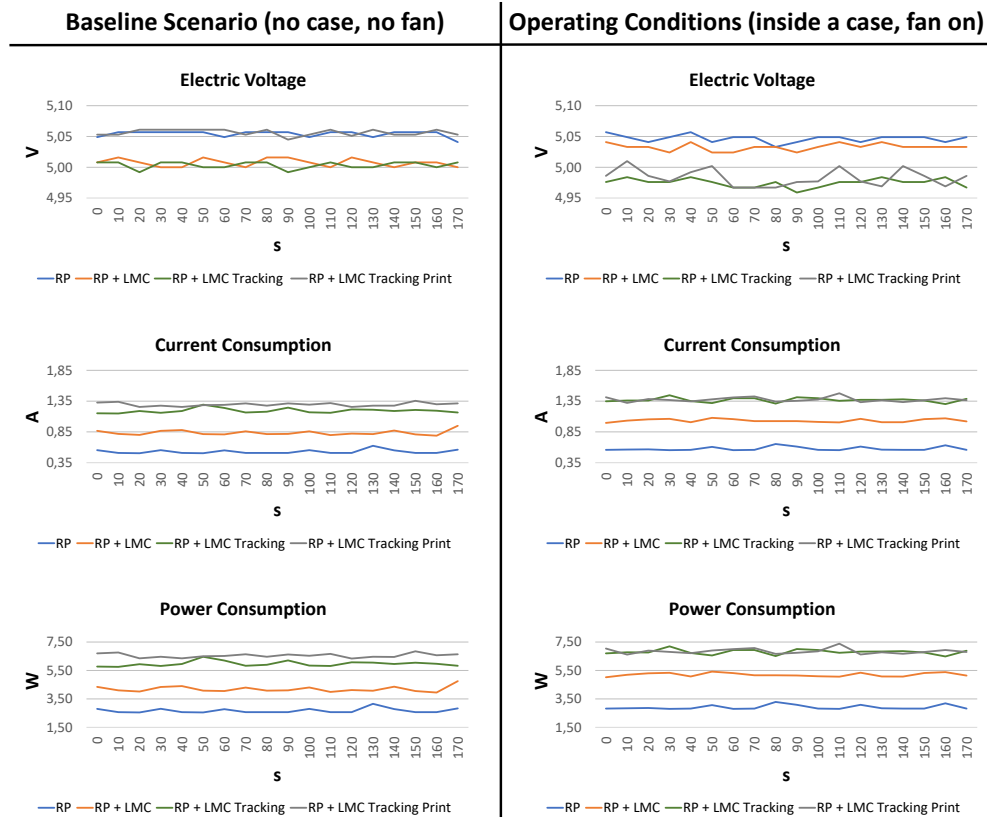


Figure 5.6: Voltage, current, and power measurements for both considered scenarios.

### 5.3.2 CPU Usage and Temperature

For each scenario, CPU usage [%] and CPU temperature [°C] were reported, as measured using the RP's on-board sensor and the GetData script, as detailed in Section 5.2. Compared to the measurements of energy consumption, these measurements required more samples because the temperature took more time

to converge to the final value. For this reason, a script was developed to acquire values every 10 s for 890 s. Table 5.2 reports the results for both the RP baseline (when the RP was not inside the case) and the proposed configuration (when the RP was inside the case with the fan).

Starting from the baseline scenario, the CPU usage was minimal in the RP configuration (the GUI was disabled). When the LeapD was running (RP + LMC), CPU usage was almost 11%, and it reached 74.02% when data were requested from an application (RP + LMC + Data). When printing was added (RP + LMC + Data + Print), the average CPU usage increased slightly. Regarding the temperature, in the RP configuration, it was 53.66 °C and increased for each task: 60.36 °C, 74.75 °C, and 75.49 °C. This demonstrates that without any active cooling, the temperatures reached were too high. In fact, during the most intense task (RP + LMC + Data + Print), there was a peak temperature of 80.34 °C. As reported in the specifications, the RP began to throttle the processor when the temperature reached 80 °C, increasing when it reached the maximum temperature of 85 °C [131]. This justifies the adoption of a fan installed on a case. In fact, in the proposed configuration, even though the CPU usage was almost identical to the baseline scenario, the drop in temperature was evident across all tasks. Even in the most intense task, the maximum temperature registered was 58.43 °C, which was almost 22 °C lower than in the same task without the fan, although it was now enclosed in a case. Additionally, in both scenarios, passive adhesive aluminum heatsinks were used, as shown in Figure 5.7.

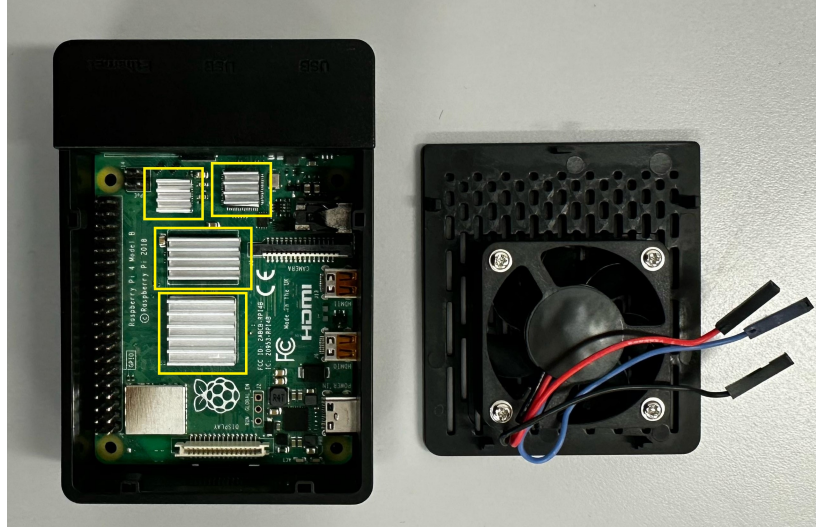
|    | Baseline Scenario |            |                 |                         | Operating Conditions (fan on) |            |                 |                         |
|----|-------------------|------------|-----------------|-------------------------|-------------------------------|------------|-----------------|-------------------------|
|    | RP                | RP + LMC   | RP + LMC + Data | RP + LMC + Data + Print | RP                            | RP + LMC   | RP + LMC + Data | RP + LMC + Data + Print |
| %  | Avg: 0.14         | Avg: 10.63 | Avg: 74.02      | Avg: 74.57              | Avg: 0.10                     | Avg: 10.67 | Avg: 73.95      | Avg: 74.09              |
|    | Min: 0.00         | Min: 9.80  | Min: 72.80      | Min: 68.10              | Min: 0.00                     | Min: 9.30  | Min: 72.70      | Min: 73.00              |
|    | Max: 1.30         | Max: 11.60 | Max: 75.10      | Max: 79.50              | Max: 0.70                     | Max: 18.40 | Max: 74.90      | Max: 75.20              |
| °C | Avg: 53.66        | Avg: 60.36 | Avg: 74.75      | Avg: 75.49              | Avg: 40.05                    | Avg: 43.77 | Avg: 56.06      | Avg: 56.04              |
|    | Min: 52.58        | Min: 56.97 | Min: 52.58      | Min: 55.50              | Min: 38.95                    | Min: 42.35 | Min: 48.69      | Min: 47.23              |
|    | Max: 55.02        | Max: 62.81 | Max: 79.85      | Max: 80.34              | Max: 41.38                    | Max: 45.28 | Max: 58.27      | Max: 58.43              |

**Table 5.2:** CPU usage (in %) and temperature (in °C), for both the considered scenarios.

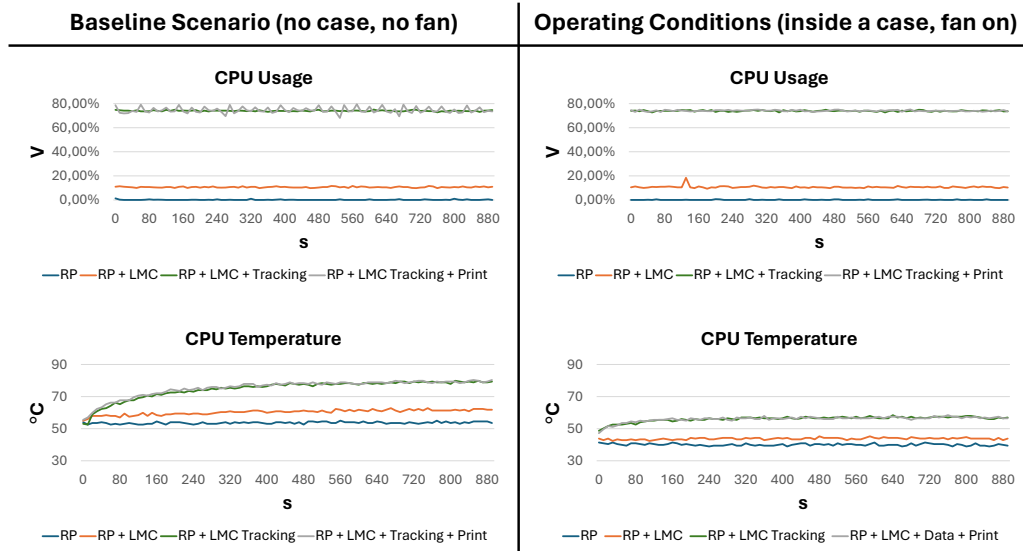
The experiments suggest that passive cooling was inadequate when computational demands were high, particularly because it could require a considerable amount of time.

Finally, the plots for the baseline scenario and the operating configuration are shown in Figure 5.8. Although CPU usage was stable, the CPU temperature significantly increased, especially in the baseline scenario. In fact, in the RP + LMC + Tracking + Print configuration, the temperature exceeded 70 °C after only 160 s and was unstable, showing an increasing trend. In contrast, in the same scenario for the final RP assembly, after the same amount of time, the temperature had already converged to a stable value at around 58 °C.





**Figure 5.7:** The aluminum heatsinks used during the measurements for the baseline and the proposed assembly are highlighted with yellow boxes. The figure also shows the RP enclosed in the case and the fan used for cooling in the proposed assembly.



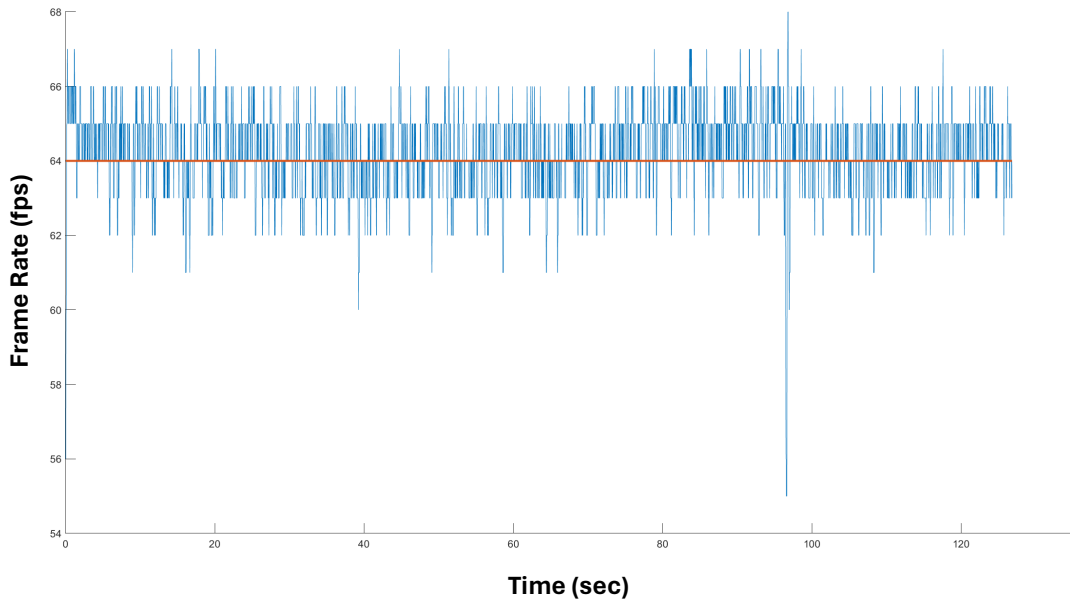
**Figure 5.8:** CPU usage (in %) and temperature (in °C) for both scenarios.

The last significant piece of information gathered from using the proposed device in both scenarios was the power bank's battery duration. For the baseline scenario, the system was left running in RP mode until shutdown occurred due to battery depletion, resulting in an observed duration of approximately 12 hours. The same procedure was repeated under operating conditions (fan on), in RP + LMC + Data + Print mode, resulting in a duration of approximately 5 hours. The results indicate that for the worst-case scenario, the duration of the battery is completely compatible with most human activities the system is involved in tracking. It is worth noting that, for the baseline scenario, measurements in other modalities were not performed, both to assess the maximum battery duration (achievable with only the RP active) and to avoid potential damage to the RP caused by overheating.

### 5.3.3 Hand Tracking

In the previous sections, the RP demonstrated its capability of running an LMC while remaining well within acceptable energy consumption and CPU temperature limits, without saturating the CPU, when an active

system is used (fan). As a final test, data from both forearms were acquired under real operating conditions (RP + LMC + Data), and the frame rate (fps) was evaluated, considering that a minimum of 25–35 fps is required for smooth real-time interaction [132, 133]. The proposed system was tested while the forearms performed various tasks: simple movements in the air (standing position), typing on a keyboard (seated position), and interacting with a manual tool (walking around the room while performing tool tapping). For the latter task, the illumination conditions varied due to the fact the person was moving near a window, then in the center of the room, and finally, near the illumination of a lamp. This is an important aspect because the LMC is based on the use of infrared light to operate: external sources of infrared light could negatively affect tracking precision. Furthermore, in this situation, the software activates specific countermeasures to compensate for this, requiring considerable CPU overload [134]. As shown in Figure 5.9, the proposed system operated at an average frame rate of 64 fps, with values ranging between 62 fps and 66 fps. Only in one isolated case did it drop to 55 fps, corresponding to a fast passage close to the window. Ultimately, the system always operated above 30 fps, thus ensuring real-time tracking. Figure 5.10 shows the spatial coordinates (X, Y, Z) of the fingertips of both hands throughout the data acquisition, corresponding to the movements analyzed for fps in Figure 5.9. To enhance the understanding of the plots, vertical lines are drawn corresponding to specific hand poses, as shown in plots (a) to (h). The plots were created using Matlab 2023a, with blue dots used to mark the joints of the hands acquired by the LMC. For the reported poses, movements were smooth except when the person was interacting with the manual tool (Figure 5.10f). In this case, the left hand was almost completely obscured by the tool, and the LMC lost track at some point (Figure 5.10g). Regarding the right hand, tracking was more stable because most of the time, the person was interacting with the index finger, and consequently, the hand was not hidden behind the tool, making it visible to the LMC. However, self-occlusions also occurred, where some fingers obstructed the view of others. In these cases, hand-tracking precision decreased significantly, and jumps occurred when the LMC recovered the right tip position.



**Figure 5.9:** The frame rate of the LMC was read from the internal data structure and reported. The original averaged values also contained decimals, which have been eliminated for convenience.



## 5.4 Conclusion

The present research proposed a portable hand-tracking system based on a head-mounted LMC connected to an RP. A 10,000 mAh power bank powers the system. Several measurements were conducted. Regarding power consumption, the system is capable of running for 5 h in tracking mode. CPU usage is not saturated, leaving space to run other applications. The temperature is very well controlled using heatsinks and a fan, even under the most intense load, which, in this case, occurs when the LMC is tracking and external software is requesting the frames and saving the data. Regarding frame rate, the system is capable of running at an average of about 60 fps without any significant drops. Tracking precision is good and stable, as demonstrated by the conducted test under changing operative conditions, although overall precision could be affected.

Considering these results, it can be confidently asserted that the proposed system holds significant promise for various precise hand-tracking applications. Head-mounted hand tracking emerges as a particularly effective design choice due to its inherent advantages: users can freely engage with their surroundings without being hindered by the position of the hand-tracking sensor, which consistently aligns with their field of view. This freedom allows users to grasp real-world objects, fostering a natural and immersive interaction experience. Additionally, the system's portability has been substantiated by its ability to seamlessly integrate the computing power of the RP with the demands of the LMC, alongside a battery life that accommodates all major tasks.

Future work will be in the direction of definitive system engineering. First of all, a button to start and stop hand tracking will be added. This will enable the possibility of pausing tracking when it is not necessary, allowing significant energy savings and, consequently, increasing usage time.

To further increase the efficiency of the system, the velocity of the fan will be controlled, taking into account the temperature of the CPU. To minimize fan usage, heatsinks with different shapes and materials with better thermal conductivity will be tested.

Another aspect is to implement data streaming using Wi-Fi. In this way, it will be possible to run intensive computations like deep learning on a dedicated workstation and retrieve the results, significantly extending the range of potential applications [135]. After these improvements, the system will be extensively tested in real-world applications involving gesture recognition and action prediction in both indoor and outdoor scenarios.

Hand-tracking technology holds promise across various practical applications, particularly in healthcare. It can enhance hand rehabilitation by providing detailed quantitative analysis and feedback on hand movements [80, 135–137]. Additionally, it can contribute to research endeavors focused on motor imagery development when hand motor execution data could be useful, especially when integrated with tools like electroencephalography [138]. Beyond healthcare, hand tracking has implications in gaming [139], human–computer interaction [140], and security/authentication systems [141].

## Chapter 6

# BNNAction-Net: Binary Neural Network on Hands Gesture Recognitions

---

The results of this chapter are published in:

F. Fontana, A. Di Matteo, L. Cinque, G. Placidi, and M. R. Marini, “BNNAction-Net: Binary Neural Network on Hands Gesture Recognition” in *ACM SIGGRAPH 2024 Posters*, pp. 1–2, 2024.

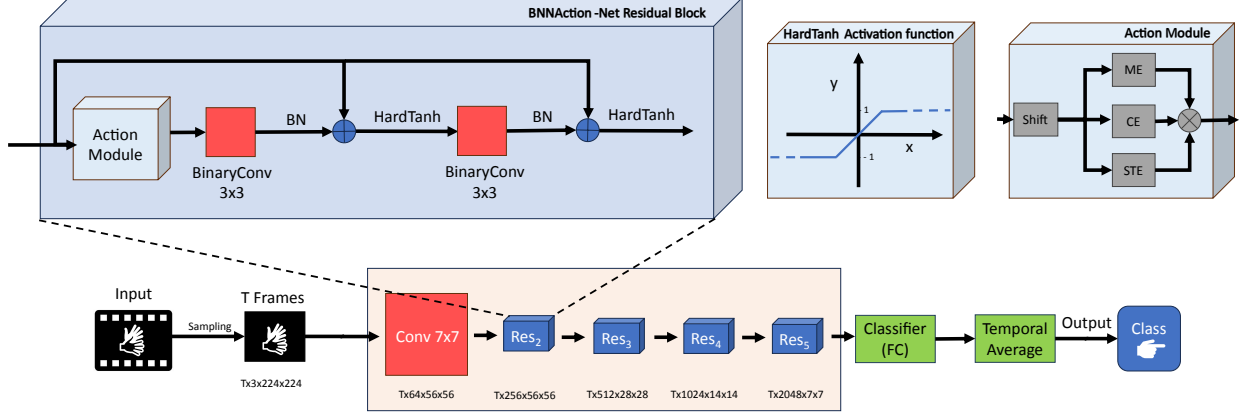
DOI: /10.1145/3641234.3671047

---

Hand Gesture Recognition (HGR) plays a crucial role in supporting motor rehabilitation and enhancing patient autonomy during Activities of Daily Living (ADL). In particular, its integration into immersive environments such as Virtual Reality (VR) enables engaging and natural interaction paradigms that can stimulate motor recovery. Within the broader domain of hand rehabilitation, gesture recognition represents a complementary aspect to movement tracking and assessment, enabling the system not only to capture motion but also to interpret intentional actions performed by the patient. This functional understanding of movement is essential for adaptive and feedback-driven therapy sessions, especially in immersive and interactive scenarios. However, the deployment of HGR systems on untethered Head-Mounted Displays (HMDs), increasingly used in portable rehabilitation settings, poses significant challenges due to limited computational and energy resources. These constraints call for the design of lightweight and efficient neural architectures capable of real-time gesture recognition while minimizing power consumption. This chapter introduces BNNAction-Net, a novel gesture recognition architecture designed specifically for resource-constrained systems. The proposed system leverages the efficiency of Binary Neural Networks (BNNs) [142], which represent both weights and activations using 1-bit values, thereby reducing memory usage and enabling faster inference through bitwise operations. BNNAction-Net integrates a binarized ResNet34 backbone into the ActionNet framework [143], originally developed for spatiotemporal action recognition, and further enhances its performance through knowledge distillation from a larger teacher model. The following sections present the BNNAction-Net architecture in detail, together with its training procedure, knowledge distillation approach, and optimization strategies. The chapter closes by summarizing the main contributions and discussing potential avenues for future work.

### 6.1 Method

The proposed method introduces innovation in hand gesture recognition (HGR) by adapting the ActionNet [143] architecture tailored to the EgoGestures dataset [144]. The EgoGesture dataset is a collection of 24161



**Figure 6.1:** Pipeline and architecture of the proposed system.

hand gesture samples grouped into 83 different classes and captured from an egocentric perspective using an RGB-D Camera in indoor and outdoor environments and taken from 50 subjects.

The key modification involves replacing the conventional backbone with a BNN, known for its efficiency. In detail, it consists of a binarized ResNet34 [145]. Model performance is further enhanced through the application of Knowledge Distillation (KD), using a pre-trained ActionNet model based on ResNet50 as the teacher.

### 6.1.1 Knowledge Distillation

KD transfers rich feature understanding from a complex ActionNet model, ResNet50 (teacher), to the modified BNNActionNet model (student), as described in [146]. Based on ResNet50, the teacher model was trained on the EgoGestures dataset. The final loss is the Mean Squared Error loss calculated between the logits of the student's network and those of the teacher's network.

$$\text{MSE Loss} = \frac{1}{N} \sum_{i=1}^N (\mathbf{T}_i - \mathbf{S}_i)^2$$

where  $N$  is the total number of instances, and  $\mathbf{T}_i$  and  $\mathbf{S}_i$  are the logits from the teacher and student for the  $i^{th}$  instance, respectively. The squared differences are summed and averaged. Results in Table 6.2 show that the KD loss is more effective when used alone, eliminating the cross-entropy loss from the model.

### 6.1.2 Implementation Details

The entire model was implemented using PyTorch<sup>1</sup>. Training was performed on the EgoGestures dataset, with the model being evaluated using accuracy. The model used as a backbone is the binarized ResNet34 [145] pre-trained on ImageNet. As illustrated in Fig. 6.1, this network design results in significant information loss. To mitigate this issue, an additional skip connections was introduced to enhance the information flow. The optimizer used is Hero Lion [147] with a learning rate of 0.01, batch size of 64, and weight decay of 0.

### 6.1.3 Evaluation

BNNAction-Net, was rigorously evaluated using the EgoGestures dataset. The official test/train split was used. The outcomes, presented in Table 6.1, demonstrate a notable enhancement in HGR, confirming the efficacy of the architectural adjustments. These modifications enabled the network to operate with ten times

<sup>1</sup><https://pytorch.org/>

fewer operations while incurring only a minimal decrease in accuracy. Additionally, Table 6.2 details the impact of each modification on the final accuracy.

**Table 6.1:** Comparison with other state-of-the-art methods, where Ops represents the number of operations for each gesture recognition inference on a sample of size  $8 \times 3 \times 224 \times 224$ .

| Method               | Ops         | Param         | Acc % |
|----------------------|-------------|---------------|-------|
| TSN[148]             | 33G         | 23.68M        | 83.1  |
| TSM[149]             | 33G         | 23.68M        | 92.1  |
| ACTION-Net[143]      | 34.7G       | 28.08M        | 94.2  |
| BNNAction-Net (Ours) | <b>3.2G</b> | <b>18.69M</b> | 92.3  |
|                      | -90.77%     | -33.44%       | -1.86 |

**Table 6.2:** Experiments investigating the effects of varying network components, loss functions, and optimizers on accuracy.

| Variation |                                                       | Ops   | Param  | Acc % |
|-----------|-------------------------------------------------------|-------|--------|-------|
| Baseline  | ActionNet with backbone<br>XNOR-Net on ResNet34 [142] | 3.21G | 18.62M | 75.6  |
|           | + Learnable Bias and<br>weights clipping [145]        | 3.29G | 18.69M | 89.3  |
|           | + Knowledge Distillation                              | 3.29G | 18.69M | 90.4  |
|           | - Cross Entropy Loss                                  | 3.29G | 18.69M | 91.5  |
|           | + Hero Lion Optimizer [147]                           | 3.29G | 18.69M | 92.3  |

## 6.2 Conclusion

An efficient system for hand gesture recognition (HGR) from an ego-view perspective was developed and evaluated using Binary Neural Networks (BNNs). The findings indicate that BNNs substantially reduce computational demands while incurring only minimal loss in accuracy. The system’s low power consumption and reduced processing needs make it ideal for standalone head-mounted displays (HMDs).

Future work will refine the system’s accuracy while maintaining its efficiency and explore its integration with real HMDs, such as using Huawei Bolt<sup>2</sup> for Android on Meta Quest. Finally, expanding the gesture library and improving user feedback mechanisms will enhance usability and user experience in practical applications.

<sup>2</sup><https://github.com/huawei-noah/bolt>

## Part III

# Physical Rehabilitation Data



## Chapter 7

# A DICOM-based Standard for Quantitative Physical Rehabilitation

---

The results of this chapter are published in:

A. Di Matteo, D. Lozzi, F. Mignosi, M. Polsinelli, G. Placidi, “A DICOM-based Standard for Quantitative Physical Rehabilitation”, in *Computational and Structural Biotechnology* 2025.

DOI: 10.1016/j.csbj.2025.01.012

---

In this part, and in particular in this chapter, the problem of a lack of a standard to handle data from quantitative physical rehabilitation data is addressed. While the previous chapters focused on developing systems for motion tracking, gesture recognition, and patient–therapist interaction, this chapter shifts the perspective toward a fundamental complementary aspect: the standardization of quantitative rehabilitation data. As the amount of information generated by sensor-based systems continues to grow, ensuring data interoperability and reusability becomes essential for advancing both research and clinical practice.

In recent years, quantitative physical rehabilitation (QPR) has gained increasing relevance as a complementary approach to traditional, subjective methods of patient assessment. Thanks to the widespread adoption of sensor-based tracking systems, it is now possible to monitor patient performance through precise measurements of movement, position, and force. However, despite these technological advances, the field still lacks a unified standard for storing and sharing such heterogeneous data, which limits interoperability between devices, institutions, and software platforms.

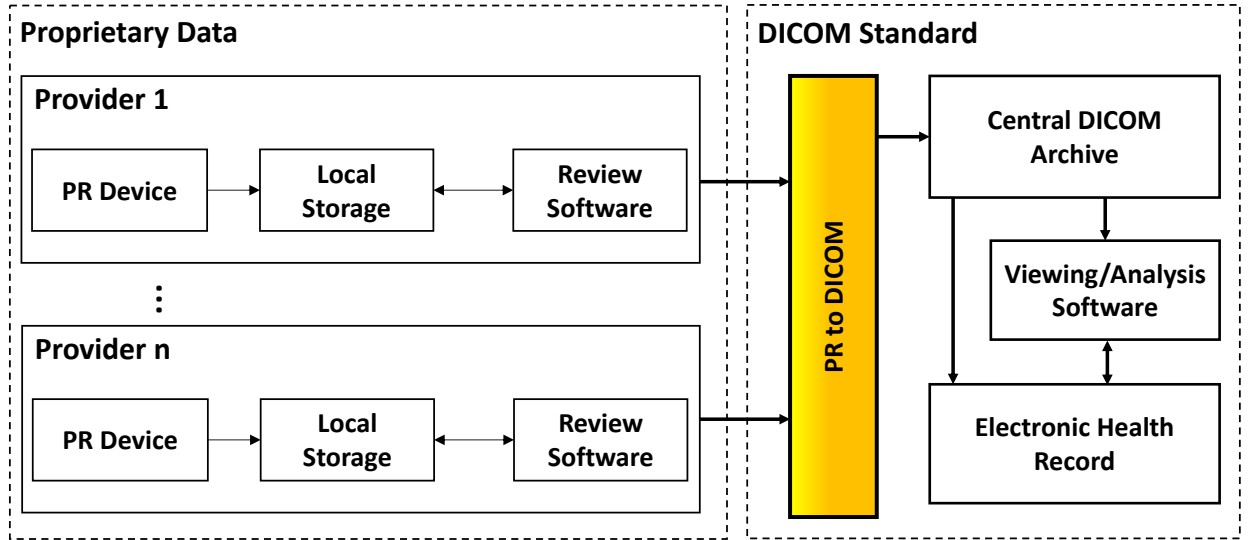
To address this limitation, the chapter explores the adoption of the Digital Imaging and Communications in Medicine (DICOM) standard as a foundation for representing, storing, and exchanging QPR data. Initially developed for medical imaging, DICOM has evolved into a versatile, extensible framework that supports a wide range of biomedical data types, including time series from physiological signals such as ECG and EEG. Its robustness, flexibility, and widespread adoption across healthcare systems make it a promising candidate for supporting QPR workflows. The choice of DICOM as the foundation for this standard is motivated by a simple clinical reality: to be useful, medical data must be visible and accessible across different systems. While modern sensors generate a wealth of information regarding a patient’s movement, these "rehabilitation time-series" are often stored in proprietary "silos" that cannot be easily shared or compared. Since DICOM is already the universal language of hospitals—with decades of proven reliability in handling complex signals like heart rhythms (ECG)—it is the most logical ecosystem for movement data. By adopting DICOM, we ensure that a patient’s recovery metrics are treated with the same clinical rigor as an X-ray or an MRI, allowing them to be integrated directly into official medical records and reviewed by any specialist, regardless of the

tracking device used. The objective of this chapter is to define a new DICOM Information Object Definition (IOD) explicitly tailored for QPR, referred to as the PR-IOD. This model is designed to manage time-series data generated by tracking systems and assistive rehabilitation technologies, enabling full interoperability and standardized storage of rehabilitation sessions. A concrete application scenario, involving data from a hand-tracking system known as the Virtual Glove, is provided to illustrate the capabilities of the proposed PR-IOD and to demonstrate how QPR data can be seamlessly integrated into the DICOM ecosystem.

## 7.1 Background

Qualitative rehabilitation, though important for direct contact with the patient, is prone to subjectivity and distorted evaluations, thus potentially leading to information loss and errors [29]. This is in contrast with the concept of personalized precision medicine [30].

However, the use of quantitative instruments in PR requires that data be standardized to improve interoperability and data sharing. Without a standard, each provider uses its data format, leading to issues



**Figure 7.1:** PR proprietary data (left) are merged, through a PR to DICOM translator, into a shared Dicom-based data format (right): in this way, data can be compared and shared easier, thus improving interoperability. This figure has been adapted from Fig.1 in [150], where a similar architecture was implemented for EEG data.

in data communication between devices and healthcare professionals. This greatly limits the possibility of creating a patient history, making it difficult to compare data from different devices, perform a temporal evaluation of the therapy or quantitative comparisons between data of different patients, create and use common evaluation scales for therapists, calculate statistics, etc. The situation is described in the left side of Fig.7.1.

PR data contains all the complexity of the medical data and in particular, the heterogeneity of the information makes it challenging to create systems that can access and integrate different databases for patient information maintenance. Over the past two decades, efforts have been made to establish medical standards for data communication between different components, including network protocols, data structures, and shared definitions [151–155]. This process has been partially performed in the context of Electronic Medical Record (EMR) and Electronic Health Record (EHR) [156, 157], which are tools for supporting physician decision-making, as well as for improving patient medication, while preserving and improving safety. However, the adoption of EMR and EHR by medical institutions is not obvious, mainly due to legal and privacy concerns and budget limitations, thus often leaving the standardization process incomplete [158–160].

Digital Imaging and COmmunications in Medicine (DICOM) was born as a voluntary standard defined by imaging companies and stakeholders, providing a framework that may be utilized independently or in conjunction with EMR and EHR. is widely recognized as a fundamental standard for the representation and sharing of medical data, and it is integrated with Picture Archiving and Communication System (PACS) [161]. Moreover, since its birth, has become the “*de facto*” standard in healthcare due to its versatility, resulting in several significant advances also in other fields of medicine [150, 162, 163].

For this reason, over the years has significantly expanded to include a variety of supplementary medical tools [164], such as those for managing time series (ECG, EEG, EMG) and others [150, 163, 165–170], although is primarily recognized as the standard for Medical Imaging. In its most recent release (3.0), DICOM has attained the status of a “*de iure*” standard that is acknowledged by the International Organization for Standardization (ISO).

QPR, whose data are collected by assistive numerical systems producing time series, is a medical domain where DICOM could offer a significant contribution to standardization. Despite DICOM’s apparent compatibility with PR, its potential use in PR is still underexplored. The result is to use DICOM as a standard data format also in PR, as reported on the right side of Fig.7.1.

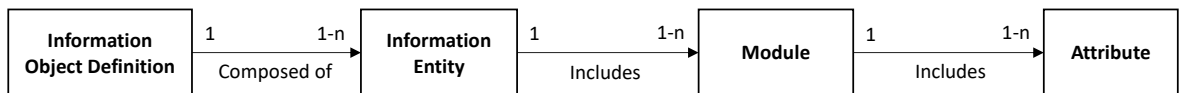
The main focus of this study is to design a new DICOM-based Information Object Definition (IOD) for PR (PR-IOD), which is an abstract template model to instantiate information about Real-World Objects, specifically designed for supporting time series deriving from tracking models of the human body or parts of it. The proposed PR-IOD has to leverage the concept of the Waveform Module (already present in DICOM) that has the potential to be used to manage time series. As a proof of concept, after the PR-IOD definition, a scenario of its application to data collected by a multi-modal hand-tracking device named Virtual Glove [54] is presented and discussed, along with its implementation.

## 7.2 The DICOM Architecture

The DICOM standard represents and connects real-world data using the Object Oriented Paradigm (OOP). Specifically, it employs the notion of **attributes** to collect and define physical world data, **methods** to delineate the procedures applied to attributes, and **classes** to integrate attributes and methods. Finally, **objects** are used to create instances of **classes** [155].

### 7.2.1 Core Information Model and Methods

Any standard has to rigorously outline how real-world data are structured and connected. In particular, for any real-world entity, such as **Patient** or **Image**, is defined as a specific contained in the IOD. Any Information Object Definition (IOD) can be Normalized if composed by a single , or Composite if consisting of a combination of IEs. [171]. The amalgamates attributes associated with the real-world entity it represents.



**Figure 7.2:** Information Object Definition Model composition.

The correlated attributes are then organized into modules, and these last are contained by the IE (Figure 7.2). Attributes are defined by a tag name, a description, and a type, which can be:

1. required;
2. required but unknown;

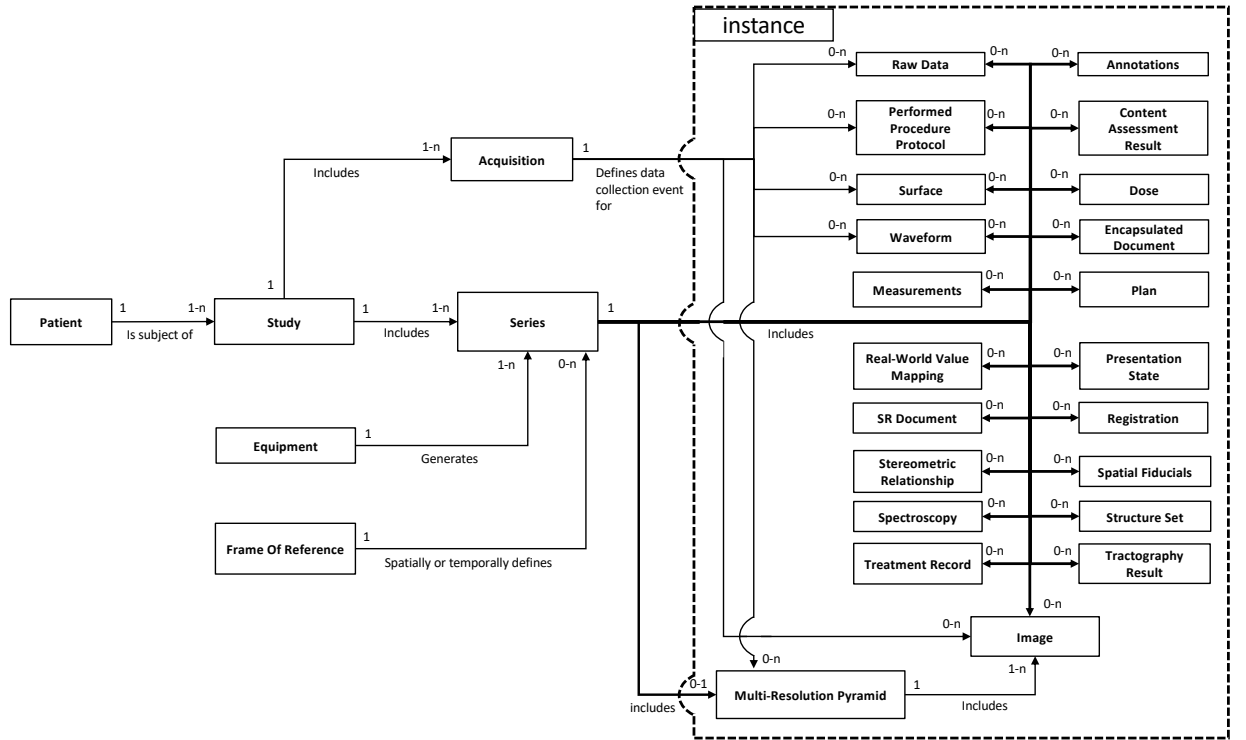
3. optional;
4. conditionally required (which means that it is required if a condition is met);
5. same as 2, but if a condition is met.

An IOD is an object-oriented abstract data model that specifies information about a class of real-world objects with similar features and is used to represent the relationship between the IEs it contains.

An example is represented by the “General Electrocardiogram (ECG) IOD”, which specifies bio-signals collected with an ECG [172].

With this structure, DICOM provides a comprehensive framework for organizing IEs and interconnecting them through constraints (Figure 7.3), in which every Composite IOD inherits entities and relationships from the framework.

An example of a composite IOD is the "Magnetic Resonance Image" IOD, which encompasses data regarding the patient, the conducted study, and the acquired image by a magnetic resonance imaging system [173]. This IOD includes all IEs from the patient block to the series block, ultimately culminating in the inclusion of the Image IE.

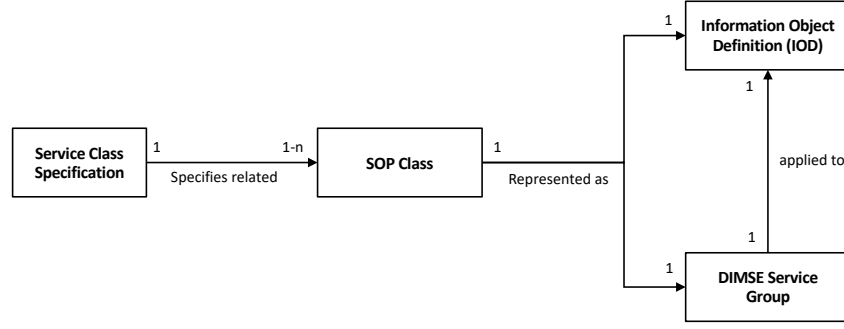


**Figure 7.3:** DICOM conceptual framework where relationships and constraints among IEs are shown [174].

In DICOM, the IOD has also associated operational services and methods called Dicom Message Service Element (DIMSE) [175]. These services can be inherited IODs specifically designed for specific applications, as well as the used structured classes.

The Class is the outcome of DICOM standard [155]. It is defined as the union of IODs and DIMSE services (Figure 7.4). This concept works as a contract since it is what is exchanged among two .

Additionally, the standard uses the client/server architecture for network communication to communicate/exchange data and services. Specifically, it designates entities that function as users as Service Class Users (SCUs), while those acting as servers are termed Service Class Providers (SCPs). Finally, at the top of the hierarchy, there is the Service Class Specification, which outlines a collection of one or more Classes



**Figure 7.4:** The Information Model delineates the interconnections among the principal components of the standard [176].

that are linked to a specific task. Examples of Service Class Specifications include storage (**Storage Service Class**), connectivity verification (**Verification Service Class**), querying and retrieval of managed instances (**Query/Retrieve Service Class**), picture printing (**Print Management Service Class**), and many more [177].

## 7.3 DICOM IOD design for Physical Rehabilitation

Before delving into the definition of the proposed Physical Rehabilitation IOD (PR-IOD), it is necessary to provide an overview of the physical rehabilitation devices and the data types they provide. After that, the structure and pipeline of Physical Rehabilitation Information Object Definition (PR-IOD) can be presented.

### 7.3.1 Equipment

As stated above, PR plays a crucial role in the recovery process. Various kinds of injuries demand different types of Physical Rehabilitation (PR) procedures [178], each involving specific devices and technologies to support QPR. The devices supporting QPR may include [179]:

- Inertial Sensors
- Pressure Sensors
- Electromyography (EMG) Sensors
- Camera-based Systems
- -based Systems

An [180] is a device that uses accelerometers, gyroscopes, magnetometers, and (sometimes) a barometer together. By mixing this multimodal information, the sensor can provide further information regarding position, orientation, velocity, acceleration, as well as any other parameter derived from them.

Pressure sensors measure the exerting pressure applied to the sensor. An example of this system is the Force Sensitive Resistor (FSR), which changes its resistive value depending on how much it is pressed[181].

The EMG systems are used to record the electrical signals of muscle activity during movements. signal recording involves the utilization of disposable concentric needle electrodes inserted inside the muscle [182].

The camera-based systems include marker placement or markerless solutions [183]. While marker-based solutions need wearing physical markers to extract kinematic data, markerless solutions need video cameras and other external devices to capture the motion of a body from which computer vision strategies and/or artificial intelligence techniques provide the extraction of kinematic data and the reconstruction of a numerical model of the body or of parts of it.

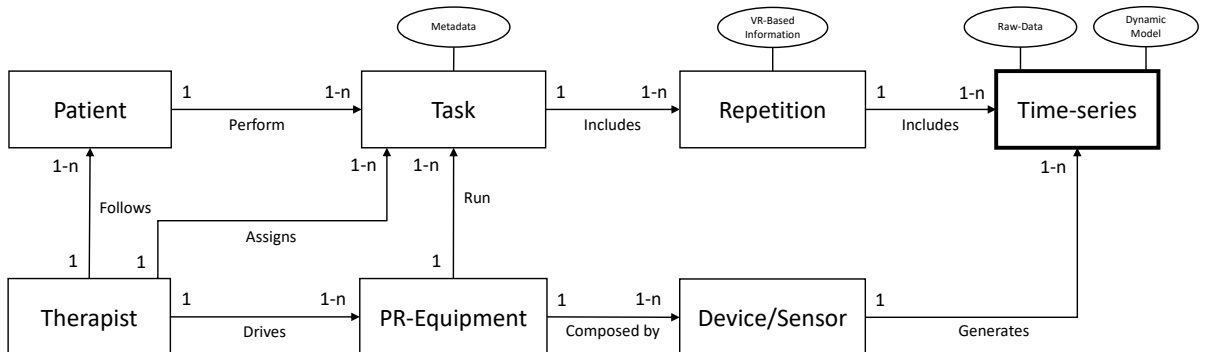
IR-based systems include the usage of commercial devices, such as the [184] and the Microsoft Kinect [185], whose behavior is very similar to the camera-based markerless solutions. In particular, the is a compact device that uses stereo vision to provide a three-dimensional representation of the movement of a human hand. It comes with two infrared cameras and three infrared LED light sources. This device represents the human hand by providing the spatial position of 25 joints assembled in a structured hand model. The Kinect is an RGB-D (Red, Green, Blue, and Depth) sensor primarily designed for real-time tracking and recognition of various parts of the human body. This system provides a diverse array of data types, which include depth data, RGB data, and IR Data.

The above systems can be used singularly as atomic tools [186, 187], or they can be mutually integrated into multiple-source systems [54, 64, 188].

These systems have the versatility to be applied and/or used to various anatomical regions, thereby facilitating monitoring and supporting the rehabilitation process. Moreover, they can be effectively integrated with for improving patient engagement [189]. Data generated by these devices predominantly takes the form of time series, providing information regarding spatial position, angles, velocity, acceleration, etc. As a final remark, all these systems have the capability of producing raw data, reconstructing dynamic numerical models, labeled and processing time series, and reports. This not only helps to create a database of the patient's activity but also provides invaluable insights into the recovery progress during rehabilitation.

### 7.3.2 Task-Based Numerical Tracking and Data Generation

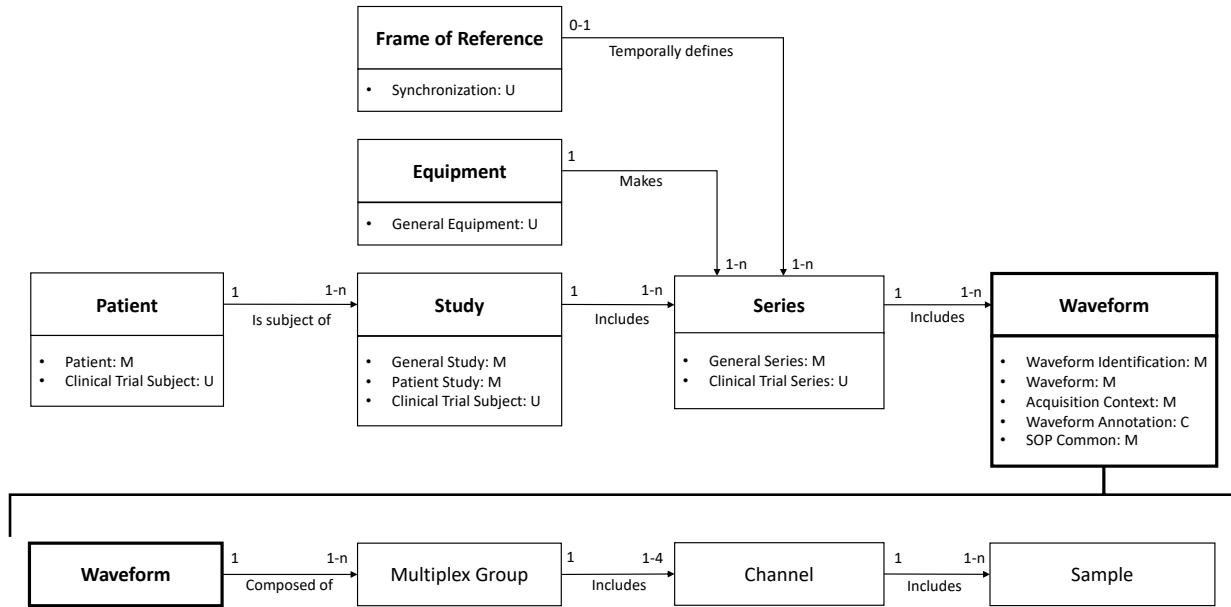
QPR is founded on the deep engagement of the patient and the (remote) interaction with the therapist. A fundamental step for numerical tracking of rehabilitation is that the patient is equipped with one of the assistive devices listed above, which allows the execution of a specifically designed rehabilitation task, accompanied by numerical data acquisition, processing, reconstruction, and storage. The therapist is responsible for the task definition and the setting of its level of difficulty, based on the patient's infirmity and residual capabilities. With the task definition, some metadata are generated including the type of task, the level of difficulty, the number of recommended repetitions, the duration of the repetitions, etc. Upon task assignment and setting, the patient proceeds to its execution by completing all the recommended repetitions, either in the presence of the therapist or at home. During the task execution, time series are generated, consisting of the raw data (temporal data directly measured by the sensors); a dynamic model of the tracked part of the body reconstructed by using the raw data; and annotations of data referred to specific time-instants of the task. VR-based information belonging to the repetition is also created, such as the score, time employed to complete the task, etc. Figure 7.5 shows the actors involved in the PR pipeline and the data generated after the execution of a rehabilitation task.



**Figure 7.5:** Actors involved in the PR Pipeline and data generation.

### 7.3.3 PR-IOD Structure

The IOD defined for PR can be directly derived from the general composite model (Figure 7.3), which is a structure used in a variety of clinical contexts to manage time series. This versatile, and redundant, structure exhibits suitability for accommodating PR data, as depicted in Figure 7.6. This approach preserves adherence to the DICOM standard while simultaneously leveraging its modules and functionalities. In what follows, a short description, for each of the involved IEs and related modules and attributes, is furnished.



**Figure 7.6:** Relationship between IEs involved in the PR-IOD Model; every is composed of modules. Every module specifies its usage, which can be Mandatory (M), User Optional (U), or Conditionally Required (C). The Waveform IE has been further divided to show its composition (structure adapted from [190]).

#### 7.3.3.1 Patient IE

The Patient block contains attributes describing its general characteristics (Patient Module). This also contains attributes in the Clinical Trial Subject Module describing the Patient as a participant in a clinical trial or research. This perfectly includes the attributes of the patient in PR.

#### 7.3.3.2 Equipment IE

The device used by the patient that generates the Series is the Equipment IE. It is important to note that the Equipment only specifies the device type employed (General Equipment Module), not the specific data format or methods utilized in producing composite data sequences or data models.

Specific equipment can create many time series, but each series is associated with one piece of equipment. Finally, this contains the attributes of the Equipment in PR.

#### 7.3.3.3 Study IE

Within this IE, the attributes delineate the specifics of a medical examination (General Study Module), whose structure is represented in Fig. 7.6. This block also includes information regarding the patient's status, residual infirmities, and specific attributes regarding the therapist and the involved clinic (Patient Study). Lastly, it contains Attributes that identify a Study in the context of a potential clinical trial or research (Clinical Trial Study). A Patient can have more than one Study, however, each Study is exclusively associated with one Patient. In PR, the Study can be viewed as the task that the patient must perform;

therefore, attributes about the type of task, the difficulty, the number of suggested repetitions, and the duration of the task are stored here. However, these metadata are not present within the General Study Module but are created as private data elements without any loss of generality (Table 7.1).

| Private Data Elements |                              |             |                      |
|-----------------------|------------------------------|-------------|----------------------|
| Name                  | Description                  | Type        | Value Representation |
| Difficulty            | Difficulty of the task       | Optional    | DS (Decimal String)  |
| Duration              | Total duration of repetition | Conditional | DS (Decimal String)  |
| Repetitions           | Number of repetitions        | Required    | DT (Date Time)       |

**Table 7.1:** The Private Data Elements pertain to the task that the patient has to carry on.

#### 7.3.3.4 Series IE

The Series is used to specify information regarding the instance Waveform (Figure 7.6). This encompasses characteristics such as the specific date and time of the series' creation, as well as details regarding the type of examination and the employed equipment (General Series Module) [171].

Additionally, this contains some attributes describing the clinical trial (Clinical Trial Series). In PR, the series are used to specify different repetitions of the task: a new series means a new execution. However, some other attributes are incorporated: the time necessary to complete the current repetition of the task, and the achieved score, when present (Table 7.2). Also, in this case, the standard does not provide these metadata which can be added using new private elements. A Study may include several Series, but a Series must be only associated with one Study.

| Private Data Elements |                     |          |                      |
|-----------------------|---------------------|----------|----------------------|
| Name                  | Description         | Type     | Value Representation |
| Final Time            | Absolute final time | Optional | DT (Date Time)       |
| Score                 | Score attained      | Optional | DS (Decimal String)  |

**Table 7.2:** Private Data Elements specific for each repetition of the task.

#### 7.3.3.5 Frame of Reference IE

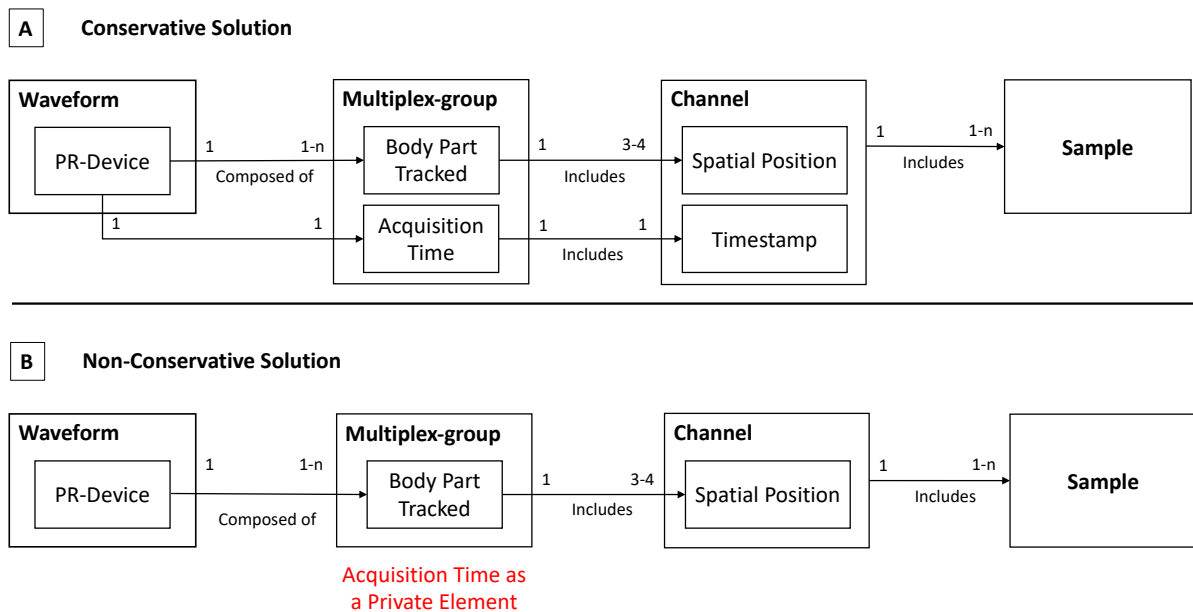
The Frame of Reference IE is used for synchronizing either different instances of a series or different sources of data collected simultaneously by different sensors. This can be associated with several Series, but each Series can be associated with just one Frame of Reference. This is depicted in Figure 7.6, and because PR devices may employ several sensors simultaneously, this IE properly meets the PR criteria.

#### 7.3.3.6 Waveform IE

The Waveform plays a pivotal role in storing multi-channel time series of data. However, some modifications are necessary to accommodate rehabilitation systems. In detail, the Waveform Module is represented as a composition of several Multiplex groups (Figure 7.6). A Multiplex group has the potential to manage several Channels. The time series in each channel is stored using a predetermined and constant number of samples. In PR, every time series produced by a sensor finds its association within a Waveform. This crucial linkage indicates that the data emanating from the sensor, about each tracked section of the body, is intricately mapped within Multiplex groups. These mappings are further defined in channels by spatial coordinates or angles, thus providing precise spatial orientation information for each section of the map. Consequently, the mapping process results in the creation of a dedicated time series for each channel. The number of Multiplex-groups included in the Waveform is unconstrained (Figure 7.6). Every Multiplexgroup can have



from 1 to 4 channels. This is because it is possible to handle not only positional data but also angles in terms of quaternions (these last require four channels). Moreover, in a Multiplex group, DICOM requires that all channels be digitized synchronously at the same sampling frequency. However, this approach does not align with rehabilitation data sources, where data from different sources can be frequently sampled at different sampling frequencies. For this reason, in PR-IOD the sampling frequency should be unconstrained. Therefore, a new data element, in which the time series stores the timestamp of a sample, is required. Unfortunately, the standard lacks this essential field for PR. To overcome this lack, two possible solutions are proposed: a conservative one and a non-conservative one (Figure 7.7). The conservative solution implies that a new Multiplex group with a single channel is created (Figure 7.7A). The non-conservative approach involves creating a new private element where the acquisition time for the entire waveform is specified (Figure 7.7B). In this last case, Table 7.3 shows the specifications of this solution.



**Figure 7.7:** PR Waveform decomposition including the time acquisition time-series for a PR-Device (that does not have a fixed sampling rate); the Conservative Solution (A) involves using an additional Multiplex-group composed of one channel; the Non-Conservative Solution (B) designates a private element for its storage.

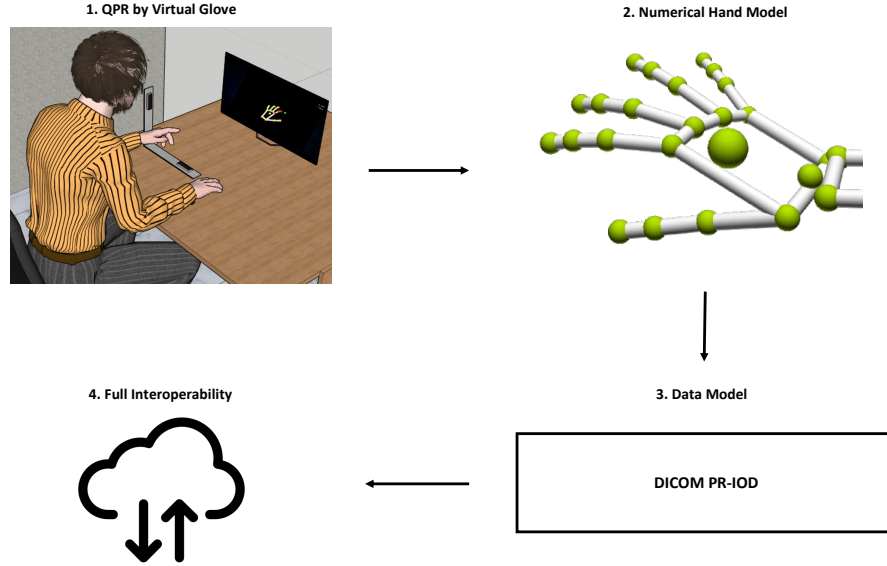
In addition, DICOM requires that every channel be composed of some mandatory attributes such as Channel Source, Channel Sensitivity, Skew, and Filter Characteristics. Moreover, this also contains some attributes defined in the Waveform Annotation Module, which allows gathering some annotations during the acquisition. It is also possible to describe the conditions present during data acquisition that can alter the measurement of PR devices (Acquisition Context Module). This IE also contains Attributes required for the correct operation and identification of the associated instance (SOP Common Module). Finally, some related Waveform attributes are used to delineate the Waveform as a separate (Waveform Identification Module)."

| Private Data Element |                                                             |             |                      |
|----------------------|-------------------------------------------------------------|-------------|----------------------|
| Name                 | Description                                                 | Type        | Value Representation |
| Acquisition Time     | Time-Series specifying the acquisition time for each sample | Conditional | OW (Other Word)      |

**Table 7.3:** New Private Data Element in the Waveform Module pertains to the sampling time if the system employed does not perform frequency acquisition at a fixed time.

## 7.4 PR-IOD for the QPR of the Hand: an example

This section presents a practical example of how data collected from the Virtual Glove (VG) [54, 67] can be seamlessly integrated into the DICOM format using the PR-IOD. The acquisition–hand model–DICOM pipeline is schematically illustrated in Figure 7.8.



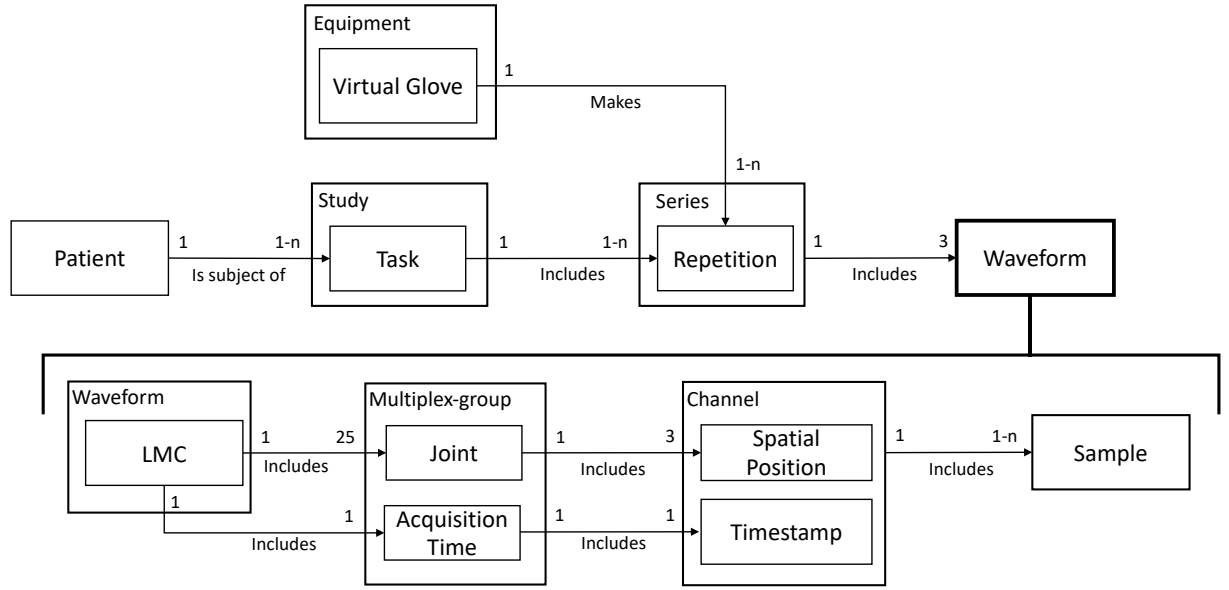
**Figure 7.8:** Quantitative physical rehabilitation of the hand: from data acquisition to PR-IOD definition, passing through the hand model construction.

The VG is a touchless hand-tracking system consisting of two mounted within an aluminum base in an orthogonal arrangement. The LMC-based solution of the VG serves to enhance the precision and efficiency of hand-tracking in the rehabilitation process. The VG is a component of a complete framework for hand telerehabilitation [64, 95] that also involves active collaboration with therapists. Within this framework, the therapist can assign specific rehabilitation tasks to the patient, set execution parameters, and conduct quantitative analyses from the resulting collected data. The data managed within the rehabilitation framework includes the following categories:

- **Patient Information, including Anamnestic Data:** This provides essential patient details and comprehensive information about the injury, residual capabilities, suggested treatment, and the required rehabilitation.
- **Therapist Metadata:** This category covers metadata associated with the therapist, including qualifications and specialties.
- **Task Metadata:** It includes details about the assigned rehabilitation task, such as the type of exercise, duration, level of difficulty, instructions, parameters, etc.
- **VG Metadata:** It includes general information about the VG device, such as the serial number, the producer, precision, etc.
- **VG-Related Task Data:** It includes all the time series generated by the VG during the executions of rehabilitation tasks.

The structured details of the data in PR-IOD for VG are reported in Figure 7.9.

The Patient’s information is conveyed within the Patient IE, specifically within the Patient Module. Metadata pertaining to anamnestic status is in the Patient Study Module, which is included in the Study



**Figure 7.9:** Overall structure of data in VG-Patient-Task within the PR-IOD: the used IEs are reported (as an example, the Frame of reference IE is not reported because not used in the VG-Patient-Task).

IE. Information regarding the Therapist can be found in the General Study Module of the Study IE. The therapist-assigned Task metadata is contained within the Study IE, which is properly located in the General Study Module. However, using this approach, specific task metadata (such as task difficulty, time duration, and others) can only be included using the non-conservative solution through private elements. Additionally, data regarding the involved therapist are stored within the Study IE, specifically in the General Study Module. Information related to the structure of the PR device, in this case, the VG, is stored within the Equipment IE. For each task, data regarding task repetitions are also available and contained in the form of private elements within the Series IE. For every repetition of the task, some metadata are extrapolated (such as final score attained, total time of execution, etc.). These data are included in the Series IE as private elements. Thus, a new Series IE means a new repetition. Within the Waveform IE, there are every time series produced by an LMC sensor during a repetition. A new Waveform IE is created for each type of LMC data.

Data from the VG can be of three types: from the horizontal LMC, the vertical LMC, and the fusion of the aforementioned. However, data from each single LMC represents raw data from which the usable output of VG is obtained through the fusion of data from both LMCs [54]. As already stated, the LMC reconstructs a numerical model of the hand composed of the numerical space of 25 joints. Therefore, the separation of each joint is made by the use of a Multiplex group, and each spatial coordinate is mapped into a channel. Finally, the time series is specified through the samples. It is important to note that, since LMC does not have a fixed frequency rate, a timestamp for each frame is specified by a specific single channel Multiplex-group. The number of samples for each Multiplex group and each channel is the same. In the depicted picture (Figure 7.9) the Frame of Reference is not included. This is because the driver of the LMC can handle it, and consequently, the two share the same temporal sequence.

#### 7.4.1 Proof of Concept

To demonstrate the effectiveness of PR-IOD, data from a repetition of a specific VR-based hand movement exercise were used, with the virtual environment implemented using Unity [94], and data collected by the

VG. The exercise, named Keyboard, involved the patient touching a keyboard with a designated joint and repeating this task for different joints at different velocities [95]. This data was collected and initially exported into the following CSV files:

- Patient File: Contains patient-specific data
- Exercise File: Contains details regarding the performed exercise, including metadata about its parameters, final score, and duration
- Therapist File: Contains data regarding the therapist who is involved in the patient's care
- Time Series Data File: Contains the spatial-temporal trajectories of the hand joints, for the whole duration of the exercise (the stick model of the hand).

A Translator program was implemented to convert the acquired data into DICOM format using the PyDICOM library [191], a widely used tool for creating and managing DICOM files. Given the nature of the study, which involved acquiring time series data for the 25 joints of the hand model, each consisting of three channels (x, y, and z), the Waveform Module was employed. Specifically, the Translator was configured to handle a DICOM file with 25 multiplex groups, each containing three channels, by adapting an existing program originally designed for 12-lead ECG studies [191]. This adaptation required only increasing the number of leads and introducing three channels per lead. To address the variable frame rate of the LMC, the conservative approach illustrated in Figure 7.7A was adopted. Additional information regarding the patient, therapist, and exercise was incorporated into specifically defined DICOM fields, most of which are natively supported by the standard. The resulting file was saved with the .dcm extension. A DICOM Viewer was also developed, implemented as a PyQt-based application [192], allowing the opening of the generated DICOM file, visualization of its contents, data analysis, and comparison. Example screenshots of the Viewer interface are shown in Figure 7.10. As can be observed, the Viewer is capable of displaying, for each selected joint, the original data and spatial trajectories and calculating and displaying velocity and acceleration. Besides, the Viewer displays the temporal movements (at true, slow, or accelerated speed) of the hand model as a whole or for some selected finger, using an interactive scatter plot. The resulting files, both containing the developed software (DICOM Translator + DICOM Viewer) and the DICOM data (example.dcm) are accessible on GitHub [193].

## 7.5 Conclusion

This chapter proposed a solution to the need for a standardized data format in Quantitative Patient Rehabilitation (QPR), based on the DICOM standard. In the proposed PR-IOD, time series data are stored using the Waveform Information Entity (IE), already available within DICOM, with the goal of preserving and extending the standard structure by introducing features specific to QPR. In this way, full compatibility with DICOM is ensured, along with all its advantages in terms of communication and interoperability. A proof of concept of the proposed model was presented using the Virtual Glove (VG) assistive system for hand-related QPR. Our primary goal was to showcase the effectiveness of utilizing DICOM format for PR data without necessitating significant modifications. The gained DICOM compatibility of PR allows a promising avenue for integrating PR data into the broader spectrum of other medical data, thereby allowing multimodality, enhancing interoperability, and the inclusion of PR data in broader clinical and research settings, in a vendor-independent data access modality. Moreover, the integration of with PR data and time series can simplify assistance for patients, and more importantly, this could allow the conjunction of multiple studies in a unique format.

This study aimed to emphasize the principle that it is generally more advantageous to utilize well-established standards and formats rather than develop novel ones. While this study provided the initial impetus, the full implementation of the complete PR-IOD should be carried out by a dedicated standardization body and is expected to extend well beyond the present effort. Once this novel IOD is established

within the official DICOM standard framework, ensuring compliance with all requisite specifications, the subsequent phase will involve the development of software aiming to assess the validity of DICOM files with PR-IOD [194, 195].



**Figure 7.10:** Viewer Interface for feature calculation and display: (A) Main Window for navigation through data and functions; (B) DICOM Field Window to show all fields, including TAG, VR, Name, and Value; (C) Trajectory feature components of a selected finger joint; (D) Velocity feature components calculated from the trajectory of the selected finger joint; (E) Acceleration feature components computed from the corresponding velocity; (F) Interactive Plot Window showing dynamic scatter plot animations of the hand as a whole or a specific finger.

## Part IV

# Assessment through MRI

## Chapter 8

# Deep Learning and Multi-Modal MRI for the Segmentation of Sub-Acute and Chronic Stroke Lesions

---

The results of this chapter are published in::

A. Di Matteo, Y. Mahé, S. Leplaideur, I. Bonan, E. Bannier, F. Galassi, “Deep Learning and Multi-Modal MRI for the Segmentation of Sub-Acute and Chronic Stroke Lesions”, in *Pattern Recognition Letters* 2025. DOI: <https://doi.org/10.1016/j.patrec.2025.11.017>

This chapter is derived from the Erasmus/Visiting Ph.D. experience at the Empenn Lab (Unité U1228, INSERM, Inria, Université de Rennes I, France) under the supervision of Prof. Francesca Galassi.

---

Accurate quantification of brain lesions during the **sub-acute and chronic phases** of stroke is crucial for understanding the extent of neurological damage, effectively monitoring patient recovery, and tailoring rehabilitation strategies. While manual segmentation by experts remains the **gold standard**, this process is inherently **labor-intensive**, susceptible to **inter-rater variability**, and impractical for large-scale clinical or research applications.

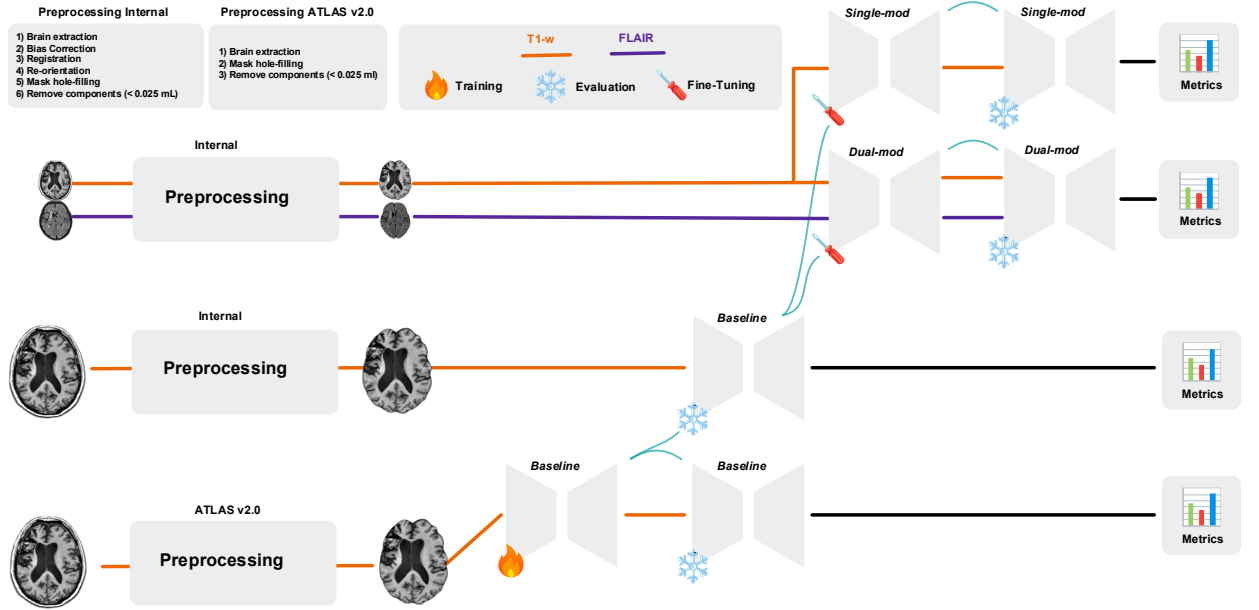
**Deep learning** methods, especially CNN, have demonstrated considerable potential in automating this task. However, much of the research focus has been on **acute stroke**, leaving the post-acute stages relatively underexplored. This gap exists partly due to the increased **complexity and variability of chronic lesions** and the limited availability of comprehensively **annotated datasets** for these later phases.

This chapter addresses these challenges by presenting a **deep learning framework for segmenting sub-acute and chronic stroke lesions**, utilizing the state-of-the-art **NN-UNet v2 architecture**. firstly a **single-modality model** was developed and trained on the publicly available **ATLAS v2.0 dataset** using **T1-W MRI** data. Subsequently, this model was **fine-tuned** using an internal clinical dataset that includes both **T1-w and flair**) scans. This allows for a systematic evaluation of performance differences between **single- versus dual-modality inputs** and assesses the impact of **domain adaptation** through fine-tuning.

The following sections detail the datasets (ATLAS v2.0 and the internal dual-modality cohort), the preprocessing pipeline, and the nnU-Net v2 training framework. We systematically compare the T1-only baseline, fine-tuned models, and T1+FLAIR configurations using early, mid, and late fusion, each evaluated with and without ensembling. Results include voxel- and lesion-level metrics with lesion-volume stratification



and statistical validation. The discussion highlights the benefits of fine-tuning and multimodal fusion, particularly late fusion, for improving cross-domain generalization and the detection of small and medium lesions.



Segmentation workflow overview. An end-to-end pipeline for post-stroke lesion segmentation encompassing data preprocessing, model training, and evaluation. The public ATLAS v2.0 dataset was used to train the baseline single-modality (T1-w) model. The internal dual-modality clinical cohort was then used to evaluate baseline generalization and to perform two subsequent adaptation stages: fine-tuning using T1-w input only, and multimodal learning using combined T1-w and FLAIR input with early, mid, and late fusion strategies. All configurations were further assessed with and without model ensembling. All models were implemented using the nnU-Net v2 framework.

## 8.1 Introduction

Magnetic Resonance Imaging (MRI) plays a central role in assessing brain damage after stroke. It is crucial both in the acute phase, immediately after onset, and in the sub-acute and chronic period occurring weeks or months later [196]. In the acute phase, rapid lesion identification is vital for guiding treatment and minimizing long-term disability, typically using diffusion-weighted and perfusion MRI. In later stages, structural MRI is preferred to evaluate residual damage and its impact on brain structure and function [197, 198]. T1-weighted (T1-w) images provide high-resolution anatomical detail essential for assessing structural damage, while Fluid-Attenuated Inversion Recovery (FLAIR) sequences, though not always acquired in clinical routine, can provide complementary information by highlighting lesion-related hyperintensities and subtle tissue changes that may not be visible on T1-w scans. Delineating stroke lesions on MRI remains challenging. Manual segmentation, while considered the reference standard, is labor-intensive, time-consuming, and prone to inter-rater variability. These limitations motivate the development of automated methods that can provide accurate and reproducible lesion delineations. Recent advances in deep learning have substantially improved medical image analysis [20, 199–201], particularly in acute stroke segmentation [8–11]. However, automated segmentation of sub-acute and chronic lesions remains considerably underexplored. Late-stage lesions are often small, heterogeneous, and poorly contrasted, and clinical imaging protocols vary widely across scanners and sites. Moreover, the scarcity of publicly available annotated datasets has limited the development and validation of robust models. Accurate segmentation of sub-acute and chronic lesions is nevertheless critical

for quantifying tissue loss, assessing functional impairment, and guiding rehabilitation strategies [202–204]. Reliable delineation enables objective tracking of lesion evolution over time, supporting prognostic assessment and evaluation of therapy outcomes. A recent review by Ahmed et al. [205] provides a comprehensive overview of automated post-stroke lesion segmentation approaches. Most of these studies rely on the ATLAS dataset [206], the first publicly available single-modality collection for sub-acute and chronic stroke. Its latest release, ATLAS v2.0, includes hundreds of T1-w MRI scans with expert lesion masks and serves as a benchmark for model training and evaluation.

Building on these foundations, this study investigates how fine-tuning and multimodal fusion can improve the segmentation of sub-acute and chronic stroke lesions. We first establish a single-modality nnU-Net v2 baseline on the public ATLAS v2.0 dataset (655 T1-w scans) using 5-fold cross-validation and evaluate its transferability to an independent clinical cohort (45 patients with paired T1-w and FLAIR MRI). On this internal dataset, we systematically compare T1-only and T1+FLAIR models across early, mid, and late fusion strategies, each assessed with metric averaging and ensembling under matched experimental conditions.

The main contributions of this work are threefold: (i) a reproducible evaluation framework that bridges public and clinical datasets for post-stroke lesion segmentation; (ii) one of the first systematic ablations of multimodal fusion strategies (early, mid, late) for sub-acute and chronic lesions; and (iii) a comprehensive clinical assessment including lesion-volume stratification.

Unlike previous works focused only on T1-w data [207, 208], this study provides an in-depth analysis of how modality integration and fine-tuning affect performance.

## 8.2 Methods and Materials

**Datasets.** Two datasets were used in this study: the public single-modality ATLAS v2.0 and an internal dual-modality clinical cohort. A summary of their characteristics is provided in Table 8.1, and lesion-volume distributions are shown in Fig. S1 (see Supplementary Material).

*ATLAS v2.0 (public, single modality).* The ATLAS v2.0 dataset [206] provides T1-w MRI scans with manually annotated stroke lesions. It includes 655 subjects for training, a public test set of 300, and a hidden test set of 316. All images are pre-aligned to the MNI-152 template and resampled to 1 mm isotropic resolution. The baseline single-modality model was trained and validated using five-fold cross-validation on the training set. Lesion volumes in this split averaged  $3.28 \times 10^1$  mL (range  $1.3 \times 10^{-2}$ – $4.97 \times 10^2$  mL).

*Internal dual-modality clinical cohort.* The internal dataset originates from a multicentre clinical investigation (NCT01677091) conducted in compliance with national ethical and data-protection regulations. All participants provided written informed consent, and the relevant ethics committee approved the protocol. Patients were enrolled in either the sub-acute (7 days–6 months) or chronic (>6 months) phase following ischemic or hemorrhagic stroke.

MRI data were acquired on Philips, Siemens, and GE scanners operating at 1.5 T or 3 T, including models such as Avanto, Skyra, Verio (Siemens); Achieva and Ingenia (Philips); and Optima, Signa HDx, and Discovery (GE). Each participant underwent paired T1-w and FLAIR acquisitions during the same session to ensure spatial correspondence. Lesions were manually segmented on FLAIR with T1-w guidance by a neuroimaging expert and verified by a neuroradiologist.

This cohort was used to fine-tune and evaluate the ATLAS-trained baseline. Models were first trained on T1-w images alone and then retrained on combined T1-w + FLAIR inputs to assess the contribution of multimodal fusion. The mean lesion volume was  $5.43 \times 10^1 \pm 5.91 \times 10^1$  mL (median  $3.52 \times 10^1$  mL; maximum  $2.67 \times 10^2$  mL).

**Preprocessing** The preprocessing pipeline is built upon our previous framework [209, 210], with the skull-stripping step based on Anima replaced by the HD-BET algorithm [211]. HD-BET was selected for its proven robustness across MRI vendors, field strengths, and acquisition protocols, making it particularly suitable for the heterogeneous multi-centre clinical dataset used in this study.

The pipeline, described here for T1-w images, comprised the following steps:

- i. *Brain extraction*: Non-brain tissue was removed using HD-BET.
- ii. *Bias correction*: Intensity non-uniformities were corrected using the N4 algorithm [212].
- iii. *Registration*: Each T1-w image was rigidly registered to the MNI-152 template using `animaPyramidalBMRegistration`. Lesion masks, when available, were transformed using nearest-neighbor interpolation and refined using hole-filling to preserve their topology.
- iv. *Reorientation*: All images and lesion masks were reoriented to the RAS (Right–Anterior–Superior) coordinate system for consistent alignment across subjects.

The same steps were applied to FLAIR images, except that each FLAIR was registered to its corresponding preprocessed T1-w scan from the same subject, which had already been aligned to MNI space. This ensured accurate within-subject correspondence between modalities while keeping interpolation minimal.

For the internal dataset, the full preprocessing pipeline was applied. For the ATLAS v2.0 dataset, which was already partially preprocessed, only HD-BET brain extraction and lesion refinement were performed. Small connected components with a volume  $< 2.5 \times 10^{-2}$  mL were removed to exclude microbleeds or interpolation artifacts, following clinical expert recommendations.

After these steps, nnU-Net v2 [20] automatically standardized voxel spacing and resampling parameters. The target spacing was set to the median per axis across the training set. Third-order spline interpolation was used for isotropic data. In contrast, for anisotropic volumes (maximum-to-minimum spacing ratio  $> 3$ ), nearest-neighbor interpolation was applied along the low-resolution axis to prevent artifacts.

Segmentation masks were one-hot encoded into background and lesion classes before resampling. Each channel was linearly interpolated and then collapsed into a discrete mask via the *argmax* operation to avoid label smoothing. Nearest-neighbor interpolation was again used for anisotropic data.

Voxel intensities were normalized on a per-image basis using z-score normalization (subtracting the mean and dividing by the standard deviation) [213]. All preprocessed data were organized in a BIDS-like directory structure [214] to ensure full traceability, reproducibility, and compatibility with downstream processing tools. Visual quality control was performed on all preprocessed images and masks to verify skull stripping, registration accuracy, and lesion-mask integrity.

**Model architectures** The segmentation pipeline was based on the nnU-Net v2 framework [20, 200], using the 3D *ResEncoder* variant (Figure 1). This architecture replaces standard convolutional blocks in the encoder with residual units to promote feature reuse and stabilize gradient flow. Each residual block consists of two consecutive Conv3d–InstanceNorm3d–LeakyReLU sequences with a residual connection before the final activation. The network comprises six resolution stages with 32, 64, 128, 256, and 320 feature maps, respectively. Downsampling is performed using strided convolutions at the start of each stage, while skip connections concatenate encoder and decoder features at matching resolutions. The decoder employs transposed convolutions for upsampling, followed by a Conv3d–InstanceNorm3d–LeakyReLU block. This configuration enables processing of  $128 \times 128 \times 128$  voxel patches with a 320-channel bottleneck representation at  $4 \times 4 \times 4$  spatial resolution. All models were trained with deep supervision using a combined soft Dice and cross-entropy loss.

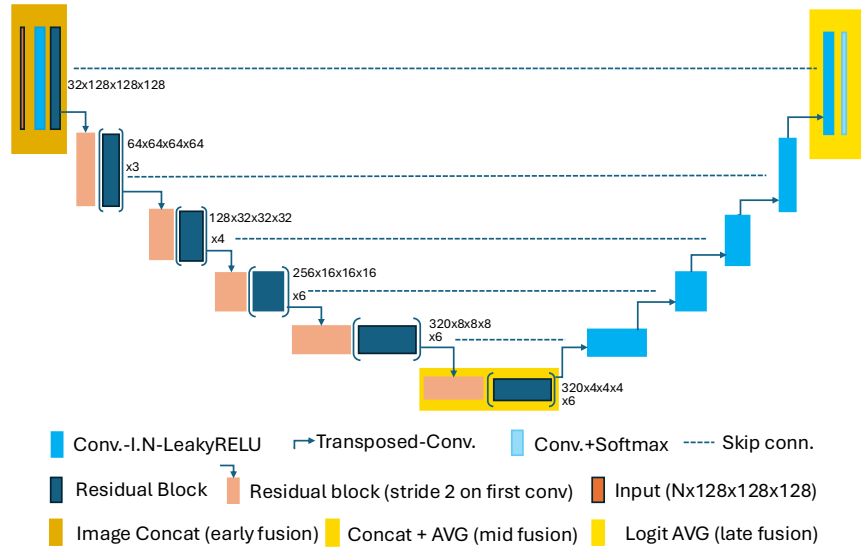
The baseline single-modality model was trained on the ATLAS v2.0 dataset and subsequently fine-tuned on the internal dataset. To investigate the added value of multimodal information, separate dual-modality models were then trained using both T1-w and FLAIR inputs under three fusion schemes described below.

*Fusion strategies.* To integrate T1-w and FLAIR information, three fusion schemes were investigated (Fig. S2, Supplementary Material):

- (i) *Early fusion*: T1-w and FLAIR volumes were concatenated as separate input channels to a single encoder–decoder model, enabling joint feature learning from the first convolutional layer. For stable optimization, the first-layer filters were initialized by duplicating the T1-w weights across both channels.

- (ii) *Mid fusion*: Two modality-specific encoders processed T1-w and FLAIR separately, and their bottleneck features were concatenated and projected through a  $1 \times 1 \times 1$  convolution before entering a shared decoder. Several fusion mechanisms were tested - including feature summation, averaging, concatenation with projection, residual weighting, squeeze-excitation gating [215], and FiLM-based modulation, to assess modality interaction [216]. The concatenation + projection design yielded the most stable and accurate results, while more complex gating or modulation schemes tended to overfit and reduce generalization. Initialization of the FLAIR encoder with T1-w pretrained weights proved essential for stable convergence and consistent performance across lesion sizes.
- (iii) *Late fusion*: Two independent encoder-decoder networks -one for each modality - were trained either separately or jointly, and their logit maps were fused voxel-wise to generate the final segmentation. Multiple fusion schemes were explored, including voxel-wise averaging, maximum, learnable scalar weighting, and learnable  $1 \times 1 \times 1$  convolutional fusion initialized as an equal-weight average. The learnable convolutional fusion achieved the best trade-off between stability and accuracy, outperforming both fixed and weighted alternatives. As in the mid-fusion setup, initializing the FLAIR branch with T1-w pretrained weights substantially improved convergence and final Dice and lesion-wise F1 scores compared with training both networks from scratch.

Only the best-performing configuration of each fusion strategy is reported in the manuscript, as these provided the most reproducible and representative results across experiments.



**Figure 8.1:** ResEnc-nnU-Net architecture. The notation  $() \times n$  indicates repeated blocks. Colored blocks highlight the type of fusion strategy applied (early, mid, or late) with the T1+FLAIR models.

**Training and inference protocol** All models were trained under a unified nnU-Net v2 protocol to ensure comparability across experiments. Training was performed on 3D input patches of  $128 \times 128 \times 128$  voxels, randomly sampled from the resampled volumes. Optimization employed stochastic gradient descent (SGD) with polynomial learning-rate decay from an initial value of 0.01 and weight decay of  $3 \times 10^{-5}$ . Each epoch comprised 250 iterations (batch size = 2), and training was run for 1,700 epochs. The loss combined soft Dice and cross-entropy terms, with deep supervision applied at all decoder scales using nnU-Net’s default decreasing weights. Dropout ( $p=0.2$ ) was applied in the encoder to reduce overfitting. Data augmentation included isotropic rescaling (scaling factor 0.85–1.25), random 3D rotations ( $\pm 15^\circ$ ), and sagittal mirroring. Each fold was trained with a fixed random seed to ensure reproducibility.

Fine-tuning and fusion experiments adopted the exact sampling, augmentation, and optimization settings as the baseline. All networks were initialized from the pretrained ATLAS baseline and fine-tuned for 200 epochs using SGD with the same parameters but a reduced learning rate of  $10^{-5}$ . The training and validation

splits were identical across all configurations to ensure a fair comparison. For each fold, the model checkpoint achieving the highest validation Dice score was selected for testing.

Baseline training on ATLAS v2.0 was performed on NVIDIA A40 and A100 GPUs, whereas fine-tuning and fusion experiments were run on NVIDIA RTX A5000 GPUs.

Inference used a sliding-window strategy with 50% overlap and Gaussian importance weighting. Test-time augmentation (TTA) was applied by mirroring the inputs along all three axes, and the logits from the augmented predictions were averaged before the final softmax activation. Post-processing removed isolated false positives using 3D connected-component analysis (minimum lesion size = 0.025 mL, consistent with preprocessing) and filled small intra-lesional holes. For anisotropic volumes, morphological operations were constrained to the high-resolution planes to mitigate through-plane artifacts.

Results were reported as mean metric values across the five folds. In addition, ensemble predictions were computed by averaging per-voxel logits from models trained with the same fold index before evaluating final performance.

**Evaluation metrics** Model performance was assessed using a combination of voxel-wise and lesion-wise metrics to provide a comprehensive evaluation and facilitate comparison with existing methods [205]. Voxel-wise metrics included precision, sensitivity, Dice score, and Jaccard index [21], as well as the Average Surface Distance (ASD), defined as the mean boundary distance between predicted and reference lesion masks. Lesion detection performance was quantified using the lesion-wise F1-score, where a predicted lesion was considered correctly detected if its connected voxels overlapped a reference lesion by at least 10%. The F1-score-1v metric, based on at least one shared voxel between prediction and reference, was also reported. Results were stratified by lesion volume - small ( $\leq 15$  mL), medium (15–70 mL), and large ( $> 70$  mL) - following established stroke lesion size definitions [217]. Statistical significance of performance differences was assessed using the Wilcoxon signed-rank test, with  $p < 0.05$  considered significant.

**Ablation study and comparative analysis** We conducted a systematic ablation study to quantify the contributions of key components, imaging modality, fusion strategy, and prediction aggregation to segmentation performance. All experiments were performed under identical training, inference, and evaluation conditions to ensure fair and controlled comparison.

For fine-tuning experiments on the internal dataset (Stages 2 and 3), each ATLAS fold served as initialization for a new five-fold cross-validation, yielding five fine-tuned models per fold and a total of 25 models per configuration. Training and validation splits were kept identical across all configurations to maintain strict comparability.

The ablation comprised three sequential stages:

*Stage 1 (Baseline transfer)* evaluated direct model transfer by applying ATLAS models to the internal dataset without fine-tuning.

*Stage 2 (T1-only fine-tuning)* fine-tuned the ATLAS models on internal T1-w data, performing a nested five-by-five cross-validation to obtain 25 models.

*Stage 3 (Dual-modality fusion)* testing three fusion strategies - early, mid, and late fusion - each following the same cross-validation procedure, yielding 25 models per strategy (75 in total).

All configurations were evaluated using two aggregation schemes: (i) *ensembling*, average per-voxel logits from models trained with the same fold index (pre-softmax); and (ii) *metric averaging*, average metrics across folds.

This framework enables a direct assessment of how fine-tuning, multimodal fusion, and prediction aggregation affect performance while holding all other factors constant.

## 8.3 Results

In this section, we detail the quantitative performance of our segmentation pipeline. The analysis transitions from a single-modality baseline trained on the ATLAS v2.0 dataset ( $n = 655$ ) to a multi-stage evaluation

on an internal dual-modality clinical cohort ( $n = 45$ ). Evaluation is performed via five-fold cross-validation using standard voxel-wise and lesion-wise metrics. Key results are stratified by lesion volume to highlight the model’s behavior across different clinical presentations, comparing the effectiveness of single-modality fine-tuning against early, mid, and late multimodal fusion strategies.

**ATLAS v2.0 cross-validation.** Table 8.2 summarizes the performance of the baseline model trained on the ATLAS v2.0 dataset using five-fold cross-validation, reporting mean, standard deviation, and median values. The baseline achieved a mean Dice score of 0.64. Voxel-wise precision and recall were 0.71 and 0.66, while lesion-wise F1 and F1-1v averaged 0.67 and 0.69, respectively. The Jaccard index was 0.53, and the ASD was 9.60 mm, reflecting good boundary alignment. Performance was consistent across folds, with low-to-moderate variability. Compared with prior work on ATLAS v2.0, our model achieved comparable or better results: Verma et al. reported Dice 0.65 and ASD 12.04 mm [207], while Huo et al. reported Dice 0.67 and F1-1v 0.56 [208].

Performance stratified by lesion volume (Table 8.3) showed a clear size-dependent trend. For small lesions (mean volume  $\approx 3.28$  mL,  $n = 423$ ), Dice was 0.55 and lesion-wise F1/F1-1v were 0.62/0.65. For medium lesions ( $\approx 38.6$  mL,  $n = 129$ ), Dice increased to 0.77 and F1/F1-1v to 0.74/0.73. For large lesions ( $\approx 152.6$  mL,  $n = 101$ ), Dice reached 0.88 and F1/F1-1v 0.82/0.80, indicating excellent overlap. These results indicate that segmentation accuracy improves with lesion size.

**Internal clinical cohort.** Table 8.4 summarizes the voxel- and lesion-wise performance metrics (mean  $\pm$  SD) of all models evaluated on the internal clinical cohort.

The directly transferred baseline model achieved a mean Dice score of 0.68, a precision of 0.78, a recall of 0.66, and a Jaccard of 0.55. Lesion-wise metrics indicated moderate detection performance (F1/F1-1v = 0.73/0.72) and an ASD of 6.95 mm. Applying ensemble averaging across folds consistently improved all metrics, increasing Dice to 0.70, Jaccard to 0.57, and F1/F1-1v to 0.79/0.76, while reducing ASD to 3.47 mm. Fine-tuning on internal T1-w data (Stage 2) further improved performance, yielding Dice 0.71 (precision 0.71, recall 0.76, Jaccard 0.58, F1/F1-1v 0.76/0.79) with ASD 7.19 mm. The ensembled T1 fine-tuned model achieved its best single-modality performance, with a Dice score of 0.72, Jaccard 0.60, F1/F1-1v 0.78/0.82, and ASD 3.37 mm. Extending to dual-modality (Stage 3) further improved segmentation accuracy. Among the fusion strategies, early fusion (joint input concatenation) achieved a Dice 0.70 and F1/F1-1v 0.75/0.78 (ASD 6.46 mm). The ensembled early-fusion model improved Dice to 0.74, Jaccard to 0.62, F1/F1-1v to 0.79/0.81, and ASD to 3.14 mm. Mid-fusion models (dual encoders merged at the bottleneck) performed slightly lower, with Dice 0.68–0.69, F1 0.74–0.77, and ASD 8.16–7.68 mm for individual and ensembled models, respectively. Late fusion, integrating modality-specific predictions at the output level, yielded the highest overall accuracy. The late-fusion model averaged reached Dice 0.73, F1/F1-1v 0.77/0.79, and ASD 5.05 mm, while the ensembled configuration achieved the best performance across all metrics: Dice 0.75, Jaccard 0.63, F1/F1-1v 0.80/0.81, and the lowest ASD, 2.94 mm.

A stratified analysis by lesion size (Table 8.3) confirmed these trends. For small lesions (mean  $\approx 4.96$  mL,  $n=14$ ), the T1–FLAIR late-fusion ensemble achieved the highest Dice (0.65) and F1-1v (0.83), while the baseline ensemble reached the best F1 (0.78). For medium lesions ( $\approx 39.65$  mL,  $n=18$ ), the T1–FLAIR late-fusion ensemble achieved the highest Dice (0.77) and F1-1v (0.85), and the T1 ensemble achieved the highest F1 (0.86). For large lesions ( $\approx 134.98$  mL,  $n=13$ ), the T1–FLAIR late-fusion ensemble achieved the highest Dice (0.83), while the T1 ensemble achieved the highest F1/F1-1v (0.83/0.81).

Statistical validation was also performed using paired, two-sided Wilcoxon signed-rank tests at the subject level ( $n=45$ ) for voxel-wise Dice and lesion-wise F1, both overall and stratified by lesion size (Tables S1–8, see Supplementary ). Ensembling significantly improved performance over single models: the ATLAS ensemble exceeded the single ATLAS baseline for Dice ( $p < 10^{-6}$ ) and F1 ( $p = 0.0004$ ), with the most significant gains for small and medium lesions. Fine-tuned T1 models also yielded significant gains over the baseline (Dice  $p = 0.0116$ ; F1  $p = 0.0102$ ), further enhanced by ensembling (Dice  $p \approx 7.3 \times 10^{-7}$ ; F1  $p = 0.0080$ ). Among multimodal configurations, early-fusion ensembles achieved significantly higher Dice ( $p \approx 4.4 \times 10^{-6}$ ) and F1 ( $p = 0.0385$ ), while mid-fusion models provided selective improvements for medium lesions ( $p = 0.0123$ ). Late

fusion achieved the most substantial and most robust gains, significantly outperforming all other baselines for both Dice (overall  $p < 10^{-6}$ ; small  $p = 0.0054$ ; medium  $p = 2.0 \times 10^{-4}$ ; large  $p = 0.0164$ ) and lesion-wise F1 (overall  $p = 0.0013$ ).

Overall, the combined quantitative and statistical analyses demonstrate that (1) fine-tuning substantially improves performance relative to direct transfer, (2) ensembling consistently enhances both voxel- and lesion-level accuracy, and (3) multimodal integration, particularly via late fusion, provides the most robust and spatially precise segmentation across lesion sizes. Qualitative examples are provided in Figures S3-4 (see Supplementary Material) for visual comparison.

## 8.4 Discussion

This study presents a reproducible deep learning framework for segmenting sub-acute and chronic stroke lesions using single-modality (T1-weighted) and dual-modality (T1-weighted + FLAIR) MRI. By systematically integrating fine-tuning, multimodal fusion, and model ensembling within the nnU-Net v2 architecture, we demonstrated how these components jointly improve segmentation accuracy and generalization across datasets and lesion sizes. All trained models and accompanying scripts are publicly released to ensure transparency and reproducibility (<https://github.com/alessandrodimatteo97/MultimodalityPostStrokeNet>).

**Fine-tuning improves domain transfer.** Models trained on the public ATLAS v2.0 dataset generalized reasonably well to our independent clinical cohort (Dice = 0.68). However, fine-tuning on local data yielded clear gains across both voxel- and lesion-level metrics (Dice = 0.72; F1 = 0.78), confirming that adapting pretrained networks to site-specific data helps overcome differences in scanners and acquisition protocols. This finding is consistent with previous neuroimaging studies showing that fine-tuning substantially improves model robustness when transferring between domains with differing intensity profiles or lesion characteristics.

**Multimodal MRI provides complementary information.** The addition of FLAIR images further improved segmentation performance. Among the evaluated fusion strategies, late fusion achieved the best overall accuracy (Dice = 0.75; F1 = 0.80; ASD = 2.94 mm), outperforming both early- and mid-fusion approaches. These results indicate that maintaining separate feature extraction for each modality before integration helps preserve complementary information - fine anatomical detail from T1-w images and lesion-related hyperintensities from FLAIR. This observation agrees with established neuroimaging evidence showing that FLAIR is particularly sensitive to chronic tissue damage, whereas T1-w imaging primarily captures structural integrity. Early fusion still provided strong results (Dice = 0.74; F1 = 0.79), suggesting that direct input concatenation remains effective when inter-modality registration is accurate.

**Ensembling improves performances.** Across all configurations, ensembling consistently improved Dice, F1, and boundary accuracy. Averaging probability maps across models produced smoother, more spatially coherent lesion maps, particularly for small and medium lesions. This improvement aligns with the well-established stabilizing effect of ensemble averaging, which reduces the influence of random initialization and local overfitting, leading to more reproducible and reliable predictions.

**Lesion size remains a key challenge.** Despite the overall performance gains, segmentation accuracy remained dependent on lesion volume. Large lesions were delineated with high precision (Dice  $\approx$  0.83), whereas smaller lesions were more difficult to detect (Dice  $\approx$  0.65), likely due to partial-volume effects and lower contrast. Fine-tuning and the inclusion of FLAIR partially mitigated this issue, but small-lesion detection remains an open problem. Future work may explore specialized loss functions, adaptive sampling, or uncertainty weighting to focus learning on low-volume lesions.

**Patient-level variability.** Examining results at the individual-subject level provided further insight into model behavior and potential *root causes* of segmentation failures. Patients with large, well-defined lesions generally showed stable performance across all configurations, while small or low-contrast lesions ac-

counted for most segmentation errors. In several cases, early fusion underperformed when minor T1–FLAIR misalignment or heterogeneous lesion intensity introduced conflicting signals, whereas late fusion remained more stable by maintaining independent modality representations before integration. Conversely, patients with strong and spatially consistent FLAIR hyperintensities benefited most from early fusion, which could exploit joint input features directly. Some apparent model errors were also attributable to data-related factors such as slight annotation inconsistencies or anisotropic resolution, particularly for small or low-contrast lesions. These observations highlight that the optimal fusion strategy may depend on lesion appearance, inter-modality alignment, and data quality, suggesting that future work should explore adaptive or confidence-weighted fusion mechanisms. Representative qualitative examples are shown in Figures S3–S4, and detailed per-patient analyses in Figures S5–S17 (see Supplementary Material).

**Comparison with previous works.** On the ATLAS v2.0 benchmark, our baseline model achieved performance comparable to or better than existing methods: lower surface distance (ASD = 9.6 mm vs. 12 mm in Verma *et al.*), higher lesion-wise F1-score (0.69 vs. 0.56 in Huo *et al.*), and similar Dice values (0.64 vs. 0.65–0.67) [207, 208]. These improvements likely stem from the standardized preprocessing adopted in our pipeline - HD-BET skull stripping, N4 bias-field correction, and intensity normalization - which were inconsistently applied in earlier studies.

**Combined insights.** Taken together, the results demonstrate that each methodological component contributed complementary strengths. Fine-tuning enhanced cross-domain adaptation, ensembling improved prediction stability and boundary coherence, and multimodal fusion enriched tissue contrast and contextual information. Among the tested fusion strategies, late fusion proved most effective, suggesting that preserving modality-specific encoders before integration better retains complementary cues from T1-weighted and FLAIR MRI. A representative example in Figure S3 (see Supplementary Material) illustrates these effects: the dual-modality ensemble achieved a Dice score of 0.81 compared to 0.60 for the single-modality model, reflecting more precise boundary delineation when FLAIR information was incorporated. Figure S4 (see Supplementary Material) further highlights the differences between the three fusion strategies, with Dice scores of 0.69, 0.41, and 0.76 for early, mid, and late fusion, respectively. Overall, early and late fusion remained competitive when modalities were well aligned, whereas mid-fusion offered limited benefits, likely due to partial feature mixing at the bottleneck stage.

**Limitations and future directions.** This study has several limitations. First, the internal dataset used for fine-tuning and evaluation, though diverse, was modest in size (45 subjects). Larger, multi-site datasets will be essential to confirm generalizability and assess robustness across scanners. Second, segmentation accuracy for small or low-contrast lesions remains limited, emphasizing the need for architectures and loss formulations tailored to class imbalance. Finally, minor T1–FLAIR misalignments occasionally degraded early-fusion performance, highlighting the importance of precise multimodal registration. Future work could extend the framework to diffusion or perfusion MRI and explore uncertainty-aware fusion methods to improve reliability in clinical deployment.

## 8.5 Conclusion

This study presented a reproducible deep learning framework for segmenting sub-acute and chronic stroke lesions from MRI, integrating fine-tuning, multimodal fusion, and model ensembling within the nnU-Net v2 architecture. A single-modality baseline trained on the public ATLAS v2.0 dataset achieved competitive performance, and fine-tuning on an independent clinical cohort significantly improved cross-domain generalization. Incorporating FLAIR alongside T1-weighted MRI further improved segmentation accuracy, with the late-fusion ensemble achieving the best overall performance (Dice = 0.75; F1 = 0.80; ASD = 2.94 mm). Early fusion also yielded consistent gains when inter-modality alignment was precise, whereas mid-fusion offered limited benefit. Across all configurations, ensembling systematically improved voxel- and lesion-level accuracy, producing smoother and more spatially coherent boundaries, particularly for small and medium



**Table 8.1:** Patient count and lesion volume statistics for each dataset. Lesion volumes are reported in milliliters (mL) as mean  $\pm$  standard deviation, median, and maximum values.

| Dataset         | Patients<br>( $n$ ) | Mean<br>(mL)                             | Median<br>(mL)     | Max<br>(mL)        |
|-----------------|---------------------|------------------------------------------|--------------------|--------------------|
| ATLAS v2.0      | 655                 | $3.28 \times 10$<br>$\pm 6.09 \times 10$ | $2.5 \times 10^1$  | $4.97 \times 10^2$ |
| Internal cohort | 45                  | $5.43 \times 10$<br>$\pm 5.91 \times 10$ | $3.52 \times 10^1$ | $2.67 \times 10^2$ |

**Table 8.2:** Performance metrics across five folds on the ATLAS v2.0 dataset. Each cell reports the mean value, followed by the standard deviation in parentheses, and the median value.

| Fold | Recall | Precision | Dice   | Jaccard | ASD (mm) | F1-score | F1-score-1v |
|------|--------|-----------|--------|---------|----------|----------|-------------|
| 0    | 0.68   | 0.69      | 0.64   | 0.53    | 9.13     | 0.66     | 0.69        |
|      | (0.30) | (0.28)    | (0.28) | (0.26)  | (15.04)  | (0.32)   | (0.29)      |
|      | 0.81   | 0.79      | 0.75   | 0.60    | 2.39     | 0.67     | 0.67        |
| 1    | 0.62   | 0.75      | 0.63   | 0.52    | 11.18    | 0.66     | 0.67        |
|      | (0.29) | (0.26)    | (0.28) | (0.27)  | (16.41)  | (0.30)   | (0.27)      |
|      | 0.75   | 0.85      | 0.73   | 0.57    | 2.88     | 0.67     | 0.67        |
| 2    | 0.63   | 0.66      | 0.60   | 0.50    | 11.43    | 0.65     | 0.66        |
|      | (0.33) | (0.33)    | (0.32) | (0.29)  | (18.50)  | (0.37)   | (0.34)      |
|      | 0.77   | 0.79      | 0.72   | 0.56    | 2.60     | 0.67     | 0.67        |
| 3    | 0.69   | 0.74      | 0.68   | 0.56    | 8.05     | 0.69     | 0.72        |
|      | (0.27) | (0.28)    | (0.25) | (0.25)  | (15.46)  | (0.31)   | (0.28)      |
|      | 0.78   | 0.85      | 0.76   | 0.61    | 2.07     | 0.67     | 0.67        |
| 4    | 0.69   | 0.72      | 0.66   | 0.54    | 7.78     | 0.71     | 0.72        |
|      | (0.29) | (0.26)    | (0.26) | (0.25)  | (13.97)  | (0.29)   | (0.28)      |
|      | 0.80   | 0.83      | 0.75   | 0.60    | 2.22     | 0.67     | 0.67        |
| Mean | 0.66   | 0.71      | 0.64   | 0.53    | 9.60     | 0.67     | 0.69        |

lesions. Nevertheless, lesion size remained a key challenge, with small lesions still more difficult to detect and delineate accurately.

These findings underscore the complementary value of multimodal MRI and demonstrate that fine-tuning and ensemble learning substantially improve lesion segmentation in heterogeneous, real-world clinical data. To further advance this work, we plan to curate and release a large, well-annotated dual-modality (T1-w + FLAIR) dataset for sub-acute and chronic stroke, addressing the current lack of publicly available multimodal resources. A larger, more diverse cohort will enable systematic analysis of factors contributing to segmentation failures, including lesion complexity, imaging quality, and scanner variability. Future efforts will also focus on refining lesion-aware loss functions, exploring uncertainty-aware fusion strategies, and promoting clinical translation through user-centered design and integration into neuroimaging workflows.

## 8.6 Data availability statement

The ATLAS v2.0 dataset is publicly available and can be accessed at [https://fcon\\_1000.projects.nitrc.org/indi/retro/atlas.html](https://fcon_1000.projects.nitrc.org/indi/retro/atlas.html). Data from internal studies are not publicly accessible but can be obtained from the corresponding author upon reasonable request.

**Table 8.3:** Stratified performance by lesion volume. Entries report mean values. Single (\*) and double asterisks (\*\*) denote the best and second-best performing models, respectively.

| Model                                 | Dataset    | Small ( $\leq 15$ mL) |               |               | Medium (15–70 mL) |               |               | Large ( $>70$ mL) |               |               |
|---------------------------------------|------------|-----------------------|---------------|---------------|-------------------|---------------|---------------|-------------------|---------------|---------------|
|                                       |            | Dice                  | F1-score      | F1-score-1v   | Dice              | F1-score      | F1-score-1v   | Dice              | F1-score      | F1-score-1v   |
| <i>Baseline</i>                       | ATLAS v2.0 | 0.55                  | 0.62          | 0.65          | <b>0.77**</b>     | 0.74          | 0.73          | <b>0.88*</b>      | <b>0.82**</b> | <b>0.80**</b> |
| <i>Baseline</i>                       | Internal   | 0.57                  | 0.73          | 0.75          | 0.70              | 0.74          | 0.73          | 0.78              | 0.72          | 0.66          |
| <i>Baseline ensemble</i>              | Internal   | <b>0.60**</b>         | <b>0.78**</b> | 0.78          | 0.71              | 0.79          | 0.78          | 0.79              | 0.78          | 0.70          |
| <i>T1 fine-tuned</i>                  | Internal   | 0.55                  | 0.65          | 0.79          | 0.75              | <b>0.83**</b> | 0.82          | 0.81              | 0.80          | 0.77          |
| <i>T1 fine-tuned ensemble</i>         | Internal   | 0.58                  | 0.65          | 0.80          | 0.76              | <b>0.86*</b>  | <b>0.84**</b> | 0.82              | <b>0.83*</b>  | <b>0.81*</b>  |
| <i>T1+FLAIR early fusion</i>          | Internal   | 0.52                  | 0.61          | 0.75          | <b>0.77**</b>     | <b>0.83**</b> | 0.84          | 0.82              | 0.80          | 0.74          |
| <i>T1+FLAIR early fusion ensemble</i> | Internal   | 0.59                  | 0.77          | <b>0.83*</b>  | <b>0.78*</b>      | 0.80          | <b>0.85*</b>  | 0.82              | 0.80          | 0.75          |
| <i>T1+FLAIR mid fusion</i>            | Internal   | 0.46                  | 0.60          | 0.71          | 0.76              | 0.82          | <b>0.85*</b>  | 0.81              | 0.77          | 0.75          |
| <i>T1+FLAIR mid fusion ensemble</i>   | Internal   | 0.48                  | 0.70          | 0.72          | <b>0.77**</b>     | <b>0.83**</b> | <b>0.85*</b>  | 0.82              | 0.79          | 0.76          |
| <i>T1+FLAIR late fusion</i>           | Internal   | <b>0.60**</b>         | 0.72          | 0.75          | <b>0.77**</b>     | 0.80          | 0.82          | 0.82              | 0.77          | 0.77          |
| <i>T1+FLAIR late fusion ensemble</i>  | Internal   | <b>0.65*</b>          | <b>0.81*</b>  | <b>0.82**</b> | <b>0.77**</b>     | 0.80          | <b>0.85*</b>  | <b>0.83**</b>     | 0.77          | 0.73          |

**Table 8.4:** Performance metrics on the Internal dataset. Metrics are reported as mean (standard deviation) and median. Single (\*) and double asterisks (\*\*) denote the best and second-best performing models, respectively.

| Model                                 | Recall                          | Precision                       | Dice                           | Jaccard                         | ASD                             | F1-score                       | F1-score-1v                    |
|---------------------------------------|---------------------------------|---------------------------------|--------------------------------|---------------------------------|---------------------------------|--------------------------------|--------------------------------|
| <i>Baseline</i>                       | 0.66<br>(0.26)<br>0.73          | <b>0.78**</b><br>(0.22)<br>0.86 | 0.68<br>(0.21)<br>0.72         | 0.55<br>(0.21)<br>0.56          | 6.95<br>(15.25)<br>2.49         | 0.73<br>(0.26)<br>0.73         | 0.72<br>(0.26)<br>0.73         |
| <i>Baseline ensemble</i>              | 0.67<br>(0.26)<br>0.74          | <b>0.80*</b><br>(0.22)<br>0.88  | 0.70<br>(0.21)<br>0.76         | 0.57<br>(0.21)<br>0.61          | 3.47<br>(3.09)<br>2.41          | <b>0.79**</b><br>(0.27)<br>1.0 | 0.76<br>(0.28)<br>0.80         |
| <i>T1 fine-tuned</i>                  | <b>0.76**</b><br>(0.24)<br>0.85 | 0.71<br>(0.25)<br>0.78          | 0.71<br>(0.21)<br>0.78         | 0.58<br>(0.21)<br>0.64          | 7.19<br>(15.58)<br>2.58         | 0.76<br>(0.29)<br>0.88         | 0.79<br>(0.25)<br>0.88         |
| <i>T1 fine-tuned ensemble</i>         | <b>0.77*</b><br>(0.24)<br>0.86  | 0.73<br>(0.24)<br>0.80          | 0.72<br>(0.21)<br>0.79         | 0.60<br>(0.21)<br>0.65          | 3.37<br>(3.09)<br>2.33          | 0.78<br>(0.31)<br>1.0          | <b>0.82*</b><br>(0.27)<br>1.0  |
| <i>T1+FLAIR early fusion</i>          | 0.73<br>(0.24)<br>0.80          | 0.74<br>(0.24)<br>0.80          | 0.70<br>(0.24)<br>0.79         | 0.59<br>(0.23)<br>0.65          | 6.46<br>(14.14)<br>2.17         | 0.75<br>(0.30)<br>0.87         | 0.78<br>(0.25)<br>0.87         |
| <i>T1+FLAIR early fusion ensemble</i> | 0.74<br>(0.24)<br>0.82          | 0.77<br>(0.24)<br>0.83          | <b>0.74*</b><br>(0.23)<br>0.82 | <b>0.62**</b><br>(0.22)<br>0.70 | <b>3.14**</b><br>(4.02)<br>1.72 | <b>0.79**</b><br>(0.30)<br>1.0 | <b>0.81**</b><br>(0.26)<br>1.0 |
| <i>T1+FLAIR mid fusion</i>            | 0.70<br>(0.27)<br>0.79          | 0.71<br>(0.26)<br>0.80          | 0.68<br>(0.25)<br>0.78         | 0.56<br>(0.24)<br>0.64          | 8.16<br>(15.81)<br>2.68         | 0.74<br>(0.29)<br>0.87         | 0.77<br>(0.27)<br>0.9          |
| <i>T1+FLAIR mid fusion ensemble</i>   | 0.70<br>(0.28)<br>0.80          | 0.74<br>(0.26)<br>0.82          | 0.69<br>(0.26)<br>0.80         | 0.58<br>(0.24)<br>0.67          | 7.68<br>(16.39)<br>2.33         | 0.77<br>(0.32)<br>1.0          | 0.78<br>(0.31)<br>1.0          |
| <i>T1+FLAIR late fusion</i>           | 0.75<br>(0.24)<br>0.84          | 0.76<br>(0.23)<br>0.84          | 0.73<br>(0.20)<br>0.80         | 0.61<br>(0.20)<br>0.67          | 5.05<br>(10.25)<br>2.21         | 0.77<br>(0.27)<br>0.88         | 0.79<br>(0.26)<br>0.87         |
| <i>T1+FLAIR late fusion ensemble</i>  | <b>0.76**</b><br>(0.24)<br>0.85 | <b>0.78**</b><br>(0.22)<br>0.85 | <b>0.75*</b><br>(0.20)<br>0.81 | <b>0.63*</b><br>(0.20)<br>0.67  | <b>2.94*</b><br>(3.30)<br>1.81  | <b>0.80*</b><br>(0.29)<br>1.0  | <b>0.81**</b><br>(0.29)<br>1.0 |

## Part V

# Conclusions

## Chapter 9

# Conclusion and Future Directions

Stroke remains a leading cause of long-term disability globally, underscoring the critical need for effective and personalized rehabilitation strategies. The research presented in this doctoral thesis aimed to address key challenges in post-stroke care by developing and integrating advanced technological systems, data standardization protocols, and AI techniques for quantitative assessment and rehabilitation, particularly focusing on upper-limb function. This work spanned multiple interconnected domains, from portable sensing and immersive telerehabilitation to standardized data management and neuroimaging analysis, converging towards a coherent, data-driven framework for enhancing post-stroke recovery.

### 9.1 Summary of Contributions

The primary contributions of this thesis are organized along four main research axes, directly addressing the research questions outlined in the introduction:

1. **Advanced Technological Systems for Hand Telerehabilitation:** To address the need for robust and usable systems for daily-life rehabilitation (**RQ1**) and personalized assessment (**RQ2**), several innovative platforms were developed.
  - The **Virtual Glove (VG) framework**, based on a dual-LMC setup, was employed for high-precision, real-time hand kinematic data acquisition in a static setting. A cooperative patient-therapist telerehabilitation architecture was built around the VG, enabling remote monitoring, task assignment, and quantitative data analysis through a dedicated server infrastructure and graphical user interfaces (GUI).
  - Leveraging the VG, the **Double Digital Twin (DDT) concept** was introduced and implemented. This methodology models both the healthy and impaired hands of a patient, allowing the healthy hand's motion (after appropriate calibration and transformation using techniques like the Kabsch algorithm) to serve as a ground truth (GT) reference during mirrored task execution in Virtual Reality (VR). This enables real-time comparison, visual guidance, and quantitative assessment of motor deficits using metrics like Dynamic Time Warping (DTW).
  - A **portable, head-mounted hand-tracking system** combining the Leap Motion Controller 2 (LMC) with a single-board computer (SBC) was designed and validated. This system enables lightweight, autonomous motion capture in unconstrained environments, overcoming limitations of traditional lab-based or PC-tethered setups. Performance evaluations confirmed its suitability regarding power consumption, thermal management, CPU usage, and real-time frame rates (FPS).
2. **Efficient AI for Real-Time Functional Analysis:** Recognizing the computational constraints

of portable and wearable systems used in rehabilitation (3), this work explored lightweight neural network architectures.

- **BNNAction-Net**, a novel architecture based on Binary Neural Networks (BNNs), was developed for Hand Gesture Recognition (HGR). By employing 1-bit representations for weights and activations and integrating a binarized ResNet backbone with the ActionNet framework enhanced by Knowledge Distillation (KD), BNNAction-Net achieves competitive accuracy on the EgoGesture dataset while drastically reducing computational complexity and memory footprint, making it suitable for deployment on resource-constrained Head-Mounted Displays (HMD).
3. **Standardization of Quantitative Rehabilitation Data:** To address the critical lack of interoperability hindering the field of Quantitative Physical Rehabilitation (QPR) (4), a novel data standard was proposed.
    - A **DICOM-based standard, termed Physical Rehabilitation Information Object Definition (PR-IOD)**, was designed specifically for QPR time-series data. Leveraging the existing Digital Imaging and Communications in Medicine (DICOM) framework, particularly the Waveform Information Entity (IE), the PR-IOD provides a structured, extensible, and interoperable format for storing kinematic and related data from various rehabilitation devices. A proof-of-concept implementation demonstrated the successful conversion and visualization of VG data into the PR-IOD format, paving the way for improved data sharing, integration with Electronic Health Records (EHR), and large-scale multi-center research.
  4. **Neuroimaging-based Assessment via Deep Learning:** Complementing the functional assessments, this thesis investigated AI methods for analyzing structural brain changes post-stroke (5).
    - **Deep learning models based on the nn-UNet framework** were developed and validated for the automated segmentation of sub-acute and chronic stroke lesions from multi-modal MRI data (T1-W and FLAIR). Training was performed on the public ATLAS v2.0 dataset, followed by fine-tuning and evaluation on an internal clinical dataset (AVCPOSTIM). The results demonstrated the effectiveness of the approach, particularly the benefits of using dual-modality input (T1-W + FLAIR) and model ensembling, which significantly improved segmentation accuracy (Dice score, ASD) and lesion detection (F1-score), especially for challenging small and medium-sized lesions, compared to baseline and single-modality models.

## 9.2 Synthesis and Significance

Together, these contributions form an integrated technological ecosystem that bridges the gap between advanced sensing, intelligent data analysis, standardized information management, and neuroimaging-based structural assessment for post-stroke rehabilitation. The framework moves beyond traditional qualitative approaches by enabling objective, quantitative monitoring of motor function through systems like the portable tracker and the VG. The introduction of the DDT paradigm offers a novel method for personalized, comparative assessment by leveraging the patient's own healthy limb as a reference. Efficient AI models like BNNAction-Net ensure that real-time functional interpretation, such as gesture recognition, is feasible even on portable devices central to home-based and immersive rehabilitation. Crucially, the proposed PR-IOD standard addresses the fundamental challenge of data fragmentation, fostering interoperability and enabling the creation of large-scale, shareable datasets. Finally, integrating advanced MRI lesion segmentation provides objective biomarkers of brain injury that can inform prognosis and personalize rehabilitation strategies, linking functional recovery to underlying neural substrates. This holistic, data-centric continuum, from capturing movement to managing data and assessing brain structure, advances the vision of precise, personalized, and quantitatively driven post-stroke care.

### 9.3 Limitations

Despite the promising results, this research has several limitations that warrant consideration. The clinical validation of the developed technological systems (VG, VG-Digital Twin, portable tracker) and AI models (BNNAction-Net, segmentation models) has been preliminary. The DT methodology, for instance, was primarily tested on healthy individuals, and its efficacy in patient populations requires further investigation. Similarly, while BNNAction-Net showed excellent efficiency, its performance was evaluated offline on a specific dataset, and real-world deployment on Head-Mounted-Display needs thorough testing. The proposed PR-IOD standard represents a significant step towards interoperability, but its adoption requires consensus-building within the wider DICOM and rehabilitation communities. The lesion segmentation models, while achieving state-of-the-art performance, were fine-tuned on a moderately sized internal dataset, and their generalizability across diverse clinical settings, scanner types, and imaging protocols needs further validation. Furthermore, the integration of the different components into a seamless clinical workflow requires significant engineering effort and usability testing with both patients and therapists.

### 9.4 Future Research Directions

Building upon the findings and limitations of this thesis, several avenues for future research emerge:

1. **Extensive Clinical Validation:** Conduct large-scale, multi-center clinical trials to rigorously evaluate the effectiveness, usability, and long-term impact of the portable hand tracker, the VG telerehabilitation framework, and the DDT methodology in diverse patient cohorts across different stages of stroke recovery (acute, sub-acute, chronic) and varying levels of impairment. Understand if the post-stroke stage make possible to test it. Longitudinal studies are needed to correlate quantitative metrics derived from these systems with established clinical scales and patient-reported outcomes.
2. **System Integration and Enhancement:** Further develop the portable head-mounted system with features like start/stop controls and adaptive fan speed for improved efficiency and create with it a new dataset for hand gesture recognition. Integrate the BNNAction-Net gesture recognition capabilities directly into the VG and DDT frameworks to enable real-time, intention-aware feedback and dynamically adaptive rehabilitation tasks. Explore integration with complementary rehabilitation technologies, such as robotics or functional electrical stimulation. Conduct formal Human-Computer Interaction (HCI) and User Experience (UX) studies involving patients and therapists to refine system interfaces and interaction paradigms.
3. **Advancement and Promotion of Data Standardization:** Actively engage with DICOM working groups and the rehabilitation research community to refine and promote the adoption of the PR-IOD standard. Develop open-source libraries and tools to facilitate the creation, validation, and visualization of PR-IOD compliant data. Extend the PR-IOD model to encompass other relevant QOR data types, such as electromyography (EMG) data, force/pressure sensor data, and data from full-body motion capture systems. Investigate the use of federated learning architectures built upon this standard to enable privacy-preserving analysis across multiple institutions.
4. **AI Model Development and Multimodal Fusion:** Enhance the robustness of stroke lesion segmentation models, particularly for small and heterogeneous lesions, by exploring different advanced deep learning architectures (e.g., transformers, attention mechanisms), self-supervised or weakly supervised learning techniques, and training on larger, more diverse multimodal datasets. Develop predictive AI models that integrate quantitative functional data (kinematics from VG/DDT, gesture performance from HGR) with structural neuroimaging data (lesion volume, location, connectivity metrics derived from segmentation) to forecast rehabilitation outcomes and personalize therapy intensity and focus. Continue research into efficient network architectures (like BNNs) to balance accuracy and deployability for real-time applications.

5. **Structure-Function Relationship Exploration:** Utilize the integrated framework to explicitly investigate the relationship between the characteristics of stroke lesions (segmented via MRI) and the patterns of functional impairment and recovery (quantified via motion tracking and gesture analysis). This could lead to a deeper understanding of neural reorganization mechanisms and help identify imaging-based biomarkers that predict response to specific rehabilitation interventions.

## Concluding Remarks

This thesis has contributed a multi-faceted technological framework aimed at advancing the field of post-stroke rehabilitation and assessment. By integrating novel hardware for motion capture, innovative concepts like the Double Digital Twin, efficient AI for real-time analysis and neuroimaging, and a foundational proposal for data standardization, this work lays the groundwork for a more quantitative, personalized, and effective approach to stroke care. While further validation and development are necessary, the presented ecosystem holds significant potential to empower clinicians with objective tools, engage patients through interactive technologies, and ultimately improve functional outcomes for stroke survivors worldwide.



## Part VI

## Lists

# List of Figures

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |    |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Illustration of a chronic stroke lesion across different imaging modalities: CT, MRI-FLAIR, T2-weighted, DWI, and ADC; highlighting the complementary diagnostic information each provides; image taken from <a href="https://radiopaedia.org/">https://radiopaedia.org/</a> ("Stroke: Evolution from acute to chronic infarction - radiology video tutorial (CT, MRI)). . . . .                                                                                                                       | 6  |
| 1.2 | Post-Stroke patient images across multiple imaging modalities: T1-W and Flair [6]. . . . .                                                                                                                                                                                                                                                                                                                                                                                                             | 8  |
| 1.3 | Schematic representation of a multilayer perceptron (MLP). The network is composed of an input layer $\mathbf{x}$ , two hidden layers $\mathbf{h}^{(1)}$ and $\mathbf{h}^{(2)}$ , and an output layer $\hat{\mathbf{y}}$ . Each neuron computes a weighted sum of its inputs, adds a bias term, and applies a nonlinear activation function. The dense connectivity between layers enables the model to learn hierarchical feature representations and approximate complex nonlinear mappings. . . . . | 9  |
| 1.4 | Convolutional Neural Network schema [18] . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 11 |
| 1.5 | Schematic representation of the 3D U-Net architecture. image taken from the original paper [19] . . . . .                                                                                                                                                                                                                                                                                                                                                                                              | 11 |
| 1.6 | Schematic overview of the <b>nnU-Net</b> pipeline [20]. The framework automatically configures preprocessing, network architecture, training, and post-processing steps based on dataset characteristics. This adaptive design allows nnU-Net to achieve state-of-the-art performance across diverse biomedical segmentation tasks without manual tuning. . . . .                                                                                                                                      | 12 |
| 2.1 | Traditional qualitative approaches versus technology-driven quantitative approaches in post-stroke rehabilitation [31] . . . . .                                                                                                                                                                                                                                                                                                                                                                       | 18 |
| 2.2 | ARMin-IV, a 7-DoF robotic device designed for upper-limb rehabilitation [40]. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                  | 20 |
| 2.3 | Wearable device manus xsense metaglove [52]. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 22 |
| 2.4 | Assembly of the Virtual Glove system, consisting of two Leap Motion Controller (V1) sensors mounted orthogonally on an aluminum base. The image shows a potential user positioning the hand above the device, demonstrating the correct placement during task execution. . . .                                                                                                                                                                                                                         | 24 |
| 3.1 | Hardware and software architecture of the framework. The hardware consists of the rehabilitation system with the VG and a server for data storage and for hosting the software. The software consists of a GUI for the Therapy and one for the Administrator. The persons involved in the process are the Therapist (or Therapists), the Patient and the system Administrator. . . . .                                                                                                                 | 31 |
| 3.2 | VG final implementation, including sensor assembly (left), NUC-based driving unit (right), and video screen for the active interaction (center). . . . .                                                                                                                                                                                                                                                                                                                                               | 32 |
| 3.3 | Skeletal hand model (skin contour is reproduced for reference) in which each joint is represented by a green sphere. . . . .                                                                                                                                                                                                                                                                                                                                                                           | 34 |
| 3.4 | Binary switch approach: depending the orientation of the hand, one sensor data are used, while the others are discarded. . . . .                                                                                                                                                                                                                                                                                                                                                                       | 34 |
| 3.5 | Use Case diagram presenting the functionalities and actors involved to the framework usage. . . . .                                                                                                                                                                                                                                                                                                                                                                                                    | 35 |
| 3.6 | Software architecture diagram representing all the system involved: Rehabilitation System, Server System, VG, Therapist System, and Administrator System. . . . .                                                                                                                                                                                                                                                                                                                                      | 37 |
| 3.7 | Swim Lane Diagram showing the pipeline execution. The pipeline describes the rehabilitation process from the beginning when the Patient is suggested to use the VG until data are processed and shared among therapists. Obviously, the process can be continued by the Therapist, assigning a new task or updating the task's difficulty, or restarting with a new Patient. . . . .                                                                                                                   | 39 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |    |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.8  | VG Data flow pipeline. Here one can see (a) the Patient performing a task, (b) The generated Hand Model, (c) The Therapist Interface for data analysis and visualization, and (d) created plots presenting joint movement. . . . .                                                                                                                                                                                                                                                       | 40 |
| 3.9  | Summary screen loaded before Patient starts the task b) Keyboard task execution c) Interceptor task execution c) Summary screen loaded when all repetitions are completed. . . . .                                                                                                                                                                                                                                                                                                       | 41 |
| 4.1  | DDT formation. First, the DT of the healthy hand (DTH) is formed by a single execution of the task, generating the Ground Truth. Then, a repeated rehabilitation sessions is performed with the impaired hand, generating the DT for the impaired hand (DTI). . . . .                                                                                                                                                                                                                    | 45 |
| 4.2  | DT Internal Model: Data provided are highlighted in black; Data processed and provided are highlighted in skyblue . . . . .                                                                                                                                                                                                                                                                                                                                                              | 46 |
| 4.3  | Qualitative evaluation of the discrepancies during the execution of the task, displaying various frames: a) Opened Hand; b) Closed Hand; c) Opened Hand ; d) Thumb Flexion; e) Index Flexion; f) Middle Flexion; g) Ring Flexion; h) Little Flexion. . . . .                                                                                                                                                                                                                             | 48 |
| 4.4  | Fingertip trajectories on the x, y, and z axes for all five fingers during the 29-second task. Blue = DTH (right hand); other colors = DTIs (left hand). . . . .                                                                                                                                                                                                                                                                                                                         | 50 |
| 5.1  | System embodiment: (a) pocket location of the SBC and power bank with respect to the head-mounted LMC; (b) details of the LMC support that enables changing the vertical orientation of the sensor. . . . .                                                                                                                                                                                                                                                                              | 54 |
| 5.2  | Timeline of the LMC: The sensor was first designed to be driven by a personal computer. In recent embodiments over the last two years, it has started to be driven by an SBC. Further, in its latest release, the ROI was increased and back-hand tracking was enabled, allowing for complete head-mounted embodiments. . . . .                                                                                                                                                          | 55 |
| 5.3  | System architecture. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 56 |
| 5.4  | System embodiment: (1) the RP enclosed in its case, (2) the power bank, (3) the LMC inside the head-mounted support. The presence of the monitor, not used in the final embodiment, is necessary during system calibration to set the values of the parameters. . . . .                                                                                                                                                                                                                  | 57 |
| 5.5  | Workflow and interactions between the hardware and software components of the proposed system. . . . .                                                                                                                                                                                                                                                                                                                                                                                   | 57 |
| 5.6  | Voltage, current, and power measurements for both considered scenarios. . . . .                                                                                                                                                                                                                                                                                                                                                                                                          | 60 |
| 5.7  | The aluminum heatsinks used during the measurements for the baseline and the proposed assembly are highlighted with yellow boxes. The figure also shows the RP enclosed in the case and the fan used for cooling in the proposed assembly. . . . .                                                                                                                                                                                                                                       | 62 |
| 5.8  | CPU usage (in %) and temperature (in °C) for both scenarios. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                     | 62 |
| 5.9  | The frame rate of the LMC was read from the internal data structure and reported. The original averaged values also contained decimals, which have been eliminated for convenience. . . . .                                                                                                                                                                                                                                                                                              | 63 |
| 5.10 | The plots of the spatial positions of each fingertip of both hands as measured by the LMC. To improve visualization, measurements for each spatial coordinate are shown in separate plots. For each coordinate, measurements of the fingertips from the same hand are shown in the same plot with different colors to facilitate comparison. The vertical dotted lines represent particular hand poses (shown in the scatter plots from (a–h) to contextualize the plots better. . . . . | 64 |
| 6.1  | Pipeline and architecture of the proposed system. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                | 67 |
| 7.1  | PR proprietary data (left) are merged, through a PR to DICOM translator, into a shared Dicom-based data format (right): in this way, data can be compared and shared easier, thus improving interoperability. This figure has been adapted from Fig.1 in [150], where a similar architecture was implemented for EEG data. . . . .                                                                                                                                                       | 71 |
| 7.2  | Information Object Definition Model composition. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                 | 72 |
| 7.3  | DICOM conceptual framework where relationships and constraints among IEs are shown [174]. . . . .                                                                                                                                                                                                                                                                                                                                                                                        | 73 |
| 7.4  | The Information Model delineates the interconnections among the principal components of the standard [176]. . . . .                                                                                                                                                                                                                                                                                                                                                                      | 74 |
| 7.5  | Actors involved in the PR Pipeline and data generation. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                          | 75 |
| 7.6  | Relationship between IEs involved in the PR-IOD Model; every is composed of modules. Every module specifies its usage, which can be Mandatory (M), User Optional (U), or Conditionally Required (C). The Waveform IE has been further divided to show its composition (structure adapted from [190]). . . . .                                                                                                                                                                            | 76 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |     |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 7.7  | PR Waveform decomposition including the time acquisition time-series for a PR-Device (that does not have a fixed sampling rate); the Conservative Solution (A) involves using an additional Multiplex-group composed of one channel; the Non-Conservative Solution (B) designates a private element for its storage. . . . .                                                                                                                                                                                                                                      | 78  |
| 7.8  | Quantitative physical rehabilitation of the hand: from data acquisition to PR-IOD definition, passing through the hand model construction. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                | 79  |
| 7.9  | Overall structure of data in VG-Patient-Task within the PR-IOD: the used IEs are reported (as an example, the Frame of reference IE is not reported because not used in the VG-Patient-Task). . . . .                                                                                                                                                                                                                                                                                                                                                             | 80  |
| 7.10 | Viewer Interface for feature calculation and display: (A) Main Window for navigation through data and functions; (B) DICOM Field Window to show all fields, including TAG, VR, Name, and Value; (C) Trajectory feature components of a selected finger joint; (D) Velocity feature components calculated from the trajectory of the selected finger joint; (E) Acceleration feature components computed from the corresponding velocity; (F) Interactive Plot Window showing dynamic scatter plot animations of the hand as a whole or a specific finger. . . . . | 83  |
| 8.1  | ResEnc-nnU-Net architecture. The notation $() \times n$ indicates repeated blocks. Colored blocks highlight the type of fusion strategy applied (early, mid, or late) with the T1+FLAIR models. . . . .                                                                                                                                                                                                                                                                                                                                                           | 89  |
| A.1  | Distribution of lesion volumes for individual subjects in each dataset. The triangle represents the mean score, and the yellow bar represents the standard deviation. . . . .                                                                                                                                                                                                                                                                                                                                                                                     | 131 |
| A.2  | Sagittal sections showing the FLAIR image with manual segmentation of Patient 43 (large lesion), along with results from the dual-modality model (middle, blue) and the single-modality model (bottom, red). The Dice scores for the dual-modality and single-modality (fine-tuned ensemble)d models for this patient are 0.81 and 0.60, respectively. . . . .                                                                                                                                                                                                    | 132 |
| A.3  | Sagittal, coronal, and axial FLAIR slices of Patient 18 (characterized by small lesions), showing the manual segmentation alongside ensemble predictions from the early-, mid-, and late-fusion models, with corresponding Dice scores of 0.69, 0.41, and 0.76, respectively. . . . .                                                                                                                                                                                                                                                                             | 133 |
| A.4  | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the Baseline model across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                                              | 134 |
| A.5  | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the T1 fine-tuned model across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                                         | 135 |
| A.6  | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble Early-fusion model across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                                 | 136 |
| A.7  | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble Mid-fusion model across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                                   | 137 |
| A.8  | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble Late-fusion model across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                                  | 138 |
| A.9  | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the Averaged models (Early fusion) across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                              | 139 |
| A.10 | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the Averaged models (Mid fusion) across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                                | 140 |
| A.11 | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the Averaged models (Late fusion) across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                               | 141 |
| A.12 | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the Averaged ensemble (T1+FLAIR Early, Mid, and Late fusion) models across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                             | 142 |
| A.13 | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble models (Early fusion) across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                              | 143 |
| A.14 | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble models (Mid fusion) across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                                | 144 |
| A.15 | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble models (Late fusion) across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                                               | 145 |
| A.16 | Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble models (T1+FLAIR Early, Mid, and Late fusion) across all lesion volume categories (All, Small, Medium, Large). . . . .                                                                                                                                                                                                                                                                                                                                                                      | 146 |

# List of Tables

|     |                                                                                                                                                                                                                                                                                        |     |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 2.1 | Comparison between Contactless, Wearable, and Marker-based Optical Motion Tracking Systems . . . . .                                                                                                                                                                                   | 23  |
| 4.1 | DTW distance (m) for each fingertip in X, Y, and Z position between DTH and each DTI for the "fist" (left) and "wrist supination" (right) gestures ( <b>Less is better</b> ) . . . . .                                                                                                 | 51  |
| 4.2 | DTW distance (m) for each fingertip in X, Y, and Z position between DTH and each DTI for the "move hand upward" (left) and "move hand downward" (right) gestures ( <b>Less is better</b> ) . . . . .                                                                                   | 51  |
| 5.1 | Voltage, current, and power consumption in two different scenarios: Baseline scenario (left) and Operating conditions (right). In the operating conditions, the use of a fan is also considered to maintain the operating temperature low. Averaged measurements are reported. . . . . | 59  |
| 5.2 | CPU usage (in %) and temperature (in °C), for both the considered scenarios. . . . .                                                                                                                                                                                                   | 61  |
| 6.1 | Comparison with other state-of-the-art methods, where Ops represents the number of operations for each gesture recognition inference on a sample of size $8 \times 3 \times 224 \times 224$ . . . . .                                                                                  | 68  |
| 6.2 | Experiments investigating the effects of varying network components, loss functions, and optimizers on accuracy. . . . .                                                                                                                                                               | 68  |
| 7.1 | The Private Data Elements pertain to the task that the patient has to carry on. . . . .                                                                                                                                                                                                | 77  |
| 7.2 | Private Data Elements specific for each repetition of the task. . . . .                                                                                                                                                                                                                | 77  |
| 7.3 | New Private Data Element in the Waveform Module pertains to the sampling time if the system employed does not perform frequency acquisition at a fixed time. . . . .                                                                                                                   | 78  |
| 8.1 | Patient count and lesion volume statistics for each dataset. Lesion volumes are reported in milliliters (mL) as mean $\pm$ standard deviation, median, and maximum values. . . . .                                                                                                     | 94  |
| 8.2 | Performance metrics across five folds on the ATLAS v2.0 dataset. Each cell reports the mean value, followed by the standard deviation in parentheses, and the median value. . . . .                                                                                                    | 94  |
| 8.3 | Stratified performance by lesion volume. Entries report mean values. Single (*) and double asterisks (**) denote the best and second-best performing models, respectively. . . . .                                                                                                     | 95  |
| 8.4 | Performance metrics on the Internal dataset. Metrics are reported as mean (standard deviation) and median. Single (*) and double asterisks (**) denote the best and second-best performing models, respectively. . . . .                                                               | 96  |
| A.1 | Statistical comparison (paired two-sided Wilcoxon signed-rank tests) for T1 fine-tuned models using voxel-wise Dice. Bold indicates $p < 0.05$ . . . . .                                                                                                                               | 133 |
| A.2 | Statistical comparison (paired two-sided Wilcoxon signed-rank tests) for early fusion models using voxel-wise Dice. Bold indicates $p < 0.05$ . . . . .                                                                                                                                | 147 |
| A.3 | Statistical comparison (paired two-sided Wilcoxon signed-rank tests) for mid fusion models using voxel-wise Dice. Bold indicates $p < 0.05$ . . . . .                                                                                                                                  | 147 |
| A.4 | Statistical comparison (paired two-sided Wilcoxon signed-rank tests) for late fusion models using voxel-wise Dice. Bold indicates $p < 0.05$ . . . . .                                                                                                                                 | 147 |
| A.5 | Statistical comparison (p-values) for non-fusion models (Lesion-wise F1-score). Bold indicates $p < 0.05$ . . . . .                                                                                                                                                                    | 148 |
| A.6 | Statistical comparison (p-values) for early fusion models (Lesion-wise F1-score). Bold indicates $p < 0.05$ . . . . .                                                                                                                                                                  | 148 |
| A.7 | Statistical comparison (p-values) for mid fusion models (Lesion-wise F1-score). Bold indicates $p < 0.05$ . . . . .                                                                                                                                                                    | 148 |

|                                                                                                                          |     |
|--------------------------------------------------------------------------------------------------------------------------|-----|
| A.8 Statistical comparison (p-values) for late fusion models (Lesion-wise F1-score). Bold indicates $p < 0.05$ . . . . . | 149 |
|--------------------------------------------------------------------------------------------------------------------------|-----|

# Acronyms

|              |                                                |
|--------------|------------------------------------------------|
| <b>ACR</b>   | American College of Radiology                  |
| <b>ADL</b>   | Activities of Daily Living                     |
| <b>AE</b>    | Application Entity                             |
| <b>AI</b>    | Artificial Intelligence                        |
| <b>API</b>   | Application Programming Interface              |
| <b>AR</b>    | Augmented Reality                              |
| <b>ARM</b>   | Advanced RISC Machine                          |
| <b>AS</b>    | Administrator System                           |
| <b>ASD</b>   | Average Surface Distance                       |
| <b>ASSD</b>  | Average Symmetric Surface Distance             |
| <b>ATLAS</b> | Anatomical Tracings of Lesions After Stroke    |
| <b>BIDS</b>  | Brain Imaging Data Structure                   |
| <b>BNN</b>   | Binary Neural Network                          |
| <b>C3D</b>   | Coordinate 3D                                  |
| <b>CAVE</b>  | Cave Automatic Virtual Environment             |
| <b>CNN</b>   | Convolutional Neural Network                   |
| <b>CPU</b>   | Central Processing Unit                        |
| <b>CRUD</b>  | create, read, update and delete                |
| <b>CSV</b>   | Comma-Separated Values                         |
| <b>CT</b>    | Computed Tomography                            |
| <b>CV</b>    | Computer Vision                                |
| <b>DDT</b>   | Double Digital Twin                            |
| <b>DICOM</b> | Digital Imaging and Communications in Medicine |
| <b>DIMSE</b> | DIcom Message Service Element                  |
| <b>DL</b>    | Deep Learning                                  |
| <b>DoF</b>   | Degrees of Freedom                             |
| <b>DSC</b>   | Dice Similarity Coefficient                    |
| <b>DT</b>    | Digital Twin                                   |
| <b>DTH</b>   | Digital Twin Healthy (Hand)                    |
| <b>DTI</b>   | Digital Twin Injured (Hand)                    |
| <b>DTW</b>   | Dynamic Time Warping                           |

**ECG** Electrocardiogram  
**EEG** Electroencephalography  
**EG** electromechanical glove  
**EHR** Electronic Health Record  
**EMG** Electromyography  
**EMR** Electronic Medical Record  
  
**FC** Fully Connected  
**FDR** False Discovery Rate  
**FLAIR** Fluid-Attenuated Inversion Recovery  
**FN** False Negative  
**FNR** False Negative Rate  
**FP** False Positive  
**FPGA** Field-Programmable Gate Array  
**FPS** Frames Per Second  
**FSR** Force Sensitive Resistor  
  
**GPU** Graphics Processing Unit  
**GRU** Gated Recurrent Unit  
**GT** Ground Truth  
**GUI** Graphical User Interface  
  
**HD** Hausdorff Distance  
**HDMI** High-Definition Multimedia Interface  
**HGR** Hand Gesture Recognition  
**HL7** Health Level 7  
**HMD** Head-Mounted Display  
**HMM** Hidden Markov Models  
  
**IE** Information Entity  
**IMU** Inertial Measurement Unit  
**IOD** Information Object Definition  
**IoU** Intersection over Union  
**IR** Infrared  
**ISO** International Organization for Standardization  
  
**JSON** Javascript Object Notation  
  
**KD** Knowledge Distillation  
**kNN** k-Nearest Neighbors  
  
**LED** Light Emitting Diode  
**LMC** Leap Motion Controller  
**LPS** Large Projection Screens



**LSTM** Long Short-Term Memory

**MLP** Multi-Layer Perceptron

**MRI** Magnetic Resonance Imaging

**MSE** Mean Squared Error

**NEMA** National Electrical Manufacturers Association

**nnU-Net** No-New-Net Framework

**OOP** Object Oriented Paradigm

**Ops** Operations

**OS** Operating System

**PACS** Picture Archiving and Communication System

**PDF** Portable Document Format

**PPV** Positive Predictive Value

**PR** Physical Rehabilitation

**PR-IOD** Physical Rehabilitation Information Object Definition

**PSA** primary-secondary approach

**QPR** Quantitative Physical Rehabilitation

**RAM** Random Access Memory

**RAS** Right-Anterior-Superior

**ReLU** Rectified Linear Unit

**RGB** Red, Green, Blue

**RGB-D** Red, Green, Blue, and Depth

**RMSD** Root Mean Squared Deviation

**rmsprop** Root Mean Square Propagation

**RNN** Recurrent Neural Network

**ROI** Region Of Interest

**RP** Raspberry Pi

**RQ** Research Question

**RS** Rehabilitation System

**SBC** Single-Board Computer

**SCP** Service Class Provider

**SCU** Service Class User

**SDO** Standards Development Organization

**SGD** Stochastic Gradient Descent

**SoC** System on a Chip

**SOP** Service Object Pair

**SQL** Structured Query Language

**SS** Server System

**STE** Straight-Through Estimator

**SVD** Singular Value Decomposition

**SVM** Support Vector Machine

**T1-w** T1-weighted

**TN** True Negative

**TP** True Positive

**TS** Therapist System

**UOL** Ultraleap Official Libraries

**USB** Universal Serial Bus

**UTS** UltraLeap Tracking Service

**VG** Virtual Glove

**VR** Virtual Reality

**WAN** Wide Area Network

**WCNR** World Congress of Neurorehabilitation

**WG** wearable glove

**Wi-Fi** Wireless Fidelity

**XML** eXtensible Markup Language

# Candidate's Publications

- [1] G. Placidi, **A. Di Matteo**, D. Lozzi, M. Polsinelli, E. Theodoridou, “Patient–therapist cooperative hand telerehabilitation through a novel framework involving the virtual glove system”, *Sensors*, vol. 23, no. 7, pp. 3463, 2023. DOI: 10.3390/s23073463
- [2] E. Theodoridou, A. Cacchio, **A. Di Matteo**, M. Polsinelli, and G. Placidi, “Materials and techniques for effective at-home rehabilitation for hand mobility restoration based on the Virtual Glove,” in *Proceedings of the 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2023, pp. 579–584. DOI: 10.1109/CBMS58004.2023.00282.
- [3] A. Bianchi, V. Zelli, A. D’Angelo, **A. Di Matteo**, G. Scoccia, K. Cannita, A. S. Dimas, S. Glentis, F. Zazzeroni, E. Alesse, *et al.*, “A method to comprehensively identify germline SNVs, INDELs and CNVs from whole exome sequencing data of BRCA1/2 negative breast cancer patients” *NAR Genomics and Bioinformatics*, vol. 6, no. 2, p. lqae033, 2024. Oxford University Press, DOI: 10.1093/nargab/lqae033
- [4] M. Polsinelli, **A. Di Matteo**, D. Lozzi, E. Mattei, F. Mignosi, L. Nazzicone, V. Stornelli, G. Placidi, “Portable Head-Mounted System for Mobile Forearm Tracking”, *Sensors*, vol. 24, no. 7, pp. 2227, 2024. DOI: 10.3390/s24072227
- [5] **A. Di Matteo**, D. Lozzi, E. Mattei, F. Mignosi, S. Montagna, M. Polsinelli, G. Placidi, “Calibration of the Double Digital Twin for the Hand Rehabilitation by the Virtual Glove”, *2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 521–526, 2024. DOI: 10.1109/CBMS61543.2024.00092
- [6] E. Mattei, D. Lozzi, **A. Di Matteo**, M. Polsinelli, C. Manes, F. Mignosi, G. Placidi, “Deep Learning Architecture analysis for EEG-Based BCI Classification under Motor Execution”, *2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 549–555, 2024. DOI: 10.1109/CBMS61543.2024.00097
- [7] F. Fontana, **A. Di Matteo**, L. Cinque, G. Placidi, and M. R. Marini, “BNNAction-Net: Binary Neural Network on Hands Gesture Recognition” in *ACM SIGGRAPH 2024 Posters*, pp. 1–2, 2024. DOI: /10.1145/3641234.3671047
- [8] E. Mattei, D. Lozzi, **A. Di Matteo**, A. Cipriani, C. Manes, G. Placidi, “MOVING: A Multi-Modal Dataset of EEG Signals and Virtual Glove Hand Tracking”, *Sensors*, vol. 24, no. 16, pp. 5207, 2024. DOI: 10.3390/s24165207
- [9] E. Mattei, D. Lozzi, **A. Di Matteo**, C. Manes, F. Mignosi, M. Polsinelli, G. Placidi, “Analysis of the EEG Resting-State Signals for BCI”, *9th Graz Brain-Computer Interface Conference 2024*, pp. 355–359, 2024. DOI: 10.3217/978-3-99161-014-4-062
- [10] D. Lozzi, E. Mattei, R. Ciuffini, **A. Di Matteo**, A. Marrelli, R. Ornello, M. Polsinelli, C. Rosignoli, S. Sacco, G. Placidi, “The challenge of driving BCI with emotional signals collected by EEG”, *9th Graz Brain-Computer Interface Conference 2024*, pp. 366–371, 2024. DOI: 10.3217/978-3-99161-014-4-064

- [11] G. D'Aloisio, **A. Di Matteo**, A. Cipriani, D. Lozzi, E. Mattei, G. Zanfardino, A. Di Marco, G. Placidi, "Engineering a Digital Twin for Diagnosis and Treatment of Multiple Sclerosis", in *International Conference on Engineering Digital Twins (EDTconf 2024)*, 2024. DOI: 10.1145/3652620.3688249
- [12] **A. Di Matteo**, D. Lozzi, E. Mattei, S. Montagna, A. Cipriani, F. Mignosi, M. Polsinelli, G. Placidi, "Proactive Hand Rehabilitation Through a Double Digital Twin Provided by the Virtual Glove", in *IEEE MetroXRaine 2024 - Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering*, 2024. DOI: 10.1109/MetroXRaine62247.2024.10796497.
- [13] D. Lozzi, **A. Di Matteo**, E. Mattei, A. Cipriani, P. Caianiello, F. Mignosi, G. Placidi, "ASIS: a smart alarm clock based on deep learning for the safety of night workers", in *IEEE MetroXRaine 2024 - Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering*, 2024. DOI: 10.1109/MetroXRaine62247.2024.10796738.
- [14] F. D'Amato, A. Cipriani, **A. Di Matteo**, D. Lozzi, E. Mattei, M. Polsinelli, G. Placidi, "A CycleGAN Model to Synthesize Missing and Unpaired MRI Sequences for Under-Represented Multiple Sclerosis Lesions", in *The 18th European Conference on Computer Vision (ECCV 2024)*, 2024. DOI: 10.1007/978-3-031-91907-7\_16.
- [15] **A. Di Matteo**, D. Lozzi, F. Mignosi, M. Polsinelli, G. Placidi, "A DICOM-based Standard for Quantitative Physical Rehabilitation", in *Computational and Structural Biotechnology* 2025. DOI: 10.1016/j.csbj.2025.01.012
- [16] **A. Di Matteo**, Y. Mahé, S. Leplaideur, I. Bonan, E. Bannier, F. Galassi, "Deep Learning and Multi-Modal MRI for the Segmentation of Sub-Acute and Chronic Stroke Lesions", in *Pattern Recognition Letters* 2025. DOI: <https://doi.org/10.1016/j.patrec.2025.11.017>

# Bibliography

- [1] Valery L Feigin, Michael Brainin, Bo Norrving, Sheila O Martins, Jeyaraj Pandian, Patrice Lindsay, Maria F Grupper, and Ilari Rautalin. World stroke organization: global stroke fact sheet 2025. *International Journal of Stroke*, 20(2):132–144, 2025.
- [2] World Health Organization – EMRO. Stroke, cerebrovascular accident: Introduction. <https://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/introduction.html>, n.d. Accessed: 2025-09-30.
- [3] Charles DA Wolfe. The impact of stroke. *British medical bulletin*, 56(2):275–286, 2000.
- [4] Cheryl Bushnell, Walter N Kernan, Anjail Z Sharrief, Seemant Chaturvedi, John W Cole, William K Cornwell III, Christine Cosby-Gaither, Sarah Doyle, Larry B Goldstein, Olive Lennon, et al. 2024 guideline for the primary prevention of stroke: a guideline from the american heart association/american stroke association. *Stroke*, 55(12):e344–e424, 2024.
- [5] Xiaohong Li, Yanjin He, Dawu Wang, and Mohammad J Rezaei. Stroke rehabilitation: from diagnosis to therapy. *Frontiers in neurology*, 15:1402729, 2024.
- [6] Alessandro Di Matteo, Youwan Mahé, Stéphanie s Leplaideur, Isabelle Bonan, Elise Bannier, and Francesca Galassi. Deep learning and multi-modal mri for the segmentation of sub-acute and chronic stroke lesions. 2025.
- [7] Jialin Luo, Peishan Dai, Zhuang He, Zhongchao Huang, Shenghui Liao, and Kun Liu. Deep learning models for ischemic stroke lesion segmentation in medical images: A survey. *Computers in biology and medicine*, 175:108509, 2024.
- [8] Hae Sol Moon, Lindsay Heffron, Ali Mahzarnia, Barnabas Obeng-Gyasi, Matthew Holbrook, Cristian T Badea, Wuwei Feng, and Alexandra Badea. Automated multimodal segmentation of acute ischemic stroke lesions on clinical mr images. *Magnetic resonance imaging*, 92:45–57, 2022.
- [9] K. K. Wong et al. Automatic segmentation in acute ischemic stroke: Prognostic significance of topological stroke volumes on stroke outcome. *Stroke*, 53:2896–2905, 2022. doi: 10.1161/STROKEAHA.121.037982.
- [10] M. R. Hernandez Petzsche et al. Isles 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Sci Data*, 9:762, 2022. doi: 10.1038/s41597-022-01875-5.
- [11] Parisa Soleimani and Navid Farezi. Utilizing deep learning via the 3d u-net neural network for the delineation of brain stroke lesions in mri image. *Scientific Reports*, 13(1):19808, 2023.
- [12] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

- 
- [13] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.
  - [14] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.
  - [15] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2): 251–257, 1991.
  - [16] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
  - [17] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
  - [18] The MathWorks, Inc. Convolutional neural network. <https://it.mathworks.com/discovery/convolutional-neural-network.html>, 2025. Accesso: 30 settembre 2025.
  - [19] O. Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. Preprint, 2015.
  - [20] F. Isensee et al. nnu-net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods*, 18:203–211, 2021. doi: 10.1038/s41592-020-01008-z.
  - [21] O. Commowick et al. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Sci Rep*, 8:13650, 2018. doi: 10.1038/s41598-018-31911-7.
  - [22] Carolee J Winstein, Joel Stein, Ross Arena, Barbara Bates, Leora R Cherney, Steven C Cramer, Frank Deruyter, Janice J Eng, Beth Fisher, Richard L Harvey, et al. Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the american heart association/american stroke association. *Stroke*, 47(6):e98–e169, 2016.
  - [23] Sandeep Dey, Aayush Arya, Arya J Raut, Shreyas Katta, and Prashant Sharma. Exploring the role of neuroplasticity in stroke rehabilitation: Mechanisms, interventions and clinical implications. *International Journal of Forensic Medicine and Related Sciences (IJFMR)*, 6(2):18483, 2024.
  - [24] Yu Chen, Kingsley Travis Abel, John T Janecek, Yunan Chen, Kai Zheng, and Steven C Cramer. Home-based technologies for stroke rehabilitation: A systematic review. *International journal of medical informatics*, 123:11–22, 2019.
  - [25] Bappaditya Debnath, Mary O’Brien, Motonori Yamaguchi, and Ardhendu Behera. A review of computer vision-based approaches for physical rehabilitation and assessment. *Multimedia Systems*, 28(1): 209–239, 2022.
  - [26] Erkan Ödemiş and Cabbar Veysel Baysal. Development of a participation assessment system based on multimodal evaluation of user responses for upper limb rehabilitation. *Biomedical Signal Processing and Control*, 70:103066, 2021.
  - [27] Cassandra Gorman and Louise Gustafsson. The use of augmented reality for rehabilitation after stroke: a narrative review. *Disability and rehabilitation: assistive technology*, 17(4):409–417, 2022.
  - [28] Kurt D Knepley, Jennifer Z Mao, Peter Wiczorek, Frederick O Okoye, Abhi P Jain, and Noam Y Harel. Impact of telerehabilitation for stroke-related deficits. *Telemedicine and e-Health*, 27(3):239–246, 2021.
  - [29] Karnica Manivasagam and Liyun Yang. Evaluation of a new simplified inertial sensor method against electrogoniometer for measuring wrist motion in occupational studies. *Sensors*, 22(4):1690, 2022.

- 
- [30] Margaret A French, Ryan T Roemmich, Kelly Daley, Meghan Beier, Sharon Penttinen, Preeti Raghavan, Peter Searson, Stephen Wegener, and Pablo Celnik. Precision rehabilitation: optimizing function, adding value to health care. *Archives of physical medicine and rehabilitation*, 103(6):1233–1239, 2022.
- [31] Sejuti Rahman, Sujun Sarker, AKM Nadimul Haque, Monisha Mushtary Uttsha, Md Fokhrul Islam, and Swakshar Deb. Ai-driven stroke rehabilitation systems and assessment: A systematic review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:192–207, 2022.
- [32] Myunghee Kim, Hyeongkeun Jeong, Praktyath Kantharaju, Dongyual Yoo, Michael Jacobson, Dongbin Shin, Changsoo Han, and James L Patton. Visual guidance can help with the use of a robotic exoskeleton during human walking. *Scientific Reports*, 12(1):3881, 2022.
- [33] Anupam Gupta, Naveen B Prakash, and Preethi R Honavar. Gait training with robotic exoskeleton assisted rehabilitation system in patients with incomplete traumatic and non-traumatic spinal cord injury: A pilot study and review of literature. *Annals of Indian Academy of Neurology*, 26(Suppl 1): S26–S31, 2023.
- [34] Patrizia Poli, Giovanni Morone, Giulio Rosati, and Stefano Masiero. Robotic technologies and rehabilitation: new tools for stroke patients’ therapy. *BioMed research international*, 2013(1):153872, 2013.
- [35] Aylar Akbari, Faezeh Haghverd, and Saeed Behbahani. Robotic home-based rehabilitation systems design: from a literature review to a conceptual framework for community-based remote therapy during covid-19 pandemic. *Frontiers in Robotics and AI*, 8:612331, 2021.
- [36] Angela Arntz, Franziska Weber, Marietta Handgraaf, Kaisa Lällä, Katariina Korniloff, Kari-Pekka Murtonen, Julija Chichaeva, Anita Kidritsch, Mario Heller, Evanthia Sakellari, et al. Technologies in home-based digital rehabilitation: scoping review. *JMIR rehabilitation and assistive technologies*, 10(1):e43615, 2023.
- [37] Neville Hogan, Hermano I Krebs, Andre Sharon, and Jain Charnnarong. Interactive robotic therapist, November 14 1995. US Patent 5,466,213.
- [38] Hermano I Krebs, Mark Ferraro, Stephen P Buerger, Miranda J Newbery, Antonio Makiyama, Michael Sandmann, Daniel Lynch, Bruce T Volpe, and Neville Hogan. Rehabilitation robotics: pilot trial of a spatial extension for mit-manus. *Journal of neuroengineering and rehabilitation*, 1(1):5, 2004.
- [39] Helen Rodgers, Helen Bosomworth, Hermano I Krebs, Frederike van Wijck, Denise Howel, Nina Wilson, Lydia Aird, Natasha Alvarado, Sreeman Andole, David L Cohen, et al. Robot assisted training for the upper limb after stroke (ratuls): a multicentre randomised controlled trial. *The Lancet*, 394(10192): 51–62, 2019.
- [40] Tobias Nef and Robert Riener. Armin-design of a novel arm rehabilitation robot. In *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.*, pages 57–60. IEEE, 2005.
- [41] Takashi Takebayashi, Yuki Uchiyama, and Kazuhisa Domen. Automatic setting optimization for robotic upper-extremity rehabilitation in patients with stroke using reogo-j: a cross-sectional clinical trial. *Scientific Reports*, 14(1):25710, 2024.
- [42] Stefan Hesse, Henning Schmidt, Cordula Werner, et al. Machines to support motor rehabilitation after stroke: 10 years of experience in berlin. *Journal of rehabilitation research and development*, 43(5):671, 2006.

- 
- [43] RJ Sanchez, Eric Wolbrecht, R Smith, J Liu, S Rao, S Cramer, T Rahman, James E Bobrow, and David J Reinkensmeyer. A pneumatic robot for re-training arm movement after stroke: Rationale and mechanical design. In *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.*, pages 500–504. IEEE, 2005.
  - [44] Albert C Lo, Peter Guarino, Hermano I Krebs, Bruce T Volpe, Christopher T Bever, Pamela W Duncan, Robert J Ringer, Todd H Wagner, Lorie G Richards, Dawn M Bravata, et al. Multicenter randomized trial of robot-assisted rehabilitation for chronic stroke: methods and entry characteristics for va robotics. *Neurorehabilitation and neural repair*, 23(8):775–783, 2009.
  - [45] Albert C Lo, Peter D Guarino, Lorie G Richards, Jodie K Haselkorn, George F Wittenberg, Daniel G Federman, Robert J Ringer, Todd H Wagner, Hermano I Krebs, Bruce T Volpe, et al. Robot-assisted therapy for long-term upper-limb impairment after stroke. *New England Journal of Medicine*, 362(19):1772–1783, 2010.
  - [46] Ammar Khan, Yahia Z Imam, Mohamed Muneer, Salman Al Jerdi, and Sumanjit K Gill. Virtual reality in stroke recovery: a meta-review of systematic reviews. *Bioelectronic Medicine*, 10(1):23, 2024.
  - [47] Rhutuja Khokale, Grace S Mathew, Somi Ahmed, Sara Maheen, Moiz Fawad, Prabhudas Bandaru, Annu Zerín, Zahra Nazir, Imran Khawaja, Imtenan Sharif, et al. Virtual and augmented reality in post-stroke rehabilitation: a narrative review. *Cureus*, 15(4), 2023.
  - [48] Ana Lúcia Faria, Andreia Andrade, Luísa Soares, and Sergi Bermúdez i Badia. Benefits of virtual reality based cognitive rehabilitation through simulated activities of daily living: a randomized controlled trial with stroke patients. *Journal of neuroengineering and rehabilitation*, 13(1):96, 2016.
  - [49] Maria Grazia Maggio, Lilla Bonanno, Amelia Rizzo, Martina Barbera, Alessandra Benenati, Federica Impellizzeri, Francesco Corallo, Rosaria De Luca, Angelo Quartarone, and Rocco Salvatore Calabrò. The role of virtual reality-based cognitive training in enhancing motivation and cognitive functions in individuals with chronic stroke. *Scientific Reports*, 15(1):25258, 2025.
  - [50] Andrea Demeco, Laura Zola, Antonio Frizziero, Chiara Martini, Arrigo Palumbo, Ruben Foresti, Giovanni Buccino, and Cosimo Costantino. Immersive virtual reality in post-stroke rehabilitation: a systematic review. *Sensors*, 23(3):1712, 2023.
  - [51] Sebastian Oltedal Thorp, Lars Morten Rimol, Simon Lervik, Hallvard Røe Evensmoen, and Simone Grassini. Comparative analysis of spatial ability in immersive and non-immersive virtual reality: The role of sense of presence, simulation sickness and cognitive load. *Frontiers in Virtual Reality*, 5:1343872, 2024.
  - [52] Movella Inc. Xsens metagloves by manus. <https://www.movella.com/motion-capture/xsens-met-agloves-by-manus>, 2025. Accessed: 2025-11-03.
  - [53] Adriana Daniela Banyai and Cornel Brişan. Robotics in physical rehabilitation: Systematic review. In *Healthcare*, volume 12, page 1720. MDPI, 2024.
  - [54] Giuseppe Placidi, Luigi Cinque, Matteo Polsinelli, and Matteo Spezialetti. Measurements by a leap-based virtual glove for the hand rehabilitation. *Sensors*, 18(3):834, 2018.
  - [55] Giuseppe Placidi, Danilo Avola, Luigi Cinque, Matteo Polsinelli, Eleni Theodoridou, and João Manuel RS Tavares. Data integration by two-sensors in a leap-based virtual glove for human-system interaction. *Multimedia Tools and Applications*, 80(12):18263–18277, 2021.



- 
- [56] Emilia Mikołajewska, Jolanta Masiak, and Dariusz Mikołajewski. Applications of artificial intelligence-based patient digital twins in decision support in rehabilitation and physical therapy. *Electronics*, 13(24):4994, 2024.
  - [57] Martin W Lauer-Schmaltz, Philip Cash, John P Hansen, and Neha Das. Human digital twins in rehabilitation: a case study on exoskeleton and serious-game-based stroke rehabilitation using the ethica methodology. *IEEE Access*, 2024.
  - [58] Yilun Chen, Wentao Wang, Junyu Diao, Daoyu Wang, Zhuo Jian, Yixi Wang, and Zhihao Jiang. Digital-twin-based patient evaluation during stroke rehabilitation. In *Proceedings of the ACM/IEEE 14th International Conference on Cyber-Physical Systems (with CPS-IoT Week 2023)*, pages 22–33, 2023.
  - [59] Md Mijanur Rahman, Ashik Uzzaman, Fatema Khatun, Md Aktaruzzaman, and Nazmul Siddique. A comparative study of advanced technologies and methods in hand gesture analysis and recognition systems. *Expert Systems with Applications*, 266:125929, 2025.
  - [60] Munir Oudah, Ali Al-Naji, and Javaan Chahl. Hand gesture recognition based on computer vision: a review of techniques. *journal of Imaging*, 6(8):73, 2020.
  - [61] Matthieu Courbariaux, Itay Hubara, Daniel Soudry, Ran El-Yaniv, and Yoshua Bengio. Binarized neural networks: Training deep neural networks with weights and activations constrained to+ 1 or-1. *arXiv preprint arXiv:1602.02830*, 2016.
  - [62] Itay Hubara, Matthieu Courbariaux, Daniel Soudry, Ran El-Yaniv, and Yoshua Bengio. Binarized neural networks. *Advances in neural information processing systems*, 29, 2016.
  - [63] Taylor Simons and Dah-Jye Lee. A review of binarized neural networks. *Electronics*, 8(6):661, 2019.
  - [64] Giuseppe Placidi, Alessandro Di Matteo, Filippo Mignosi, Matteo Polsinelli, and Matteo Spezialetti. Compact, accurate and low-cost hand tracking system based on leap motion controllers and raspberry pi. In *ICPRAM*, pages 652–659, 2022.
  - [65] Eleni Theodoridou, Luigi Cinque, Filippo Mignosi, Giuseppe Placidi, Matteo Polsinelli, João Manuel RS Tavares, and Matteo Spezialetti. Hand tracking and gesture recognition by multiple contactless sensors: a survey. *IEEE Transactions on Human-Machine Systems*, 2022.
  - [66] Giuseppe Placidi. A smart virtual glove for the hand telerehabilitation. *Computers in Biology and Medicine*, 37(8):1100–1107, 2007.
  - [67] Giuseppe Placidi, Danilo Avola, Daniela Iacoviello, and Luigi Cinque. Overall design and implementation of the virtual glove. *Computers in biology and medicine*, 43(11):1927–1940, 2013.
  - [68] Giuseppe Placidi, Luigi Cinque, Andrea Petracca, Matteo Polsinelli, and Matteo Spezialetti. A virtual glove system for the hand rehabilitation based on two orthogonal leap motion controllers. In *ICPRAM*, pages 184–192, 2017.
  - [69] Salih Ertug Ovur, Hang Su, Wen Qi, Elena De Momi, and Giancarlo Ferrigno. Novel adaptive sensor fusion methodology for hand pose estimation with multileap motion. *IEEE Transactions on Instrumentation and Measurement*, 70:1–8, 2021.
  - [70] Andrew Houston, Vanessa Walters, Thomas Corbett, and Russell Coppack. Evaluation of a multi-sensor leap motion setup for biomechanical motion capture of the hand. *Journal of Biomechanics*, page 110713, 2021.
-

- 
- [71] Tomas Novacek, Christian Marty, and Marcel Jirina. Project multileap: Fusing data from multiple leap motion sensors. In *2021 IEEE 7th International Conference on Virtual Reality (ICVR)*, pages 19–25. IEEE, 2021.
- [72] Jože Guna, Grega Jakus, Matevž Pogačnik, Sašo Tomažič, and Jaka Sodnik. An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. *Sensors*, 14(2):3702–3720, 2014.
- [73] Ela Tarakci, Nilay Arman, Devrim Tarakci, and Ozgur Kasapcopur. Leap motion controller-based training for upper extremity rehabilitation in children and adolescents with physical disabilities: A randomized controlled trial. *Journal of Hand Therapy*, 33(2):220–228, 2020.
- [74] Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, and Umapada Pal. A lexicon-free approach for 3d handwriting recognition using classifier combination. *Pattern Recognition Letters*, 103:1–7, 2018.
- [75] Kostas Nizamis, Noortje Rijken, Ana Mendes, Mariska Janssen, Arjen Bergsma, and Bart Koopman. A novel setup and protocol to measure the range of motion of the wrist and the hand. *Sensors*, 18(10):3230, September 2018. doi: 10.3390/s18103230. URL <https://doi.org/10.3390/s18103230>.
- [76] Rogério S. Gonçalves, Marcus R. S. B. de Souza, and Giuseppe Carbone. Analysis of the leap motion controller’s performance in measuring wrist rehabilitation tasks using an industrial robot arm reference. *Sensors*, 22(13):4880, June 2022. doi: 10.3390/s22134880. URL <https://doi.org/10.3390/s22134880>.
- [77] Daniel Bachmann, Frank Weichert, and Gerhard Rinkenauer. Evaluation of the leap motion controller as a new contact-free pointing device. *Sensors*, 15(1):214–233, 2014.
- [78] Andrew Houston, Vanessa Walters, Thomas Corbett, and Russell Coppack. Evaluation of a multi-sensor leap motion setup for biomechanical motion capture of the hand. *Journal of Biomechanics*, 127:110713, October 2021. doi: 10.1016/j.jbiomech.2021.110713. URL <https://doi.org/10.1016/j.jbiomech.2021.110713>.
- [79] Danilo Avola, Matteo Spezialetti, and Giuseppe Placidi. Design of an efficient framework for fast prototyping of customized human–computer interfaces and virtual environments for rehabilitation. *Computer Methods and Programs in Biomedicine*, 110(3):490–502, June 2013. doi: 10.1016/j.cmpb.2013.01.009. URL <https://doi.org/10.1016/j.cmpb.2013.01.009>.
- [80] A Petracca, M Carrieri, D Avola, S Basso Moro, S Brigadoi, S Lancia, M Spezialetti, M Ferrari, V Quaresima, and G Placidi. A virtual ball task driven by forearm movements for neuro-rehabilitation. In *2015 International Conference on Virtual Rehabilitation (ICVR)*. IEEE, June 2015. doi: 10.1109/icvr.2015.7358600. URL <https://doi.org/10.1109/icvr.2015.7358600>.
- [81] Roberto Minerva, Gyu Myoung Lee, and Noël Crespi. Digital twin in the iot context: A survey on technical features, scenarios, and architectural models. *Proceedings of the IEEE*, 108(10):1785–1824, 2020.
- [82] Michael Grieves and John Vickers. *Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems*, pages 85–113. Springer International Publishing, Cham, 2017.
- [83] Evangelia Katsoulakis, Qi Wang, Huanmei Wu, Leili Shahriyari, Richard Fletcher, Jinwei Liu, Luke Achenie, Hongfang Liu, Pamela Jackson, Ying Xiao, Tanveer Syeda-Mahmood, Richard Tuli, and Jun Deng. Digital twins for health: a scoping review. *npj Digital Medicine*, 7(77):1–11, 2024.

- 
- [84] Alessandro Ricci, Angelo Croatti, and Sara Montagna. Pervasive and connected digital twins—a vision for digital health. *IEEE Internet Computing*, 26(5):26–32, 2022.
- [85] Genevieve Coorey, Gemma A. Figtree, David F. Fletcher, Victoria J. Snelson, Stephen Thomas Vernon, David Winlaw, Stuart M. Grieve, Alistair McEwan, Jean Yee Hwa Yang, Pierre Qian, Kieran O’Brien, Jessica Orchard, Jinman Kim, Sanjay Patel, and Julie Redfern. The health digital twin to tackle cardiovascular disease—a review of an emerging interdisciplinary field. *npj Digital Medicine*, 5(1):126, 2022.
- [86] Patrizio Armeni, Irem Polat, Leonardo Maria De Rossi, Lorenzo Diaferia, Severino Meregalli, and Anna Gatti. Digital twins in healthcare: Is it the beginning of a new era of evidence-based medicine? a critical review. *Journal of personalized medicine.*, 12(8), 2022-8-30. ISSN 2075-4426.
- [87] Genevieve Coorey, Gemma A. Figtree, David F. Fletcher, Victoria J. Snelson, Stephen Thomas Vernon, David Winlaw, Stuart M. Grieve, Alistair McEwan, Jean Yee Hwa Yang, Pierre Qian, Kieran O’Brien, Jessica Orchard, Jinman Kim, Sanjay Patel, and Julie Redfern. The health digital twin to tackle cardiovascular disease—a review of an emerging interdisciplinary field. *npj Digital Medicine*, 5(1):126, 2022.
- [88] Bergthor Björnsson, Carl Borrebaeck, Nils Elander, Thomas Gasslander, Danuta R. Gawel, Mika Gustafsson, Rebecka Jörnsten, Eun Jung Lee, Xinxiu Li, Sandra Lilja, David Martínez-Enguita, Andreas Matussek, Per Sandström, Samuel Schäfer, Margaretha Stenmarker, X. F. Sun, Oleg Sysoev, Huan Zhang, Mikael Benson, and on behalf of the Swedish Digital Twin Consortium. Digital twins to personalize medicine. *Genome Medicine*, 12(1):4, 2019.
- [89] M. W. Lauer-Schmaltz, P. Cash, J. P. Hansen, and A. Maier. Designing human digital twins for behaviour-changing therapy and rehabilitation: A systematic review. *Proceedings of the Design Society*, 2:1303–1312, 2022. doi: 10.1017/pds.2022.132.
- [90] Safa Elkefi and Onur Asan. Digital twins for managing health care systems: Rapid literature review. *J Med Internet Res*, 24(8):e37641, Aug 2022. ISSN 1438-8871. doi: 10.2196/37641.
- [91] Md. Mahafuzur Rahaman Khan, Md. Samiul Haque Sunny, Tanvir Ahmed, Md. Tanzil Shahria, Preet Parag Modi, Md. Ishrak Islam Zarif, Javier Dario Sanjuan De Caro, Sheikh Iqbal Ahamed, Helal Uddin Ahmed, Erin M. McGonigle Ketchum, and Mohammad H. Rahman. Development of a robot-assisted telerehabilitation system with integrated iiot and digital twin. *IEEE Access*, 11:70174–70189, 2023. doi: 10.1109/ACCESS.2023.3291803.
- [92] Alessio Cellupica, Marco Cirelli, Giovanni Saggio, Emanuele Gruppioni, and Pier Paolo Valentini. An interactive digital-twin model for virtual reality environments to train in the use of a sensorized upper-limb prosthesis. *Algorithms*, 17(1):35, 2024.
- [93] Margarida F Pereira, Cosima Prahm, Jonas Kolbensschlag, Eva Oliveira, and Nuno F Rodrigues. Application of ar and vr in hand rehabilitation: A systematic review. *Journal of Biomedical Informatics*, 111:103584, 2020.
- [94] Unity Technologies. Unity - game engine. <https://unity.com/>.
- [95] Giuseppe Placidi, Alessandro Di Matteo, Daniele Lozzi, Matteo Polsinelli, and Eleni Theodoridou. Patient–therapist cooperative hand telerehabilitation through a novel framework involving the virtual glove system. *Sensors*, 23(7):3463, 2023.
- [96] Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(5):922–923, 1976.

- 
- [97] Wolfgang Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 34(5):827–828, 1978.
  - [98] MongoDB, Inc. Mongoddb homepage, 2023. URL <https://www.mongodb.com/>.
  - [99] Python Software Foundation. Python language reference, version 3.x, 2023. URL <https://www.python.org>.
  - [100] Plotly Technologies Inc. Collaborative data science, 2015. URL <https://plot.ly>.
  - [101] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.
  - [102] Romain Tavenard, Johann Faouzi, Gilles Vandewiele, Felix Divo, Guillaume Androz, Chester Holtz, Marie Payne, Roman Yurchak, Marc Rußwurm, Kushal Kolar, et al. Tslearn, a machine learning toolkit for time series data. *Journal of machine learning research*, 21(118):1–6, 2020.
  - [103] Jan Hendrik Hammer and Jürgen Beyerer. Robust hand tracking in realtime using a single head-mounted rgb camera. In *Human-Computer Interaction. Interaction Modalities and Techniques: 15th International Conference, HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part IV 15*, pages 252–261. Springer, 2013.
  - [104] Li Sun, Guizhong Liu, and Yiqing Liu. 3d hand tracking with head mounted gaze-directed camera. *IEEE Sensors Journal*, 14(5):1380–1390, 2013.
  - [105] Oytun Akman, Ronald Poelman, Wouter Caarls, and Pieter Jonker. Multi-cue hand detection and tracking for a head-mounted augmented reality system. *Machine vision and applications*, 24:931–946, 2013.
  - [106] Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5):6380–6393, 2013.
  - [107] Ana de Los Reyes-Guzmán, Vicente Lozano-Berrio, Maria Alvarez-Rodriguez, Elisa Lopez-Dolado, Silvia Ceruelo-Abajo, Francisco Talavera-Diaz, and Angel Gil-Agudo. Rehabhand: Oriented-tasks serious games for upper limb rehabilitation by using leap motion controller and target population in spinal cord injury. *NeuroRehabilitation*, 48(3):365–373, 2021.
  - [108] Hajar Sharif, Ahmadreza Eslaminia, Pramod Chembrammell, and Thenkurussi Kesavadas. Classification of activities of daily living based on grasp dynamics obtained from a leap motion controller. *Sensors*, 22(21):8273, 2022.
  - [109] Rafael Martins and Pollyana Notargiacomo. Evaluation of leap motion controller effectiveness on 2d game environments using usability heuristics. *Multimedia Tools and Applications*, 80:5539–5557, 2021.
  - [110] Shahin Fereidouni, Mohsen Sheikh Hassani, Alireza Talebi, and Amir Hossein Rezaie. A novel design and implementation of wheelchair navigation system using leap motion sensor. *Disability and Rehabilitation: Assistive Technology*, 17(4):442–448, 2022.
  - [111] Ing-Jr Ding and Meng-Chuan Hsieh. A hand gesture action-based emotion recognition system by 3d image sensor information derived from leap motion sensors for the specific group with restlessness emotion problems. *Microsystem Technologies*, 28(1):403–415, 2022.
  - [112] Basma Hisham and Alaa Hamouda. Arabic sign language recognition using ada-boosting based on a leap motion controller. *International Journal of Information Technology*, 13:1221–1234, 2021.
-

- 
- [113] Summer Lindsey. Evaluation of low cost controllers for mobile based virtual reality headsets. 2017.
  - [114] Kenneth R Moser and J Edward Swan. Evaluation of user-centric optical see-through head-mounted display calibration using a leap motion controller. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 159–167. IEEE, 2016.
  - [115] Trinette Wright, Sandrine De Ribaupierre, and Roy Eagleson. Design and evaluation of an augmented reality simulator using leap motion. *Healthcare technology letters*, 4(5):210–215, 2017.
  - [116] Pu-Hsuan Chien, Yang-Cheng Lin, et al. Gesture-based head-mounted augmented reality game development using leap motion and usability evaluation. In *15th International Conference on Interfaces and Human Computer Interaction, IHCI 2021 and 14th International Conference on Game and Entertainment Technologies, GET*, pages 149–56, 2021.
  - [117] Hui Zhang. Head-mounted display-based intuitive virtual reality training system for the mining industry. *International Journal of Mining Science and Technology*, 27(4):717–722, 2017.
  - [118] Elisa Gusai, Chiara Bassano, Fabio Solari, and Manuela Chessa. Interaction in an immersive collaborative virtual reality environment: A comparison between leap motion and htc controllers. In *New Trends in Image Analysis and Processing-ICIAP 2017: ICIAP International Workshops, WBICV, SSPandBE, 3AS, RGBD, NIVAR, IWBAAS, and MADiMa 2017, Catania, Italy, September 11-15, 2017, Revised Selected Papers 19*, pages 290–300. Springer, 2017.
  - [119] Alexander Masurovsky, Paul Chojecki, Detlef Runde, Mustafa Lafci, David Przewozny, and Michael Gaebler. Controller-free hand tracking for grab-and-place tasks in immersive virtual reality: Design elements and their empirical study. *multimodal technologies and interaction* 4, 4 (2020), 2020.
  - [120] Dorota BELANOVÁ and Kaori YOSHIDA. Hand position tracking correction of leap motion controller attached to the virtual reality headset. *International Journal of Biomedical Soft Computing and Human Sciences: the official journal of the Biomedical Fuzzy Systems Association*, 25(1):29–37, 2020.
  - [121] Microsoft Corporation. Microsoft hololens 2. <https://www.microsoft.com/en-us/hololens>, 2024. Accessed: 2024-01-10.
  - [122] Siwei Zhang, Qianli Ma, Yan Zhang, Zhiyin Qian, Taein Kwon, Marc Pollefeys, Federica Bogo, and Siyu Tang. Egobody: Human body shape and motion of interacting people from head-mounted devices. In *European Conference on Computer Vision*, pages 180–200. Springer, 2022.
  - [123] Mitchell Doughty and Nilesh R Ghugre. Hmd-egopose: Head-mounted display-based egocentric marker-less tool and hand pose estimation for augmented surgical guidance. *International Journal of Computer Assisted Radiology and Surgery*, 17(12):2253–2262, 2022.
  - [124] Alexander Schäfer, Gerd Reis, and Didier Stricker. Controlling continuous locomotion in virtual reality with bare hands using hand gestures. In *International Conference on Virtual Reality and Mixed Reality*, pages 191–205. Springer, 2022.
  - [125] Diar Abdulkarim, Massimiliano Di Luca, Poppy Aves, Mohamed Maaroufi, Sang-Hoon Yeo, R Chris Miall, Peter Holland, and Joseph M Galea. A methodological framework to assess the accuracy of virtual reality hand-tracking systems: A case study with the meta quest 2. *Behavior Research Methods*, pages 1–12, 2023.
  - [126] Meta Platforms, Inc. Meta quest 3: New mixed reality vr headset. <https://www.meta.com/quest/quest-3/>, 2024. Accessed: 2024-01-10.

- 
- [127] Jennifer Challenor, David White, and David Murphy. Hand-controlled user interfacing for head-mounted augmented reality learning environments. *Multimodal Technologies and Interaction*, 7(6):55, 2023.
- [128] Raspberry pi and python users - we've got news for you! [https://www.reddit.com/r/Ultraleap/comments/181boh/raspberry\\_pi\\_and\\_python\\_users\\_weve\\_got\\_news\\_for/](https://www.reddit.com/r/Ultraleap/comments/181boh/raspberry_pi_and_python_users_weve_got_news_for/), . Accessed: 2024-02-27.
- [129] Leap motion controller software for raspberry pi. <https://leap2.ultraleap.com/gemini-downloads/#tab-desktop>. Accessed: 2024-02-27.
- [130] Leap motion controller 2 datasheet. [https://www.ultraleap.com/datasheets/leap-motion-controller-2-datasheet\\_issue13.pdf](https://www.ultraleap.com/datasheets/leap-motion-controller-2-datasheet_issue13.pdf), . Accessed: 2024-02-03.
- [131] The operating temperature for a raspberry pi. <https://copperhilltech.com/content/The%20operating%20Temperature%20For%20A%20Raspberry%20Pi%20%E2%80%93%20Technologist%20Tips.pdf>. Accessed: 2024-02-03.
- [132] Chen Qian, Xiao Sun, Yichen Wei, Xiaou Tang, and Jian Sun. Realtime and robust hand tracking from depth. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1106–1113, 2014.
- [133] Koushik Roy and Md Akiful Hoque Akif. Real time hand gesture based user friendly human computer interaction system. In *2022 International Conference on Innovations in Science, Engineering and Technology (ICISSET)*, pages 260–265. IEEE, 2022.
- [134] What is the leap motion controller's operating environment range?, . URL <https://support.ultraleap.com/hc/en-us/articles/360004328878-What-is-the-Leap-Motion-Controller-s-operating-environment-range>. Accessed: 2024-03-02.
- [135] Giuseppe Placidi, Alessandro Di Matteo, Daniele Lozzi, Matteo Polsinelli, and Eleni Theodoridou. Patient–therapist cooperative hand telerehabilitation through a novel framework involving the virtual glove system. *Sensors*, 23(7):3463, March 2023. ISSN 1424-8220. doi: 10.3390/s23073463. URL <http://dx.doi.org/10.3390/s23073463>.
- [136] Giuseppe Placidi, Luigi Cinque, Andrea Petracca, Matteo Polsinelli, and Matteo Spezialetti. A virtual glove system for the hand rehabilitation based on two orthogonal leap motion controllers. In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*. SCITEPRESS - Science and Technology Publications, 2017. doi: 10.5220/0006197801840192. URL <http://dx.doi.org/10.5220/0006197801840192>.
- [137] Marika Carrieri, Andrea Petracca, Stefania Lancia, Sara Basso Moro, Sabrina Brigadoi, Matteo Spezialetti, Marco Ferrari, Giuseppe Placidi, and Valentina Quaresima. Prefrontal cortex activation upon a demanding virtual hand-controlled task: A new frontier for neuroergonomics. *Frontiers in Human Neuroscience*, 10, February 2016. ISSN 1662-5161. doi: 10.3389/fnhum.2016.00053. URL <http://dx.doi.org/10.3389/fnhum.2016.00053>.
- [138] Rob HJ Van der Lubbe, Jagna Sobierajewicz, Marijtje LA Jongsma, Willem B Verwey, and Anna Przekoracka-Krawczyk. Frontal brain areas are more involved during motor imagery than during motor execution/preparation of a response sequence. *International journal of psychophysiology*, 164: 71–86, 2021.
- [139] Jan-Niklas Voigt-Antons, Tanja Kojic, Danish Ali, and Sebastian Möller. Influence of hand tracking as a way of interaction in virtual reality on user experience. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–4. IEEE, 2020.

- 
- [140] Jie Li, Xiaofeng Liu, Zhelong Wang, Tingting Zhang, Sen Qiu, Hongyu Zhao, Xu Zhou, Huili Cai, Rongrong Ni, and Angelo Cangelosi. Real-time hand gesture tracking for human–computer interface based on multi-sensor data fusion. *IEEE Sensors Journal*, 21(23):26642–26654, 2021.
  - [141] Aveen Dayal, Naveen Paluru, Linga Reddy Cenkeramaddi, and Phaneendra K Yalavarthy. Design and implementation of deep learning based contactless authentication system using hand gestures. *Electronics*, 10(2):182, 2021.
  - [142] Mohammad Rastegari, Vicente Ordonez, Joseph Redmon, and Ali Farhadi. Xnor-net: Imagenet classification using binary convolutional neural networks. In *European conference on computer vision*, pages 525–542. Springer, 2016.
  - [143] Zhengwei Wang, Qi She, and Aljosa Smolic. Action-net: Multipath excitation for action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13214–13223, 2021.
  - [144] Yifan Zhang, Congqi Cao, Jian Cheng, and Hanqing Lu. Egogesture: a new dataset and benchmark for egocentric hand gesture recognition. *IEEE Transactions on Multimedia*, 20(5):1038–1050, 2018.
  - [145] Zihan Xu, Mingbao Lin, Jianzhuang Liu, Jie Chen, Ling Shao, Yue Gao, Yonghong Tian, and Rongrong Ji. Recu: Reviving the dead weights in binary neural networks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5198–5208, 2021.
  - [146] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.
  - [147] Xiangning Chen, Chen Liang, Da Huang, Esteban Real, Kaiyuan Wang, Yao Liu, Hieu Pham, Xuanyi Dong, Thang Luong, Cho-Jui Hsieh, Yifeng Lu, and Quoc V. Le. Symbolic discovery of optimization algorithms, 2023.
  - [148] Ji Lin, Chuang Gan, and Song Han. Tsm: Temporal shift module for efficient video understanding. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7083–7093, 2019.
  - [149] Boyuan Jiang, MengMeng Wang, Weihao Gan, Wei Wu, and Junjie Yan. Stm: Spatiotemporal and motion encoding for action recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2000–2009, 2019.
  - [150] Clemens Lang, Silvia Winkler, Johannes Koren, Martin Huber, Tilmann Kluge, and Christoph Baumgartner. Dicom® integrated eeg data: A first clinical implementation of the new dicom standard for neurophysiology data. *Clinical Neurophysiology*, 155:107–112, 2023.
  - [151] Robert Campbell, Ed Pentz, and Ian Borthwick. Academic and professional publishing. 2012.
  - [152] ANSI (American National Standards Institute). Ansi essential requirements. <https://www.ansi.org/american-national-standards/ans-introduction/essential-requirements>, 2022. Online; accessed 29 June 2023.
  - [153] UNIDO (United Nations Industrial Development Organization). Private standards. <https://www.unido.org/our-focus/advancing-economic-competitiveness/meeting-standards/private-standards>, 2023. Online; accessed 29 June 2023.
  - [154] H Edward Shortliffe and J James Cimino. *Biomedical informatics: computer applications in health care and biomedicine*. Springer, (2014).
-

- 
- [155] Alex AT Bui and Ricky K Taira. *Medical imaging informatics*. Springer Science & Business Media, (2009).
- [156] Min-Fang Tsai, Shin-Yuan Hung, Wen-Ju Yu, Charlie C Chen, and David C Yen. Understanding physicians' adoption of electronic medical records: Healthcare technology self-efficacy, service level and risk perspectives. *Computer Standards & Interfaces*, 66:103342, 2019.
- [157] Janita FJ Vos, Albert Boonstra, Arjen Kooistra, Marc Seelen, and Marjolein van Offenbeek. The influence of electronic health record use on collaboration among medical specialties. *BMC health services research*, 20(1):1–11, 2020.
- [158] Paolo Marcheschi. Relevance of ehealth standards for big data interoperability in radiology and beyond. *La radiologia medica*, 122(6):437–443, 2017.
- [159] Erik J Zhang, Heng Tan, Joseph A Sanford, James D Michelson, Brian M Waldschmidt, and Mitchell H Tsai. Rebooting the electronic health record. *Journal of Medical Systems*, 46(7):48, 2022.
- [160] Obinwa Ozonze, Philip J Scott, and Adrian A Hopgood. Automating electronic health record data quality assessment. *Journal of Medical Systems*, 47(1):23, 2023.
- [161] Liting Yan. Dicom standard and its application in pacs system. *Medical Imaging Process & Technology*, 1(1), 2018.
- [162] NEMA (National Electrical Manufacturers Association ). Dicom standard. <https://www.dicomstandard.org/current>, 2023. Online; accessed 29 June 2023.
- [163] Jonathan J Halford, David A Clunie, Benjamin H Brinkmann, Dagmar Krefting, Jan Rémi, Felix Rosenow, Aatif Husain, Franz Fürbass, J Andrew Ehrenberg, and Silvia Winkler. Standardization of neurophysiology signal data into the dicom® standard. *Clinical Neurophysiology*, 132(4):993–997, 2021.
- [164] NEMA (National Electrical Manufacturers Association ). Strategic document. <https://dicom.nema.org/dicom/geninfo/Strategy.pdf>, 2023. Online; accessed 1 September 2023.
- [165] DICOM Part 3: Information Object Definitions with Waveform Module - Section A.34 . [https://dicom.nema.org/medical/dicom/current/output/html/part03.html#sect\\_A.34](https://dicom.nema.org/medical/dicom/current/output/html/part03.html#sect_A.34), . Accessed: 2024-02-20.
- [166] Wei-Lun Huang, Shuya Liu, Jun Kang, Amir Gandjbakhche, and Mehran Armand. Dicom file for total body photography: a work item proposal. In *Photonics in Dermatology and Plastic Surgery 2022*, volume 11934, pages 64–74. SPIE, 2022.
- [167] Oliver Burgert, Thomas Treichel, Christian Dressler, and Michael Gessat. Storing data generated by optical surface scanners using dicom: a work item proposal. In *Medical Imaging 2010: Advanced PACS-based Imaging Informatics and Therapeutic Applications*, volume 7628, pages 87–96. SPIE, 2010.
- [168] Mahmoud Ismail and James Philbin. Multi-series dicom: an extension of dicom that stores a whole study in a single object. *Journal of digital imaging*, 26:691–697, 2013.
- [169] Innolitics, LLC. DICOM Standard Browser. <https://dicom.innolitics.com/ciods/>, 2020. Accessed: 2024-02-20.
- [170] Approved Supplements - DICOM Standard. <https://www.dicomstandard.org/approved-supplements>, . Accessed: 2024-02-20.



- 
- [171] Bas Revet. Dicom cook book for implementations in modalities. *Philips Medical Systems, Eindhoven, Netherlands*, 1997.
- [172] NEMA (National Electrical Manufacturers Association ). Dicom ecg iod. [https://dicom.nema.org/medical/dicom/current/output/chtml/part03/sect\\_A.34.4.html](https://dicom.nema.org/medical/dicom/current/output/chtml/part03/sect_A.34.4.html), 2023. Online; accessed September 1, 2023.
- [173] NEMA (National Electrical Manufacturers Association ). Dicom mri iod. [https://dicom.nema.org/dicom/2013/output/chtml/part03/sect\\_A.4.htm](https://dicom.nema.org/dicom/2013/output/chtml/part03/sect_A.4.htm), 2023. Online; accessed September 1, 2023.
- [174] NEMA (National Electrical Manufacturers Association ). Dicom part real world composite iod structure. [https://dicom.nema.org/MEDICAL/Dicom/current/output/chtml/part03/chapter\\_A.html#sect\\_A.1.2](https://dicom.nema.org/MEDICAL/Dicom/current/output/chtml/part03/chapter_A.html#sect_A.1.2), 2023. Online; accessed September 1, 2023.
- [175] NEMA (National Electrical Manufacturers Association ). Dicom part 7. [https://dicom.nema.org/medical/dicom/current/output/chtml/part07/sect\\_7.5.html](https://dicom.nema.org/medical/dicom/current/output/chtml/part07/sect_7.5.html), 2023. Online; accessed September 1, 2023.
- [176] NEMA (National Electrical Manufacturers Association ). Dicom information model. [https://dicom.nema.org/medical/dicom/current/output/chtml/part04/chapter\\_6.html](https://dicom.nema.org/medical/dicom/current/output/chtml/part04/chapter_6.html), 2023. Online; accessed September 1, 2023.
- [177] NEMA (National Electrical Manufacturers Association ). Dicom part 4. <https://dicom.nema.org/medical/dicom/current/output/html/part04.html>, 2023. Online; accessed September 1, 2023.
- [178] Susan B O’Sullivan, Thomas J Schmitz, and George Fulk. *Physical rehabilitation*. FA Davis, (2019). ISBN 0803661622.
- [179] Lucas Medeiros Souza do Nascimento, Lucas Vacilotto Bonfati, Melissa La Banca Freitas, José Jair Alves Mendes Junior, Hugo Valadares Siqueira, and Sergio Luiz Stevan Jr. Sensors and systems for physical rehabilitation and health monitoring—a review. *Sensors*, 20(15):4063, 2020.
- [180] Jorge Garza Ulloa. *Applied Biomechatronics Using Mathematical Models*. Academic press, (2018).
- [181] Zhen Gang Xiao and Carlo Menon. A review of force myography research and development. *Sensors*, 19(20):4557, 2019.
- [182] Kerry R Mills. The basics of electromyography. *Journal of Neurology, Neurosurgery & Psychiatry*, 76 (suppl 2):ii32–ii35, 2005.
- [183] Thomas Hellsten, Jonny Karlsson, Muhammed Shamsuzzaman, and Göran Pulkkis. The potential of computer vision-based marker-less human motion analysis for rehabilitation. *Rehabilitation Process and Outcome*, 10:11795727211022330, 2021.
- [184] Ultraleap. Leap motion controller (lmc). <https://leap2.ultraleap.com/leap-motion-controller-2/>, 2023. Online; accessed September 1, 2023.
- [185] Microsoft. Microsoft kinect. <https://azure.microsoft.com/it-it/products/kinect-dk>, 2023. Online; accessed September 1, 2023.
- [186] Nieves Pavón-Pulido, Juan Antonio López-Riquelme, and Jorge J Feliú-Batlle. Iot architecture for smart control of an exoskeleton robot in rehabilitation by using a natural user interface based on gestures. *Journal of Medical Systems*, 44:1–10, 2020.

- 
- [187] Enjie Ghorbel, Renato Baptista, Abdelrahman Shabayek, Djamila Aouada, Maialen Gorostiza Ormaeche, Janire Orcajo Lago, and Leire Ortiz Fernandez. Home-based rehabilitation system for stroke survivors: a clinical evaluation. *Journal of medical systems*, 44:1–11, 2020.
  - [188] Giuseppe Placidi, Danilo Avola, Daniela Iacoviello, and Luigi Cinque. Overall design and implementation of the virtual glove. *Computers in Biology and Medicine*, 43(11):1927–1940, 2013. ISSN 0010-4825. doi: <https://doi.org/10.1016/j.compbimed.2013.08.026>. URL <https://www.sciencedirect.com/science/article/pii/S0010482513002515>.
  - [189] Sara Basso Moro, Marika Carrieri, Danilo Avola, Sabrina Brigadoi, Stefania Lancia, Andrea Petracca, Matteo Spezialetti, Marco Ferrari, Giuseppe Placidi, and Valentina Quaresima. A novel semi-immersive virtual reality visuo-motor task activates ventrolateral prefrontal cortex: a functional near-infrared spectroscopy study. *Journal of Neural Engineering*, 13(3):036002, March 2016. ISSN 1741-2552. doi: 10.1088/1741-2560/13/3/036002. URL <http://dx.doi.org/10.1088/1741-2560/13/3/036002>.
  - [190] NEMA (National Electrical Manufacturers Association ). Dicom waveform ie. [https://dicom.nema.org/medical/dicom/current/output/chtml/part17/sect\\_C.5.html](https://dicom.nema.org/medical/dicom/current/output/chtml/part17/sect_C.5.html), 2023. Online; accessed September 1, 2023.
  - [191] D. L. Mason et al. pydicom: An open source DICOM library. <https://github.com/pydicom/pydicom>. Version: 2.4.4.
  - [192] Riverbank Computing Limited. Pyqt. <https://riverbankcomputing.com/software/pyqt>. Version: 5.15.10.
  - [193] Alessandro Di Matteo. DICOM-based Standard for Quantitative Physical Rehabilitation. <https://github.com/alessandrodimatteo97/DICOM-based-Standard-for-Quantitative-Physical-Rehabilitation>.
  - [194] Jorge Miguel Silva, Tiago Marques Godinho, David Silva, and Carlos Costa. A community-driven validation service for standard medical imaging objects. *Computer Standards & Interfaces*, 61:121–128, 2019.
  - [195] Manisha Mantri, Shruti Taran, and Gaur Sunder. Dicom integration libraries for medical image interoperability: a technical review. *IEEE Reviews in Biomedical Engineering*, 15:247–259, 2020.
  - [196] J. Bernhardt et al. Agreed definitions and a shared vision for new standards in stroke recovery research: The stroke recovery and rehabilitation roundtable taskforce. *Int J Stroke*, 12:444–450, 2017. doi: 10.1177/1747493017711816.
  - [197] Isabelle Bonan, Stephanie Leplaideur, and Philippe Carson. Rééducation de l’équilibre après accident vasculaire cérébral. In *Accident vasculaire cérébral et médecine physique et de réadaptation: Actualités en 2010*, pages 37–44. Springer, 2010.
  - [198] Karim Jamal, Sébastien Cordillet, Stephanie Leplaideur, Hélène Rauscent, Melanie Cogne, and Isabelle Bonan. Reliability and minimal detectable change of body-weight distribution and body sway between right and left brain-damaged patients at a chronic stage. *Disability and Rehabilitation*, 45(2):260–265, 2023.
  - [199] Matteo Polsinelli, Luigi Cinque, and Giuseppe Placidi. A light cnn for detecting covid-19 from ct scans of the chest. *Pattern recognition letters*, 140:95–100, 2020.
  - [200] Fabian Isensee, Tassilo Wald, Constantin Ulrich, Michael Baumgartner, Saikat Roy, Klaus Maier-Hein, and Paul F Jaeger. nnu-net revisited: A call for rigorous validation in 3d medical image segmentation.

- In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 488–498. Springer, 2024.
- [201] Lucas W Remedios, Han Liu, Samuel W Remedios, Lianrui Zuo, Adam M Saunders, Shunxing Bao, Yuankai Huo, Alvin C Powers, John Virostko, and Bennett A Landman. Influence of early through late fusion on pancreas segmentation from imperfectly registered multimodal magnetic resonance imaging. *Journal of Medical Imaging*, 12(2):024008–024008, 2025.
  - [202] Francisco Gomez-Donoso, Sergio Orts-Escolano, Alberto Garcia-Garcia, Jose Garcia-Rodriguez, John Alejandro Castro-Vargas, Sergiu Ovidiu-Oprea, and Miguel Cazorla. A robotic platform for customized and interactive rehabilitation of persons with disabilities. *Pattern Recognition Letters*, 99: 105–113, 2017.
  - [203] Huaigang Yang, Dong Zhang, Ping Xie, and Xiaoling Chen. Dsc-grunet: A lightweight neural network model for multimodal gesture recognition based on depthwise separable convolutions and gru. *Pattern Recognition Letters*, 190:35–44, 2025.
  - [204] Alessandro Di Matteo, Daniele Lozzi, Enrico Mattei, Filippo Mignosi, Sara Montagna, Matteo Polsinelli, and Giuseppe Placidi. Calibration of the double digital twin for the hand rehabilitation by the virtual glove. In *2024 IEEE 37th international symposium on computer-based medical systems (CBMS)*, pages 521–526. IEEE, 2024.
  - [205] R. Ahmed et al. An appraisal of the performance of ai tools for chronic stroke lesion segmentation. *Comput Biol Med*, 164:107302, 2023. doi: 10.1016/j.compbimed.2023.107302.
  - [206] S.-L. Liew et al. Open source dataset of stroke anatomical brain images and manual lesion segmentations. *Sci Data*, 5:180011, 2018. doi: 10.1038/sdata.2018.11.
  - [207] K. Verma et al. Automatic segmentation and quantitative assessment of stroke lesions on mr images. *Diagnostics*, 12:2055, 2022. doi: 10.3390/diagnostics12092055.
  - [208] J. Huo et al. Mapping: Model average with post-processing for stroke lesion segmentation. Preprint, 2022.
  - [209] A. Masson et al. A nnunet implementation of new lesions segmentation from serial flair images of ms patients. In *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 45–51, 2021. URL <https://hal.inria.fr/hal-03358968v1>.
  - [210] Olivier Commowick, Frédéric Cervenansky, François Cotton, and Michel Dojat. Msseg-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure. In *MICCAI 2021-24th international conference on medical image computing and computer assisted intervention*, page 126, 2021.
  - [211] F. Isensee et al. Automated brain extraction of multisequence mri using artificial neural networks. *Hum Brain Mapp*, 40:4952–4964, 2019. doi: 10.1002/hbm.24750.
  - [212] N. J. Tustison et al. N4itk: Improved n3 bias correction. *IEEE Trans Med Imaging*, 29:1310–1320, 2010. doi: 10.1109/TMI.2010.2046908.
  - [213] A. Carré et al. Standardization of brain mr images across machines and protocols: Bridging the gap for mri-based radiomics. *Sci Rep*, 10:12340, 2020. doi: 10.1038/s41598-020-69298-z.

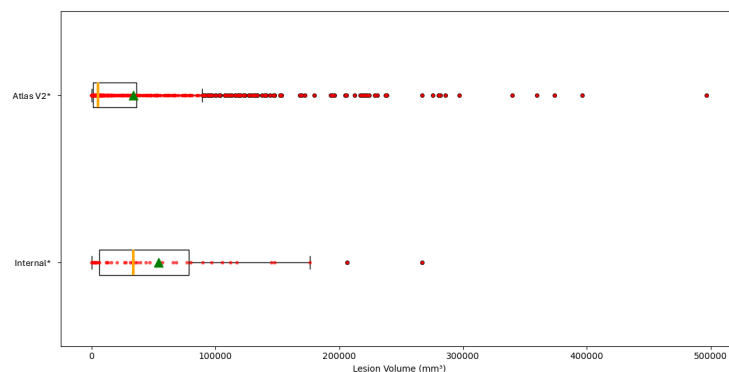
- [214] Krzysztof J Gorgolewski, Tibor Auer, Vince D Calhoun, R Cameron Craddock, Samir Das, Eugene P Duff, Guillaume Flandin, Satrajit S Ghosh, Tristan Glatard, Yaroslav O Halchenko, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, 3(1):1–9, 2016.
- [215] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
- [216] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.
- [217] Gisle Berg Helland, Mona Kristiansen Beyer, Brian Anthony B Enriquez, Hege Ihle-Hansen, Håkon Ihle-Hansen, Stein Andersson, Esten Høyland Leonardsen, Helle Stangeland, Bettina Üjhelyi, Guri Hagberg, et al. Lesion volume as a predictor for return to work after endovascular treatment: A 4-year prospective cohort study. *Stroke: Vascular and Interventional Neurology*, 4(6):e001494, 2024.

## Part VII

# Appendix

## Appendix A

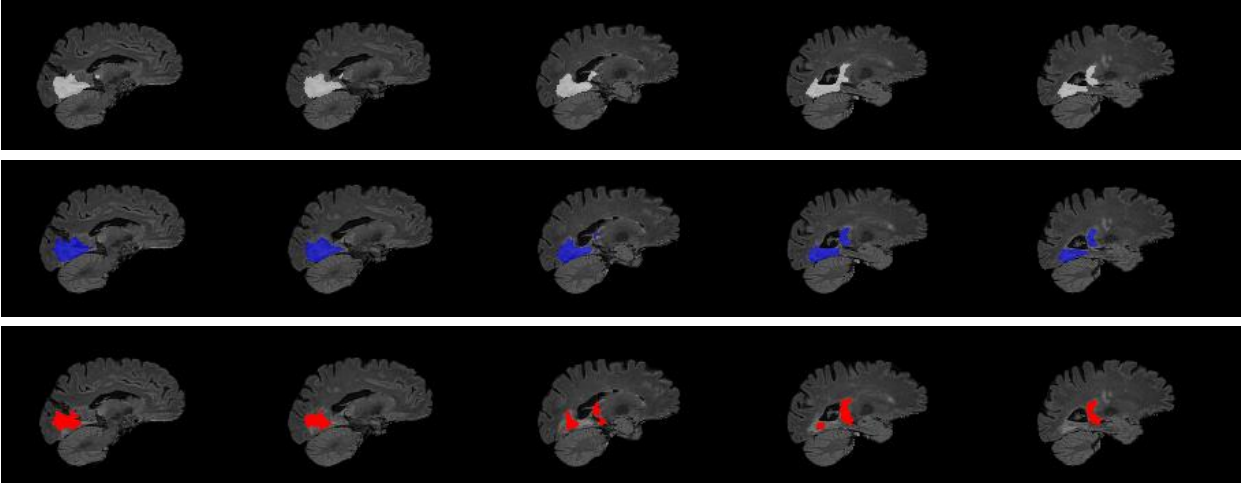
# Supplementary Material Chapter 8. Deep Learning and Multi-Modal MRI for the Segmentation of Sub-Acute and Chronic Stroke Lesions



**Figure A.1:** Distribution of lesion volumes for individual subjects in each dataset. The triangle represents the mean score, and the yellow bar represents the standard deviation.

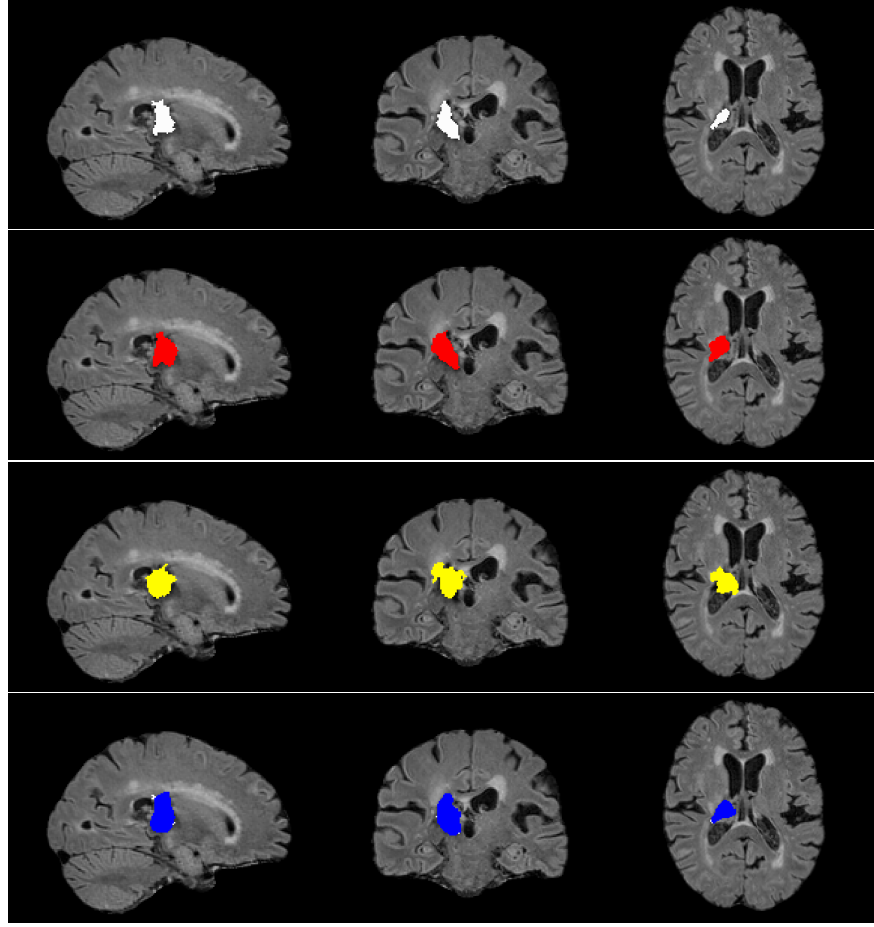
### Per-subject performance analysis

A detailed per-subject analysis, summarized in the boxplots shown in Figures A.4–A.16, revealed several informative cases. Patients 44 and 32 (with small and medium lesions, respectively) consistently emerged as the most challenging, recording zero values for both voxel-wise Dice and lesion-wise F1-scores across all configurations, including fine-tuned and ensembled variants (see **Figures A.4–A.16**). This indicates that none of the fusion strategies successfully segmented these extreme cases, likely because the lesions were very small or poorly defined. Patient 20 (with a small lesion) exhibited an anomalous behavior: performance in non-ensembled models was already poor (see Figures 9B,9F,10B,10F) and dropped to zero in the T1+FLAIR early-fusion ensemble (see Figures A.13B, A.13F). This failure likely reflects a combination of weak lesion contrast, small lesion size, and minor T1–FLAIR misalignment, which together led to destructive interference when the modalities were combined at the input level. Mid- and late-fusion configurations,



**Figure A.2:** Sagittal sections showing the FLAIR image with manual segmentation of Patient 43 (large lesion), along with results from the dual-modality model (middle, blue) and the single-modality model (bottom, red). The Dice scores for the dual-modality and single-modality (fine-tuned (ensemble)d) models for this patient are 0.81 and 0.60, respectively.

which encode modalities independently before integration, partially recovered performance: the late-fusion ensemble achieved a Dice of 0.425 and a lesion-wise F1 score of 0.648 (see Figures A.15B, 15F). In comparison, the mid-fusion ensemble reached Dice = 0.24 and F1 = 0.40 (see Figures A.14B, A.14F). In contrast, patients 6, 17, and 28 (with small lesions) exhibited heterogeneous responses across fusion strategies. Patient 6 maintained stable performance with early (Dice = 0.635, F1 = 1.0; see Figures A.6B, A.6F) and late fusion (Dice = 0.478, F1 = 1.0; see Figures A.8B, A.8F), but performance dropped sharply with mid fusion (Dice = 0.085, F1 = 0.0; see Figures A.7B, A.7F). Patient 17 achieved the best results with early fusion (Dice = 0.854, F1 = 0.867; see Figures A.6B, A.6F), whereas patient 28 showed consistently high scores across all configurations, with late fusion slightly outperforming early (Dice = 0.758 vs. 0.722; see Figures A.13B, A.15B). Overall, FLAIR provided complementary contrast, but the effectiveness of each fusion strategy depended on lesion characteristics and the spatial alignment between modalities. Patient 42 (with a small lesion) displayed notable variability across fusion strategies. The early-fusion average underperformed (Dice = 0.478, F1 = 0.233; see Figures A.9B, A.9F), while the late-fusion average showed improvement (Dice = 0.682, F1 = 0.497; see Figures A.11B, A.11F), and the late-fusion ensemble achieved the highest accuracy (Dice = 0.754, F1 = 1.0; see Figures A.15B, A.15F). This variability likely resulted from slight T1–FLAIR misalignment and irregular lesion morphology, which made the model sensitive to fold-specific differences. Averaging predictions across folds mitigated this effect, demonstrating that ensembling can stabilize multimodal integration when image quality or spatial correspondence is suboptimal. On the other hand, patients 4, 5, and 29 (with medium, small, and small lesion sizes, respectively) showed relatively stable or improved performance across all models, with high scores in both Dice and lesion-wise F1-score. For patient 29 in particular, the late-fusion strategy (avg Dice = 0.640, F1 = 0.733; ensemble Dice = 0.774, F1 = 1.0; see Figures A.8B, A.8E) clearly outperformed early fusion (avg Dice = 0.250, F1 = 0.267; ensemble Dice = 0.424, F1 = 0.5; see Figures A.6B, A.6F). For these subjects, the integration of FLAIR appeared clearly beneficial. Overall, the per-subject analysis indicates that segmentation performance varies with lesion size, contrast, and inter-modality alignment. Cases with small or poorly contrasted lesions (e.g., patients 20, 32, and 44) showed the lowest accuracy, while those with strong and spatially consistent FLAIR contrast (e.g., patients 17 and 28) benefited from early fusion. Conversely, patients with slight T1–FLAIR misalignment or irregular lesion morphology (e.g., patient 42) achieved better stability with late fusion. These trends suggest that early fusion favors well-aligned, high-contrast inputs, whereas late fusion offers greater robustness when modality signals differ in quality or registration.

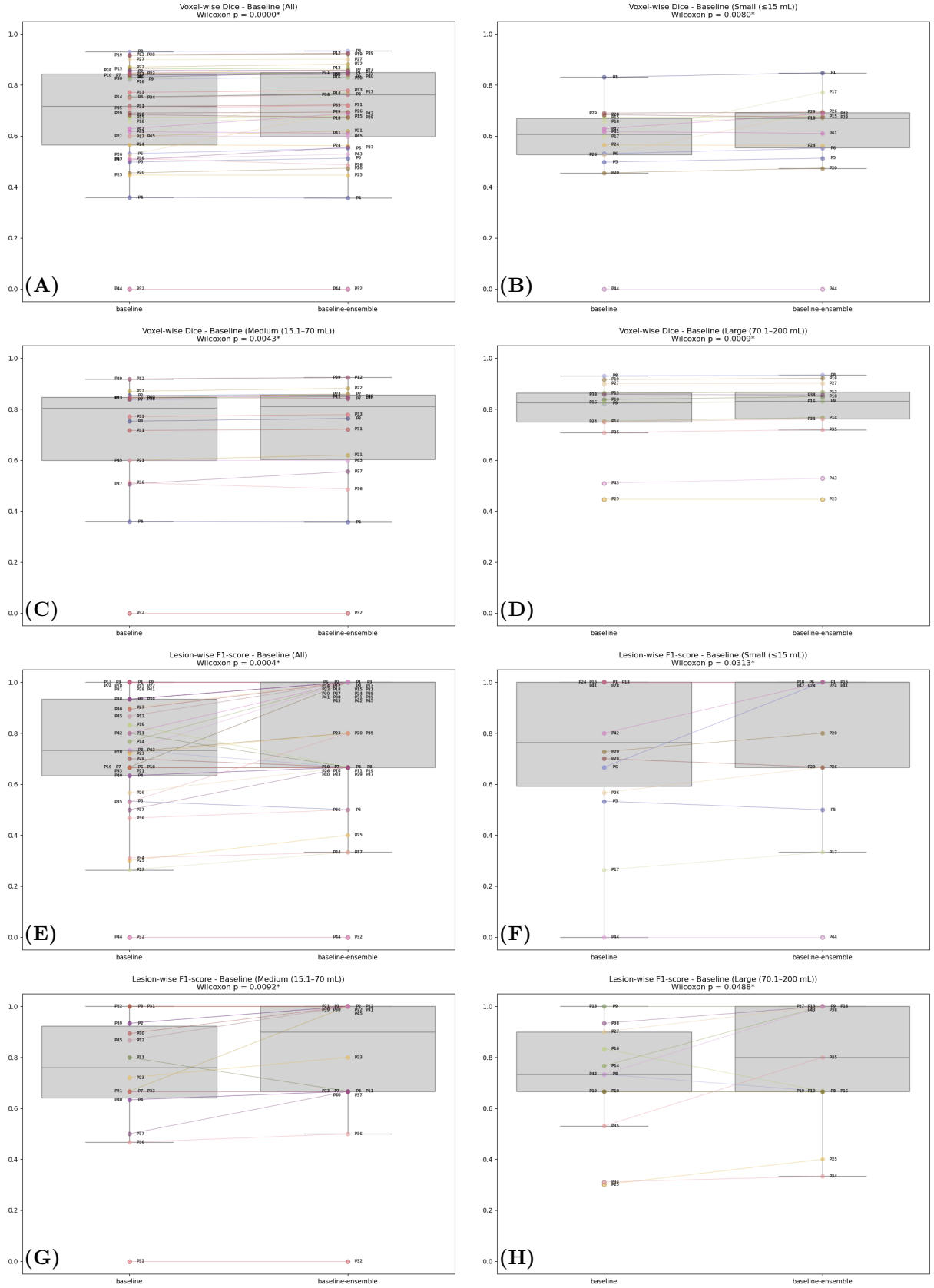


**Figure A.3:** Sagittal, coronal, and axial FLAIR slices of Patient 18 (characterized by small lesions), showing the manual segmentation alongside ensemble predictions from the early-, mid-, and late-fusion models, with corresponding Dice scores of 0.69, 0.41, and 0.76, respectively.

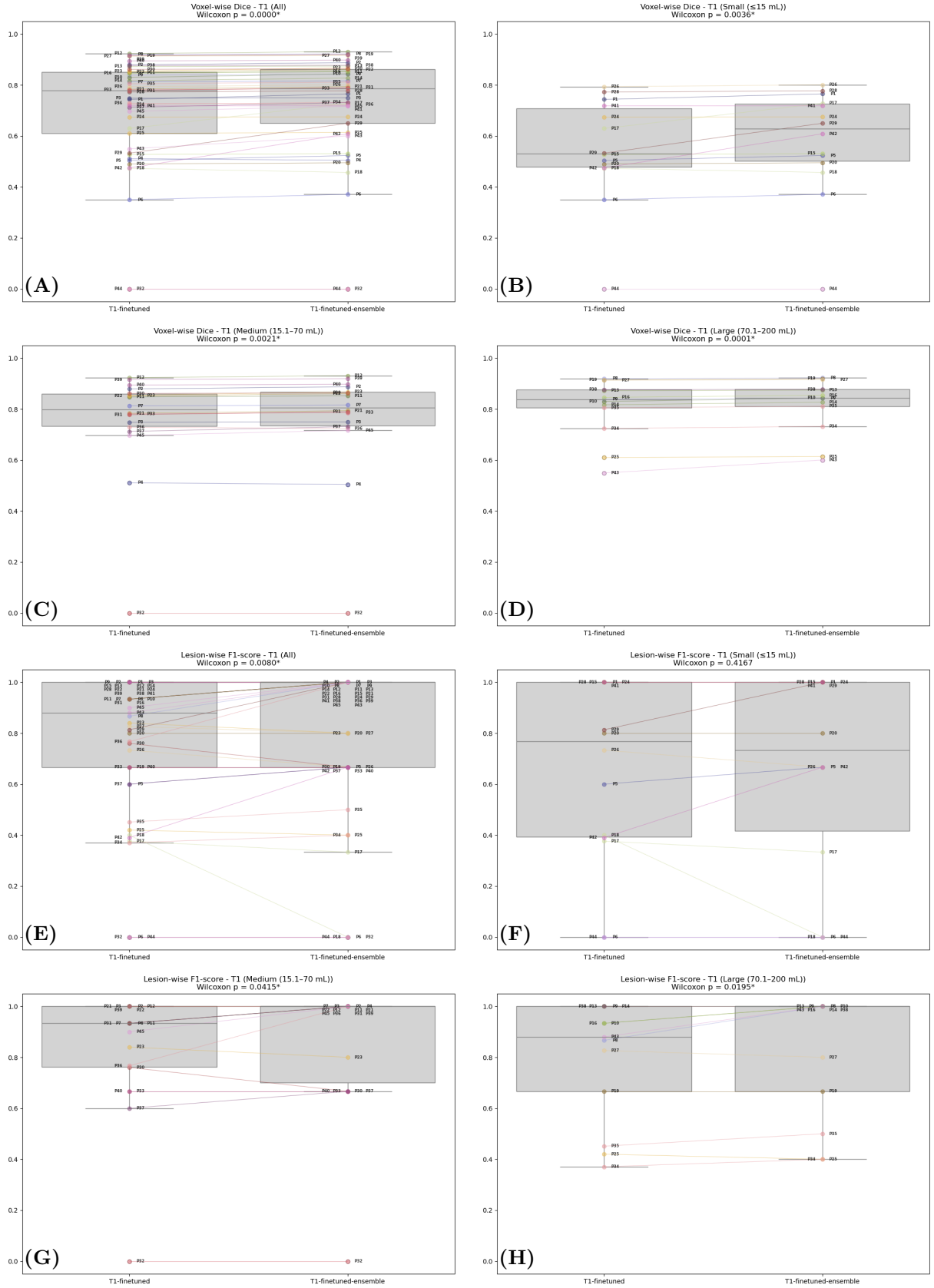
**Table A.1:** Statistical comparison (paired two-sided Wilcoxon signed-rank tests) for T1 fine-tuned models using voxel-wise Dice. Bold indicates  $p < 0.05$ .

| Model Comparison                                | Overall        | Small         | Medium        | Large          |
|-------------------------------------------------|----------------|---------------|---------------|----------------|
| Baseline (ensemble) vs Baseline                 | <b>0.0e+00</b> | <b>0.0080</b> | <b>0.0043</b> | <b>9.0e-04</b> |
| T1 fine-tuned vs Baseline                       | <b>0.0116</b>  | 0.7238        | <b>0.0037</b> | <b>0.0199</b>  |
| T1 fine-tuned vs Baseline (ensemble)            | 0.1581         | 0.9497        | <b>0.0308</b> | 0.0732         |
| T1 fine-tuned (ensemble) vs Baseline            | <b>9.0e-04</b> | 0.4307        | <b>0.0013</b> | <b>0.0040</b>  |
| T1 fine-tuned (ensemble) vs Baseline (ensemble) | <b>0.0219</b>  | 0.7238        | <b>0.0065</b> | <b>0.0239</b>  |
| T1 fine-tuned (ensemble) vs T1 fine-tuned       | <b>0.0e+00</b> | <b>0.0036</b> | <b>0.0021</b> | <b>1.0e-04</b> |

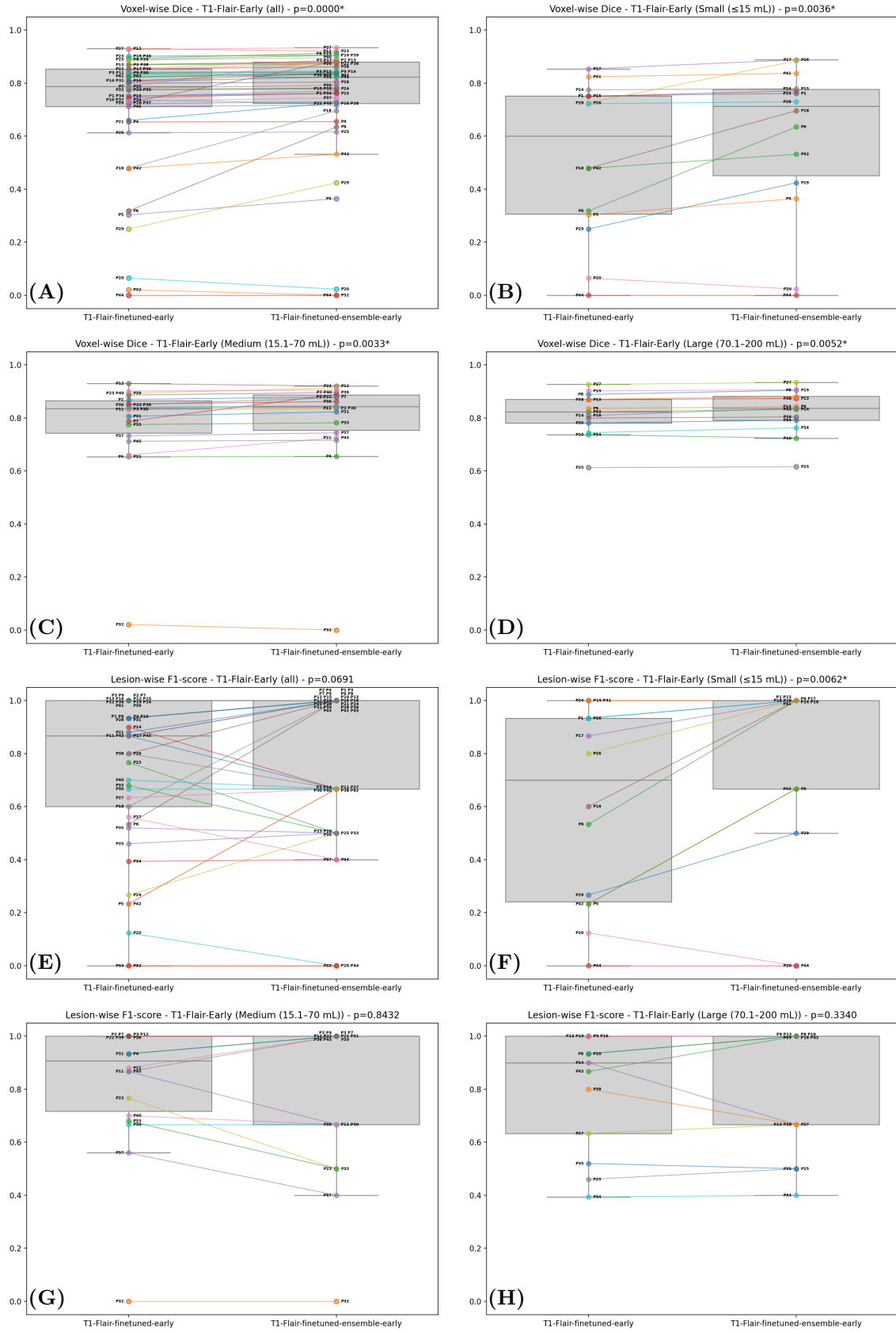




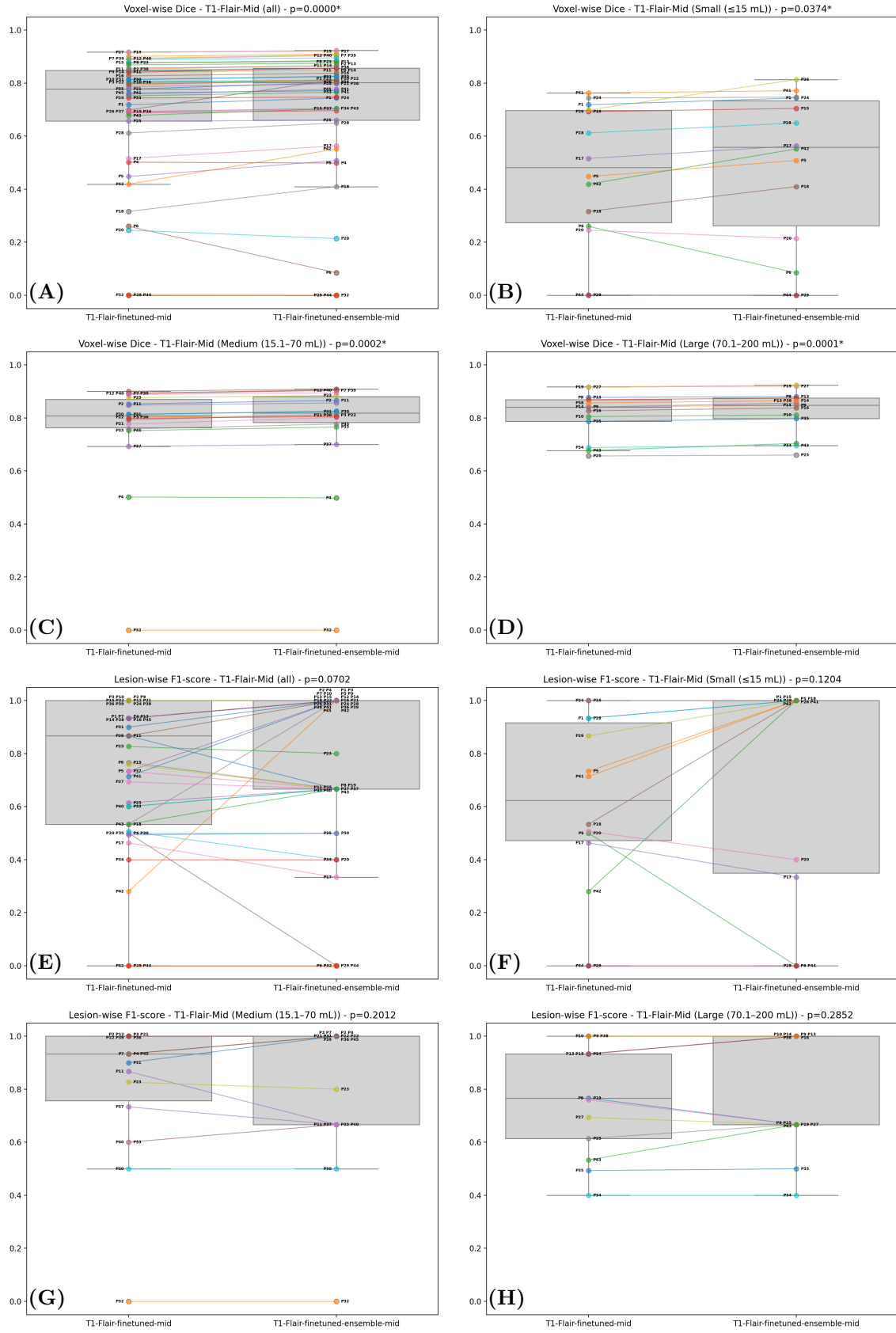
**Figure A.4:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the Baseline model across all lesion volume categories (All, Small, Medium, Large).



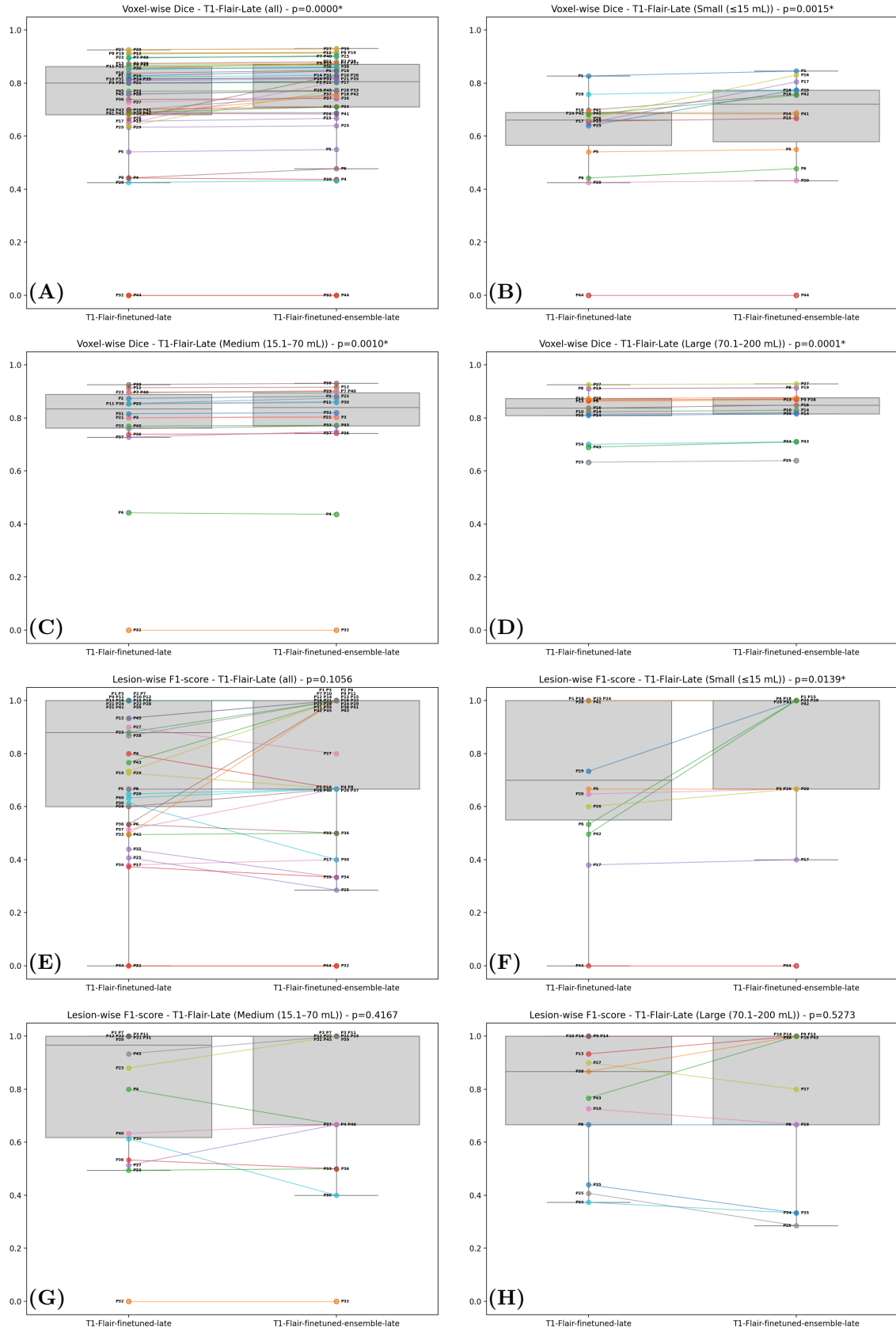
**Figure A.5:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the T1 fine-tuned model across all lesion volume categories (All, Small, Medium, Large).



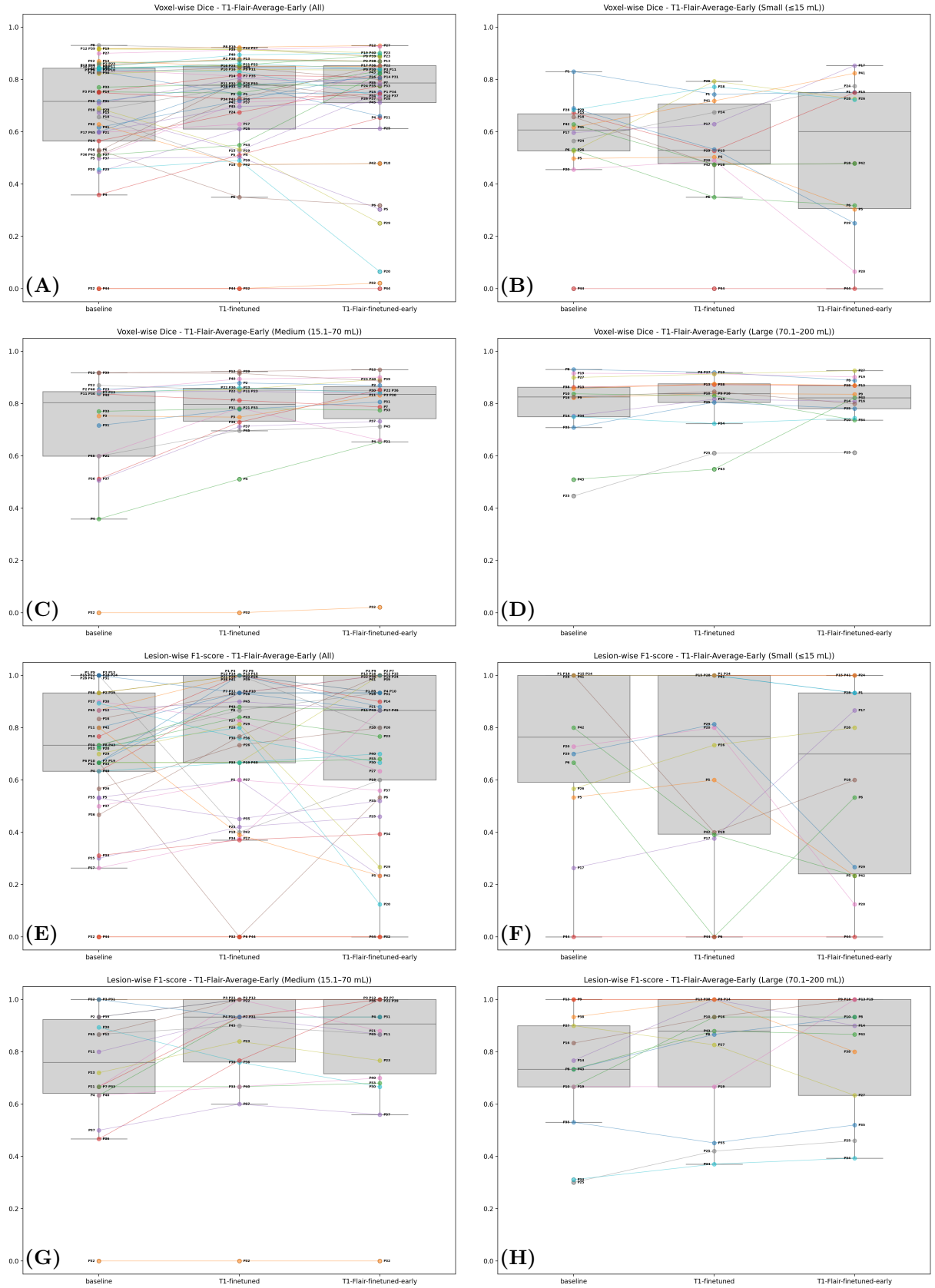
**Figure A.6:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble Early-fusion model across all lesion volume categories (All, Small, Medium, Large).



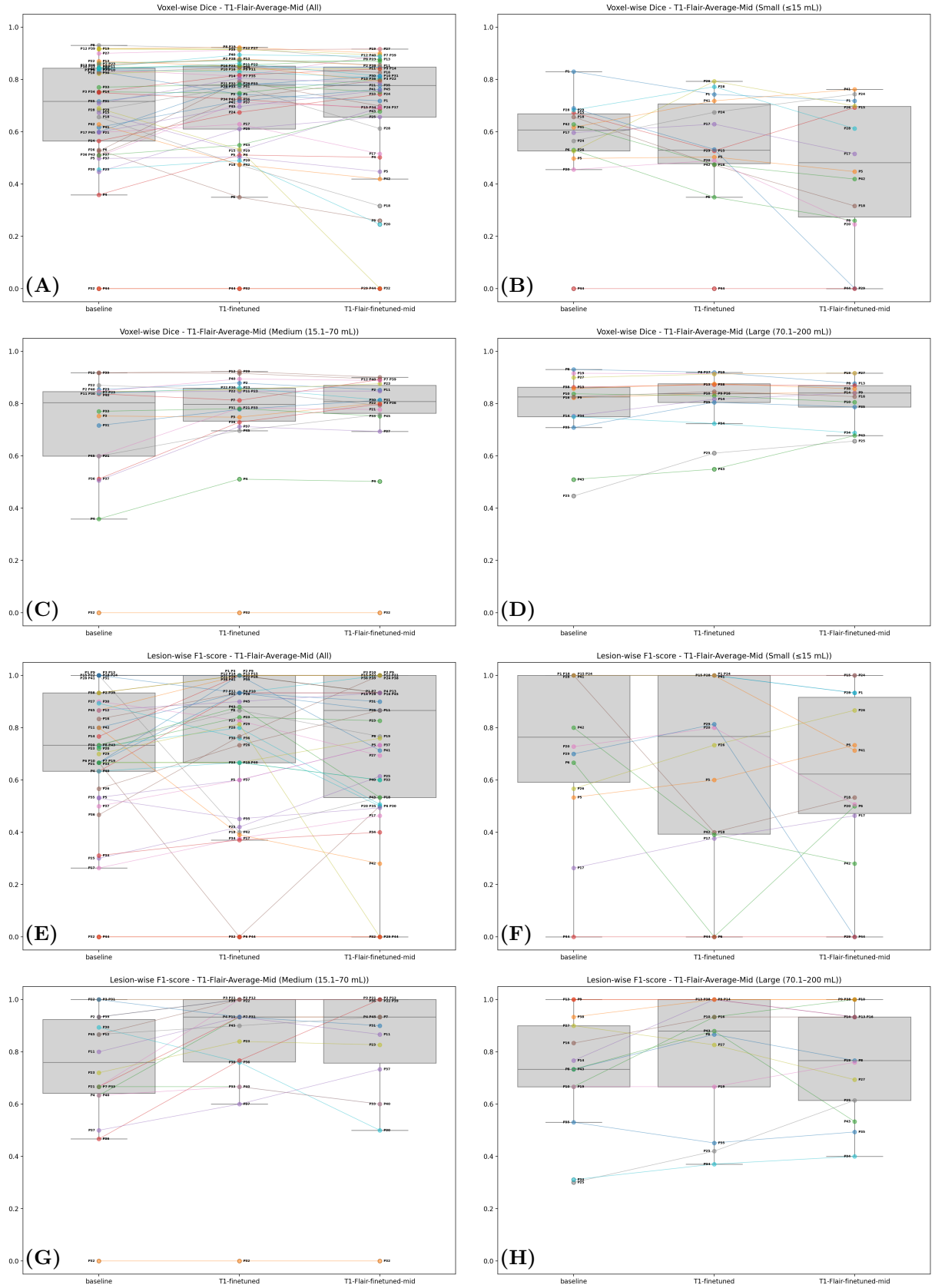
**Figure A.7:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble Mid-fusion model across all lesion volume categories (All, Small, Medium, Large).



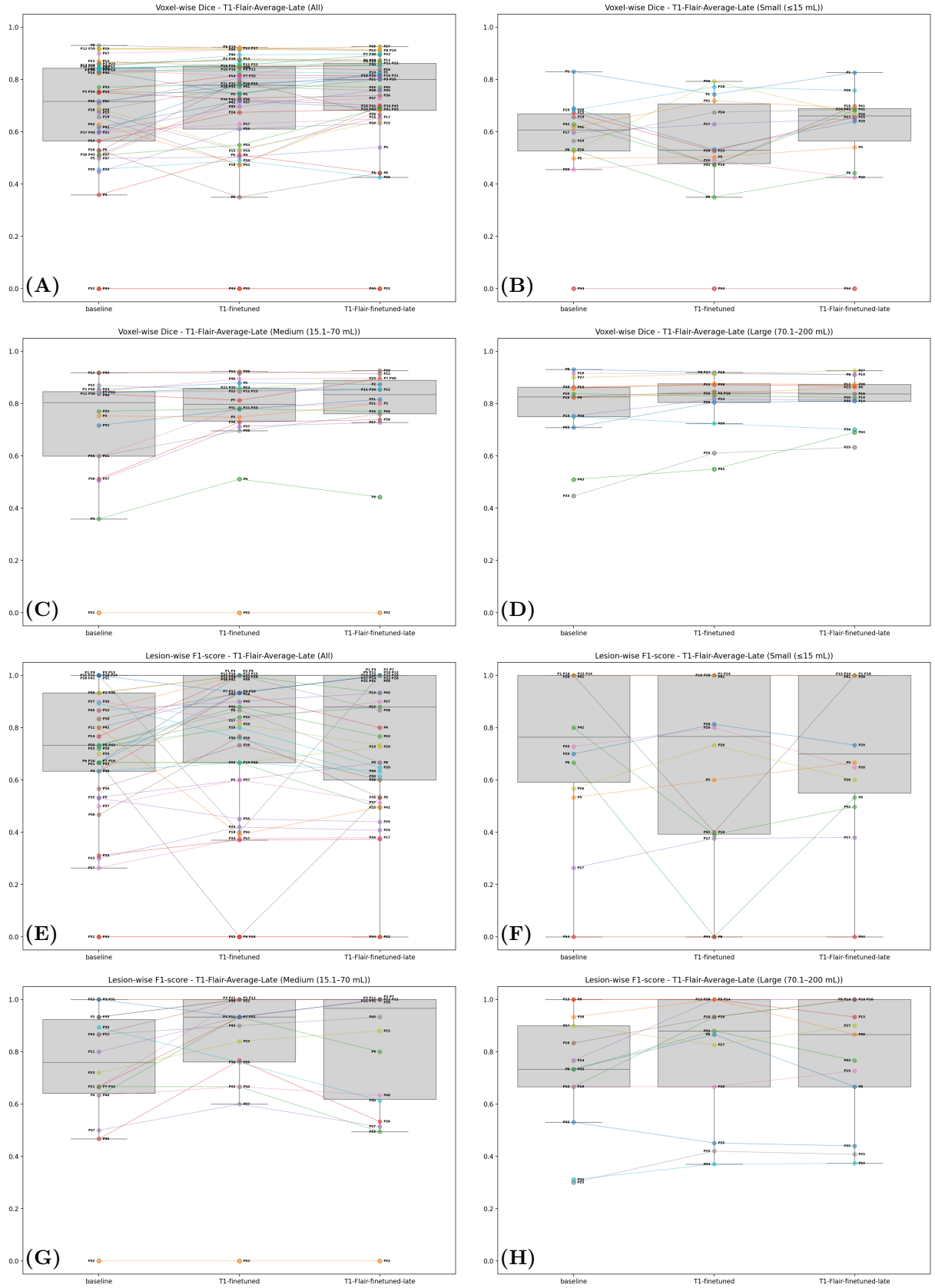
**Figure A.8:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble Late-fusion model across all lesion volume categories (All, Small, Medium, Large).



**Figure A.9:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the Averaged models (Early fusion) across all lesion volume categories (All, Small, Medium, Large).

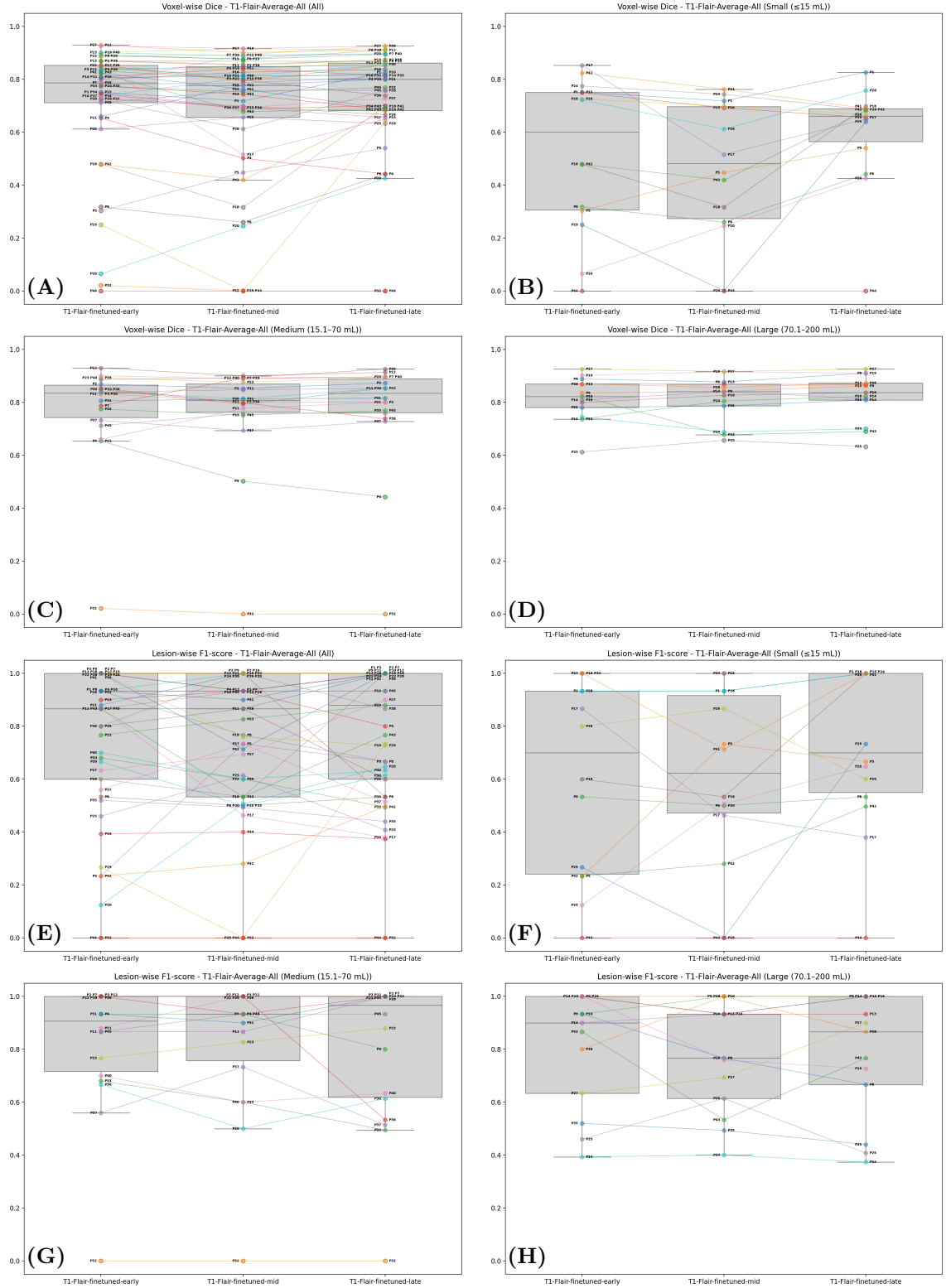


**Figure A.10:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the Averaged models (Mid fusion) across all lesion volume categories (All, Small, Medium, Large).

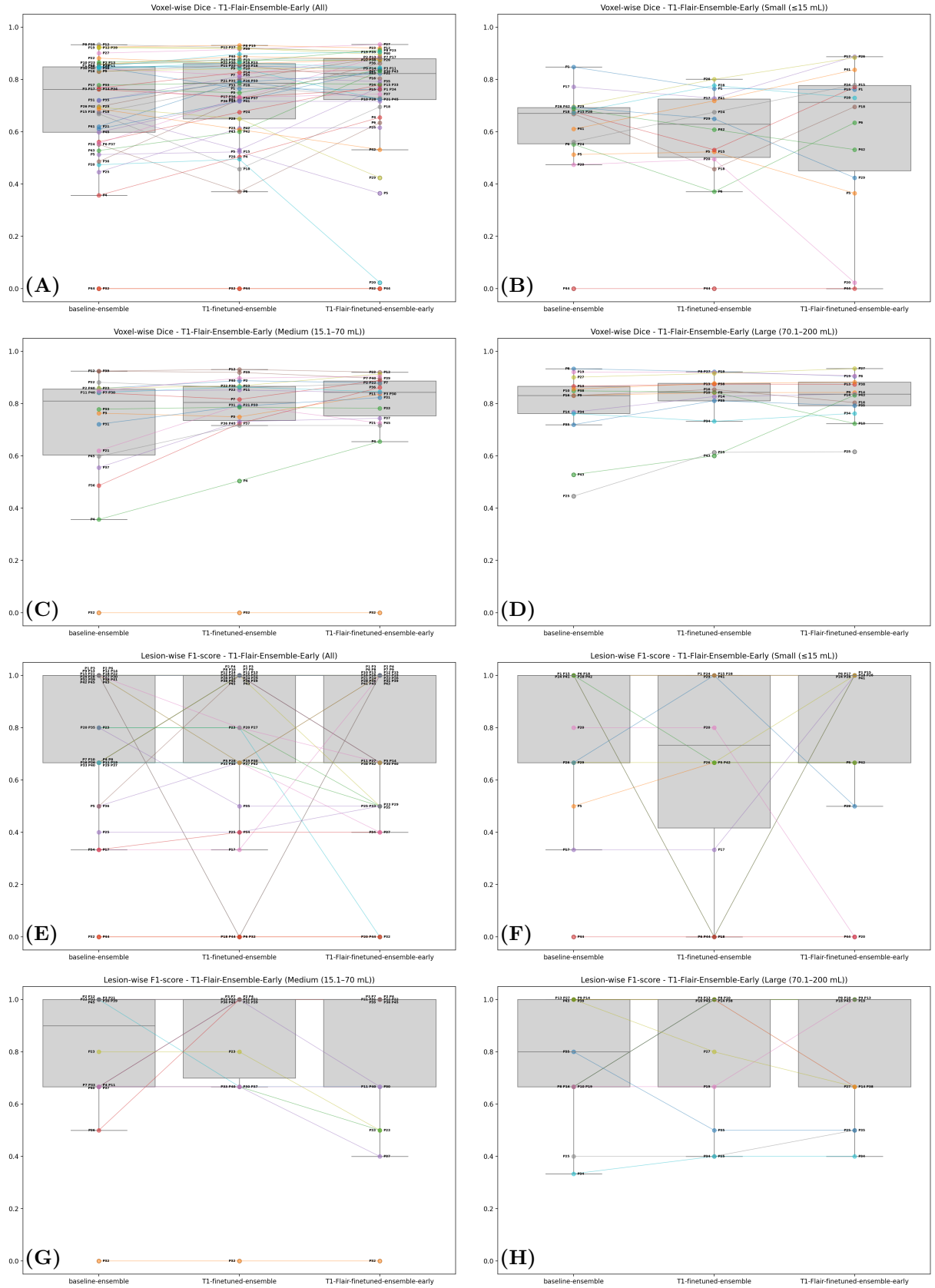


**Figure A.11:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the Averaged models (Late fusion) across all lesion volume categories (All, Small, Medium, Large).

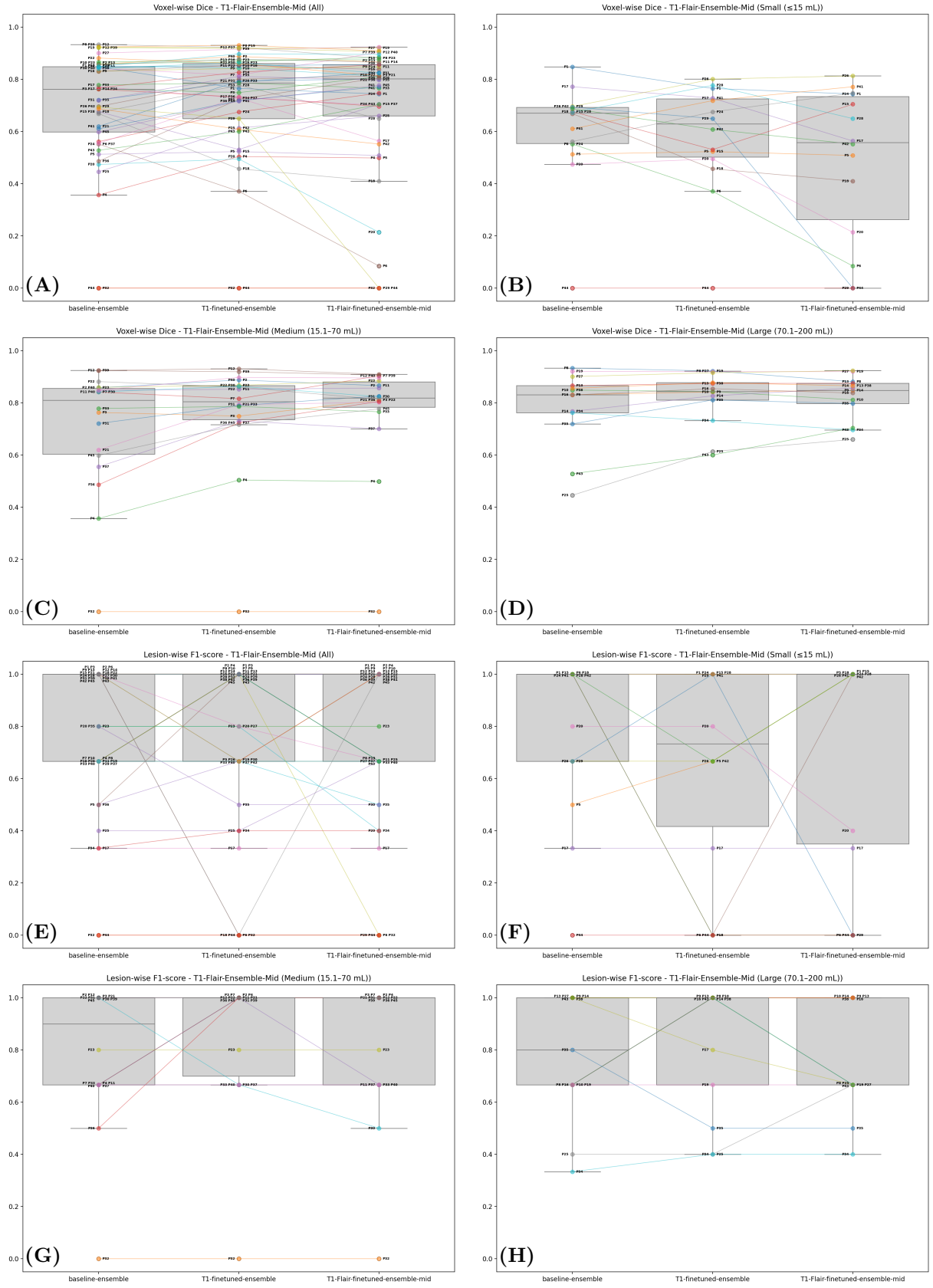




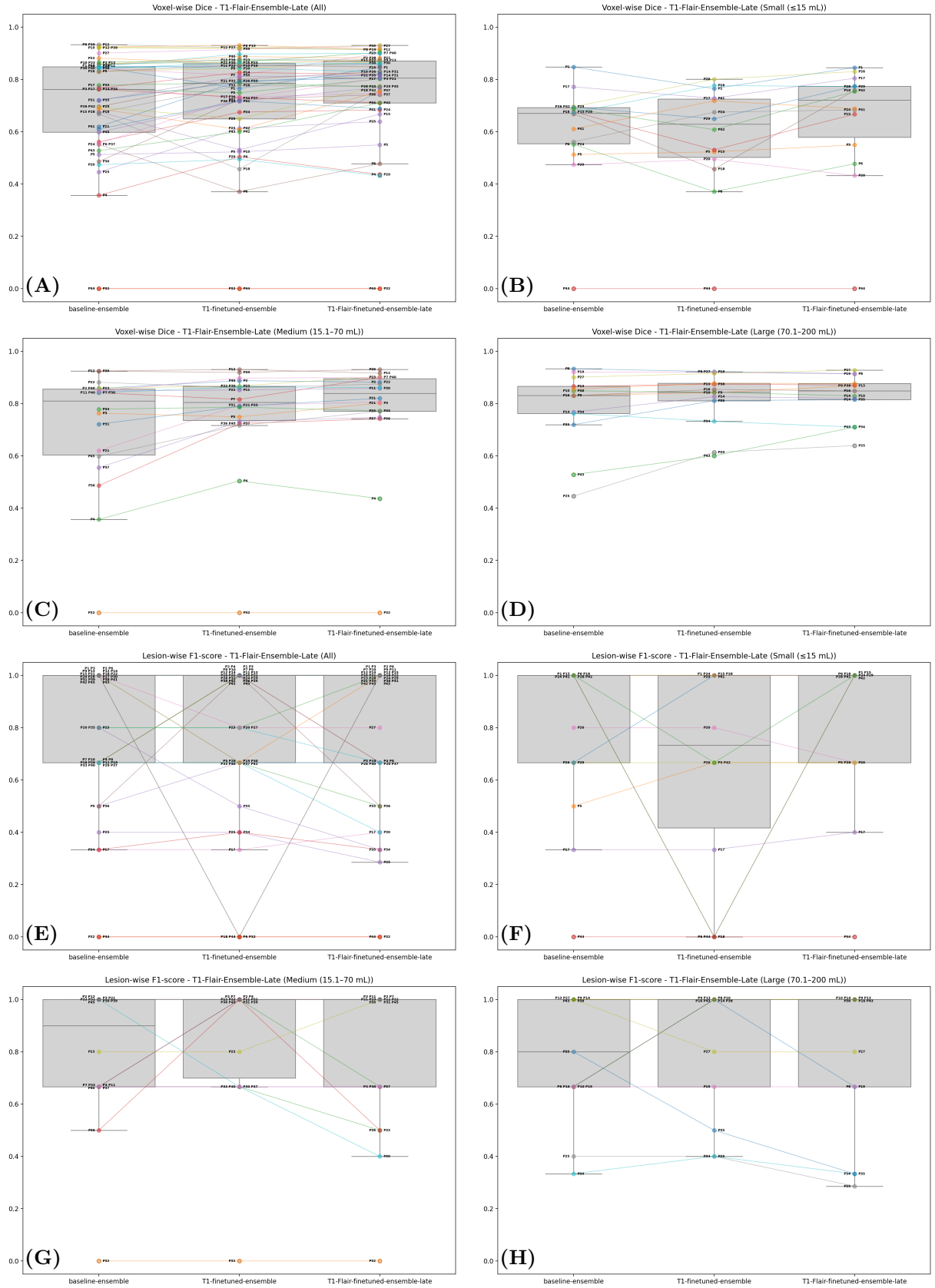
**Figure A.12:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the Averaged ensemble (T1+FLAIR Early, Mid, and Late fusion) models across all lesion volume categories (All, Small, Medium, Large).



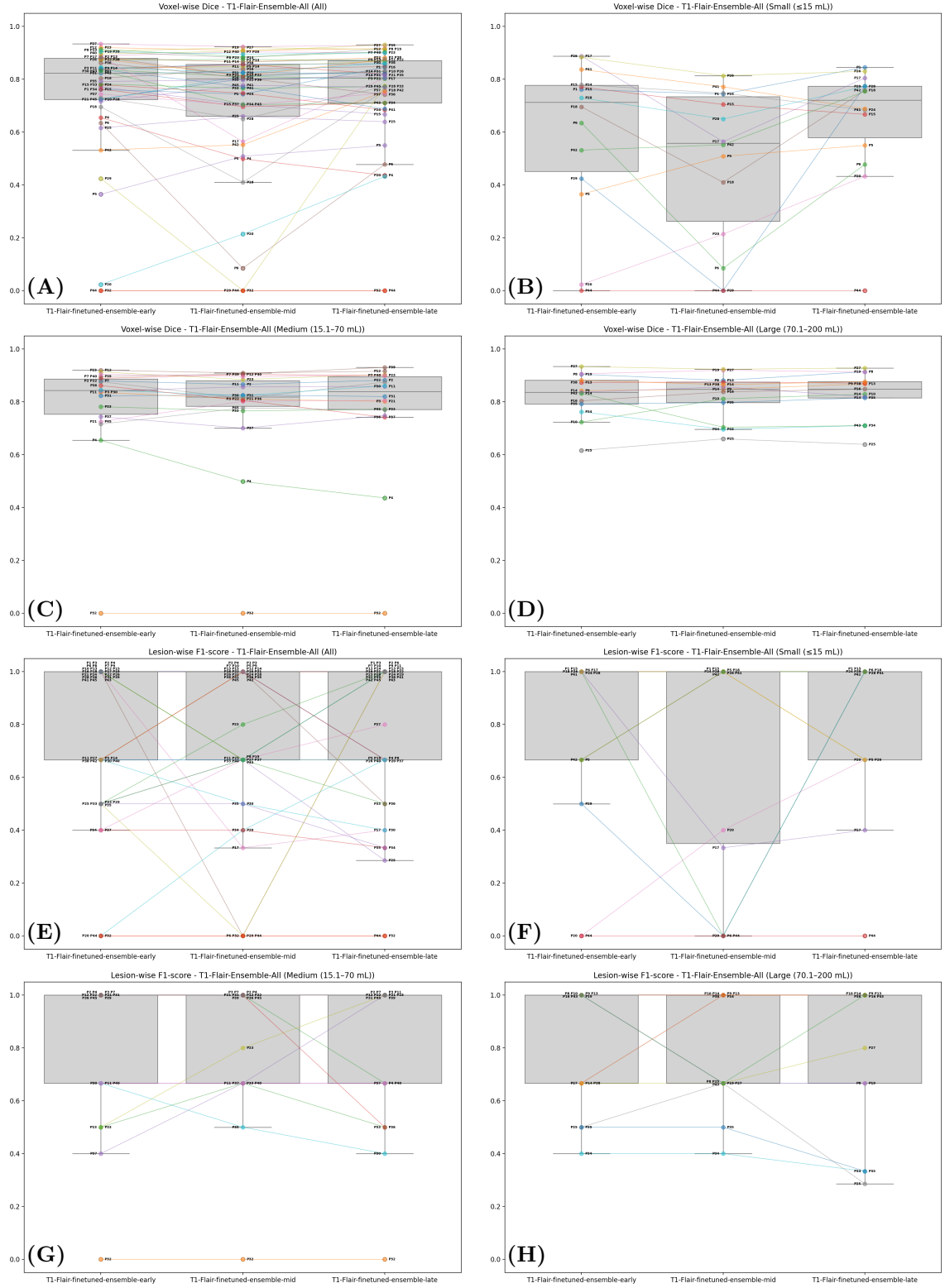
**Figure A.13:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble models (Early fusion) across all lesion volume categories (All, Small, Medium, Large).



**Figure A.14:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble models (Mid fusion) across all lesion volume categories (All, Small, Medium, Large).



**Figure A.15:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble models (Late fusion) across all lesion volume categories (All, Small, Medium, Large).



**Figure A.16:** Voxel-wise Dice (A-D) and Lesion-wise F1-score (E-H) for the ensemble models (T1+FLAIR Early, Mid, and Late fusion) across all lesion volume categories (All, Small, Medium, Large).

**Table A.2:** Statistical comparison (paired two-sided Wilcoxon signed-rank tests) for early fusion models using voxel-wise Dice. Bold indicates  $p < 0.05$ .

| Model Comparison                                                      | Overall        | Small         | Medium         | Large         |
|-----------------------------------------------------------------------|----------------|---------------|----------------|---------------|
| T1+FLAIR (early fusion) vs Baseline                                   | 0.0617         | 0.6999        | <b>0.0045</b>  | 0.1698        |
| T1+FLAIR (early fusion) vs Baseline (ensemble)                        | 0.1550         | 0.8756        | <b>0.0134</b>  | 0.2709        |
| T1+FLAIR (early fusion) vs T1 fine-tuned                              | 0.5326         | 0.5966        | 0.1733         | 0.8781        |
| T1+FLAIR (early fusion) vs T1 fine-tuned (ensemble)                   | 0.8876         | 0.8088        | 0.5507         | 0.9364        |
| T1+FLAIR (early fusion) vs T1+FLAIR (mid fusion)                      | <b>0.0145</b>  | <b>0.0374</b> | 0.0770         | 0.6576        |
| T1+FLAIR (early fusion) vs T1+FLAIR (late fusion)                     | 0.9193         | 0.9020        | 0.6170         | 0.9045        |
| T1+FLAIR (early fusion) vs T1+FLAIR (early fusion, ensemble)          | 1.0000         | 0.9964        | 0.9972         | 0.9960        |
| T1+FLAIR (early fusion) vs T1+FLAIR (mid fusion, ensemble)            | 0.2798         | 0.2108        | 0.2475         | 0.8473        |
| T1+FLAIR (early fusion) vs T1+FLAIR (late fusion, ensemble)           | 0.9908         | 0.9835        | 0.8376         | 0.9529        |
| T1+FLAIR (early fusion, ensemble) vs Baseline                         | <b>0.0015</b>  | 0.2762        | <b>5.0e-04</b> | 0.0732        |
| T1+FLAIR (early fusion, ensemble) vs Baseline (ensemble)              | <b>0.0067</b>  | 0.4583        | <b>0.0032</b>  | 0.1219        |
| T1+FLAIR (early fusion, ensemble) vs T1 fine-tuned                    | <b>0.0124</b>  | 0.1244        | <b>0.0176</b>  | 0.3934        |
| T1+FLAIR (early fusion, ensemble) vs T1 fine-tuned (ensemble)         | 0.1495         | 0.2315        | 0.1048         | 0.5537        |
| T1+FLAIR (early fusion, ensemble) vs T1+FLAIR (early fusion)          | <b>0.0e+00</b> | <b>0.0036</b> | <b>0.0033</b>  | <b>0.0052</b> |
| T1+FLAIR (early fusion, ensemble) vs T1+FLAIR (mid fusion)            | <b>7.0e-04</b> | <b>0.0139</b> | <b>0.0156</b>  | 0.2709        |
| T1+FLAIR (early fusion, ensemble) vs T1+FLAIR (late fusion)           | 0.3018         | 0.4583        | 0.1906         | 0.5270        |
| T1+FLAIR (early fusion, ensemble) vs T1+FLAIR (mid fusion, ensemble)  | <b>0.0189</b>  | <b>0.0320</b> | 0.0813         | 0.6066        |
| T1+FLAIR (early fusion, ensemble) vs T1+FLAIR (late fusion, ensemble) | 0.7066         | 0.7466        | 0.5658         | 0.7513        |

**Table A.3:** Statistical comparison (paired two-sided Wilcoxon signed-rank tests) for mid fusion models using voxel-wise Dice. Bold indicates  $p < 0.05$ .

| Model Comparison                                                     | Overall        | Small         | Medium         | Large          |
|----------------------------------------------------------------------|----------------|---------------|----------------|----------------|
| T1+FLAIR (mid fusion) vs Baseline                                    | 0.2533         | 0.9497        | <b>0.0139</b>  | 0.0955         |
| T1+FLAIR (mid fusion) vs Baseline (ensemble)                         | 0.4952         | 0.9804        | <b>0.0276</b>  | 0.2072         |
| T1+FLAIR (mid fusion) vs T1 fine-tuned                               | 0.9374         | 0.9727        | 0.4529         | 0.6576         |
| T1+FLAIR (mid fusion) vs T1 fine-tuned (ensemble)                    | 0.9860         | 0.9904        | 0.6563         | 0.8633         |
| T1+FLAIR (mid fusion) vs T1+FLAIR (early fusion)                     | 0.9855         | 0.9626        | 0.9292         | 0.3677         |
| T1+FLAIR (mid fusion) vs T1+FLAIR (late fusion)                      | 0.9980         | 0.9934        | 0.9803         | 0.9161         |
| T1+FLAIR (mid fusion) vs T1+FLAIR (early fusion, ensemble)           | 0.9993         | 0.9861        | 0.9844         | 0.7513         |
| T1+FLAIR (mid fusion) vs T1+FLAIR (mid fusion, ensemble)             | 1.0000         | 0.9626        | 0.9998         | 1.0000         |
| T1+FLAIR (mid fusion) vs T1+FLAIR (late fusion, ensemble)            | 0.9999         | 0.9971        | 0.9824         | 0.9933         |
| T1+FLAIR (mid fusion, ensemble) vs Baseline                          | 0.0626         | 0.8756        | <b>0.0043</b>  | 0.0636         |
| T1+FLAIR (mid fusion, ensemble) vs Baseline (ensemble)               | 0.1731         | 0.9421        | <b>0.0123</b>  | 0.0839         |
| T1+FLAIR (mid fusion, ensemble) vs T1 fine-tuned                     | 0.3722         | 0.7685        | 0.1544         | 0.3934         |
| T1+FLAIR (mid fusion, ensemble) vs T1 fine-tuned (ensemble)          | 0.8077         | 0.9421        | 0.3437         | 0.5803         |
| T1+FLAIR (mid fusion, ensemble) vs T1+FLAIR (early fusion)           | 0.7202         | 0.7892        | 0.7659         | 0.1698         |
| T1+FLAIR (mid fusion, ensemble) vs T1+FLAIR (mid fusion)             | <b>0.0e+00</b> | <b>0.0374</b> | <b>2.0e-04</b> | <b>1.0e-04</b> |
| T1+FLAIR (mid fusion, ensemble) vs T1+FLAIR (late fusion)            | 0.9039         | 0.9727        | 0.5283         | 0.5537         |
| T1+FLAIR (mid fusion, ensemble) vs T1+FLAIR (early fusion, ensemble) | 0.9811         | 0.9680        | 0.9187         | 0.4197         |
| T1+FLAIR (mid fusion, ensemble) vs T1+FLAIR (late fusion, ensemble)  | 0.9954         | 0.9920        | 0.8341         | 0.8781         |

**Table A.4:** Statistical comparison (paired two-sided Wilcoxon signed-rank tests) for late fusion models using voxel-wise Dice. Bold indicates  $p < 0.05$ .

| Model Comparison                                                      | Overall        | Small         | Medium         | Large          |
|-----------------------------------------------------------------------|----------------|---------------|----------------|----------------|
| T1+FLAIR (late fusion) vs Baseline                                    | <b>1.0e-04</b> | 0.0579        | <b>7.0e-04</b> | <b>0.0471</b>  |
| T1+FLAIR (late fusion) vs Baseline (ensemble)                         | <b>0.0067</b>  | 0.5966        | <b>0.0015</b>  | 0.0839         |
| T1+FLAIR (late fusion) vs T1 fine-tuned                               | <b>0.0099</b>  | 0.0503        | <b>0.0420</b>  | 0.4463         |
| T1+FLAIR (late fusion) vs T1 fine-tuned (ensemble)                    | 0.2770         | 0.3248        | 0.2174         | 0.7061         |
| T1+FLAIR (late fusion) vs T1+FLAIR (early fusion)                     | 0.0807         | 0.0980        | 0.3994         | 0.1082         |
| T1+FLAIR (late fusion) vs T1+FLAIR (mid fusion)                       | <b>0.0020</b>  | <b>0.0066</b> | <b>0.0197</b>  | 0.0955         |
| T1+FLAIR (late fusion) vs T1+FLAIR (early fusion, ensemble)           | 0.6982         | 0.5417        | 0.8094         | 0.5000         |
| T1+FLAIR (late fusion) vs T1+FLAIR (mid fusion, ensemble)             | 0.0961         | <b>0.0273</b> | 0.4717         | 0.4730         |
| T1+FLAIR (late fusion) vs T1+FLAIR (late fusion, ensemble)            | 1.0000         | 0.9985        | 0.9990         | 1.0000         |
| T1+FLAIR (late fusion, ensemble) vs Baseline                          | <b>0.0e+00</b> | <b>0.0054</b> | <b>2.0e-04</b> | <b>0.0164</b>  |
| T1+FLAIR (late fusion, ensemble) vs Baseline (ensemble)               | <b>0.0e+00</b> | <b>0.0165</b> | <b>6.0e-04</b> | <b>0.0402</b>  |
| T1+FLAIR (late fusion, ensemble) vs T1 fine-tuned                     | <b>1.0e-04</b> | <b>0.0054</b> | <b>0.0074</b>  | 0.0636         |
| T1+FLAIR (late fusion, ensemble) vs T1 fine-tuned (ensemble)          | <b>0.0020</b>  | <b>0.0096</b> | <b>0.0308</b>  | 0.4730         |
| T1+FLAIR (late fusion, ensemble) vs T1+FLAIR (early fusion)           | <b>0.0092</b>  | <b>0.0165</b> | 0.1733         | 0.0549         |
| T1+FLAIR (late fusion, ensemble) vs T1+FLAIR (mid fusion)             | <b>1.0e-04</b> | <b>0.0029</b> | <b>0.0176</b>  | <b>0.0085</b>  |
| T1+FLAIR (late fusion, ensemble) vs T1+FLAIR (late fusion)            | <b>0.0e+00</b> | <b>0.0015</b> | <b>0.0010</b>  | <b>1.0e-04</b> |
| T1+FLAIR (late fusion, ensemble) vs T1+FLAIR (early fusion, ensemble) | 0.2934         | 0.2534        | 0.4342         | 0.2709         |
| T1+FLAIR (late fusion, ensemble) vs T1+FLAIR (mid fusion, ensemble)   | <b>0.0046</b>  | <b>0.0080</b> | 0.1659         | 0.1367         |

**Table A.5:** Statistical comparison (p-values) for non-fusion models (Lesion-wise F1-score). Bold indicates  $p < 0.05$ .

| Model Comparison                                | Overall        | Small         | Medium        | Large         |
|-------------------------------------------------|----------------|---------------|---------------|---------------|
| Baseline (ensemble) vs Baseline                 | <b>4.0e-04</b> | <b>0.0313</b> | <b>0.0092</b> | <b>0.0488</b> |
| T1 fine-tuned vs Baseline                       | <b>0.0102</b>  | 0.6630        | <b>0.0065</b> | <b>0.0186</b> |
| T1 fine-tuned vs Baseline (ensemble)            | 0.4658         | 0.7505        | 0.1548        | 0.3555        |
| T1 fine-tuned (ensemble) vs Baseline            | <b>0.0031</b>  | 0.6103        | <b>0.0053</b> | <b>0.0137</b> |
| T1 fine-tuned (ensemble) vs Baseline (ensemble) | 0.2632         | 0.8618        | 0.0786        | 0.1562        |
| T1 fine-tuned (ensemble) vs T1 fine-tuned       | <b>0.0080</b>  | 0.4167        | <b>0.0415</b> | <b>0.0195</b> |

**Table A.6:** Statistical comparison (p-values) for early fusion models (Lesion-wise F1-score). Bold indicates  $p < 0.05$ .

| Model Comparison                                                      | Overall       | Small         | Medium        | Large  |
|-----------------------------------------------------------------------|---------------|---------------|---------------|--------|
| T1+FLAIR (early fusion) vs Baseline                                   | 0.1773        | 0.9304        | <b>0.0151</b> | 0.0508 |
| T1+FLAIR (early fusion) vs Baseline (ensemble)                        | 0.8350        | 0.9539        | 0.3918        | 0.4414 |
| T1+FLAIR (early fusion) vs T1 fine-tuned                              | 0.7482        | 0.7130        | 0.8069        | 0.4229 |
| T1+FLAIR (early fusion) vs T1 fine-tuned (ensemble)                   | 0.9671        | 0.6395        | 0.9897        | 0.8975 |
| T1+FLAIR (early fusion) vs T1+FLAIR (mid fusion)                      | 0.2421        | 0.4295        | 0.4295        | 0.3108 |
| T1+FLAIR (early fusion) vs T1+FLAIR (late fusion)                     | 0.7119        | 0.9137        | 0.4120        | 0.2524 |
| T1+FLAIR (early fusion) vs T1+FLAIR (early fusion, ensemble)          | 0.9309        | 0.9938        | 0.1568        | 0.6816 |
| T1+FLAIR (early fusion) vs T1+FLAIR (mid fusion, ensemble)            | 0.8229        | 0.8072        | 0.7126        | 0.5098 |
| T1+FLAIR (early fusion) vs T1+FLAIR (late fusion, ensemble)           | 0.8138        | 0.9704        | 0.3282        | 0.3125 |
| T1+FLAIR (early fusion, ensemble) vs Baseline                         | <b>0.0385</b> | 0.3060        | 0.1649        | 0.0820 |
| T1+FLAIR (early fusion, ensemble) vs Baseline (ensemble)              | 0.3381        | 0.5420        | 0.3048        | 0.3975 |
| T1+FLAIR (early fusion, ensemble) vs T1 fine-tuned                    | 0.2980        | 0.1635        | 0.8075        | 0.3413 |
| T1+FLAIR (early fusion, ensemble) vs T1 fine-tuned (ensemble)         | 0.5454        | 0.1721        | 0.9661        | 0.7812 |
| T1+FLAIR (early fusion, ensemble) vs T1+FLAIR (early fusion)          | 0.0691        | <b>0.0062</b> | 0.8432        | 0.3340 |
| T1+FLAIR (early fusion, ensemble) vs T1+FLAIR (mid fusion)            | 0.0582        | <b>0.0251</b> | 0.6954        | 0.3760 |
| T1+FLAIR (early fusion, ensemble) vs T1+FLAIR (late fusion)           | 0.1836        | 0.3001        | 0.3835        | 0.2705 |
| T1+FLAIR (early fusion, ensemble) vs T1+FLAIR (mid fusion, ensemble)  | 0.3770        | 0.1721        | 0.9013        | 0.5000 |
| T1+FLAIR (early fusion, ensemble) vs T1+FLAIR (late fusion, ensemble) | 0.5884        | 0.7060        | 0.5000        | 0.3516 |

**Table A.7:** Statistical comparison (p-values) for mid fusion models (Lesion-wise F1-score). Bold indicates  $p < 0.05$ .

| Model Comparison                                                     | Overall       | Small  | Medium        | Large  |
|----------------------------------------------------------------------|---------------|--------|---------------|--------|
| T1+FLAIR (mid fusion) vs Baseline                                    | 0.2167        | 0.9227 | <b>0.0250</b> | 0.1372 |
| T1+FLAIR (mid fusion) vs Baseline (ensemble)                         | 0.8071        | 0.9691 | 0.2664        | 0.5083 |
| T1+FLAIR (mid fusion) vs T1+FLAIR (early fusion)                     | 0.7579        | 0.5705 | 0.5705        | 0.7156 |
| T1+FLAIR (mid fusion) vs T1+FLAIR (late fusion)                      | 0.8191        | 0.9453 | 0.3606        | 0.4922 |
| T1+FLAIR (mid fusion) vs T1+FLAIR (early fusion, ensemble)           | 0.9418        | 0.9749 | 0.3046        | 0.6445 |
| T1+FLAIR (mid fusion) vs T1+FLAIR (mid fusion, ensemble)             | 0.9298        | 0.8796 | 0.7988        | 0.7520 |
| T1+FLAIR (mid fusion) vs T1+FLAIR (late fusion, ensemble)            | 0.8919        | 0.9836 | 0.3944        | 0.4600 |
| T1+FLAIR (mid fusion) vs T1 fine-tuned                               | 0.6894        | 0.5707 | 0.7241        | 0.7217 |
| T1+FLAIR (mid fusion) vs T1 fine-tuned (ensemble)                    | 0.9576        | 0.6390 | 0.9845        | 0.9004 |
| T1+FLAIR (mid fusion, ensemble) vs Baseline                          | <b>0.0292</b> | 0.6323 | <b>0.0271</b> | 0.0850 |
| T1+FLAIR (mid fusion, ensemble) vs Baseline (ensemble)               | 0.5828        | 0.8274 | 0.2887        | 0.5469 |
| T1+FLAIR (mid fusion, ensemble) vs T1 fine-tuned                     | 0.2242        | 0.3360 | 0.2963        | 0.5078 |
| T1+FLAIR (mid fusion, ensemble) vs T1 fine-tuned (ensemble)          | 0.6546        | 0.4158 | 0.9101        | 0.8750 |
| T1+FLAIR (mid fusion, ensemble) vs T1+FLAIR (early fusion)           | 0.1771        | 0.1928 | 0.2874        | 0.5098 |
| T1+FLAIR (mid fusion, ensemble) vs T1+FLAIR (mid fusion)             | 0.0702        | 0.1204 | 0.2012        | 0.2852 |
| T1+FLAIR (mid fusion, ensemble) vs T1+FLAIR (late fusion)            | 0.2985        | 0.6323 | 0.1871        | 0.3711 |
| T1+FLAIR (mid fusion, ensemble) vs T1+FLAIR (early fusion, ensemble) | 0.6230        | 0.8279 | 0.0987        | 0.6562 |
| T1+FLAIR (mid fusion, ensemble) vs T1+FLAIR (late fusion, ensemble)  | 0.4811        | 0.7697 | 0.2641        | 0.4062 |

**Table A.8:** Statistical comparison (p-values) for late fusion models (Lesion-wise F1-score). Bold indicates  $p < 0.05$ .

| Model Comparison                                                      | Overall       | Small         | Medium        | Large  |
|-----------------------------------------------------------------------|---------------|---------------|---------------|--------|
| T1+FLAIR (late fusion) vs Baseline                                    | <b>0.0334</b> | 0.6004        | <b>0.0373</b> | 0.1763 |
| T1+FLAIR (late fusion) vs Baseline (ensemble)                         | 0.8033        | 0.8242        | 0.5000        | 0.6777 |
| T1+FLAIR (late fusion) vs T1 fine-tuned                               | 0.8175        | 0.2877        | 0.8938        | 0.7681 |
| T1+FLAIR (late fusion) vs T1 fine-tuned (ensemble)                    | 0.9740        | 0.5000        | 0.9750        | 0.9180 |
| T1+FLAIR (late fusion) vs T1+FLAIR (early fusion)                     | 0.2881        | 0.0863        | 0.5880        | 0.7612 |
| T1+FLAIR (late fusion) vs T1+FLAIR (mid fusion)                       | 0.1809        | 0.0547        | 0.6394        | 0.5303 |
| T1+FLAIR (late fusion) vs T1+FLAIR (early fusion, ensemble)           | 0.8164        | 0.6999        | 0.6165        | 0.7471 |
| T1+FLAIR (late fusion) vs T1+FLAIR (mid fusion, ensemble)             | 0.7015        | 0.3677        | 0.8129        | 0.6797 |
| T1+FLAIR (late fusion) vs T1+FLAIR (late fusion, ensemble)            | 0.8944        | 0.9861        | 0.5833        | 0.5273 |
| T1+FLAIR (late fusion, ensemble) vs Baseline                          | <b>0.0013</b> | <b>0.0140</b> | <b>0.0297</b> | 0.1377 |
| T1+FLAIR (late fusion, ensemble) vs Baseline (ensemble)               | 0.2744        | 0.1367        | 0.3422        | 0.5625 |
| T1+FLAIR (late fusion, ensemble) vs T1 fine-tuned                     | 0.3282        | 0.0534        | 0.7469        | 0.7578 |
| T1+FLAIR (late fusion, ensemble) vs T1 fine-tuned (ensemble)          | 0.6701        | 0.0681        | 0.9310        | 1.0000 |
| T1+FLAIR (late fusion, ensemble) vs T1+FLAIR (early fusion)           | 0.1862        | <b>0.0296</b> | 0.6718        | 0.7217 |
| T1+FLAIR (late fusion, ensemble) vs T1+FLAIR (mid fusion)             | 0.1081        | <b>0.0164</b> | 0.6056        | 0.5771 |
| T1+FLAIR (late fusion, ensemble) vs T1+FLAIR (late fusion)            | 0.1056        | <b>0.0139</b> | 0.4167        | 0.5273 |
| T1+FLAIR (late fusion, ensemble) vs T1+FLAIR (early fusion, ensemble) | 0.4116        | 0.2940        | 0.5000        | 0.6719 |
| T1+FLAIR (late fusion, ensemble) vs T1+FLAIR (mid fusion, ensemble)   | 0.5189        | 0.2303        | 0.7359        | 0.6875 |



## Part VIII

# Acknowledgement

## Acknowledgements

The research presented in this thesis was conducted at the A<sup>2</sup>VI-Lab (MeSVA), under the guidance of Prof. Giuseppe Placidi, and in collaboration with Prof. Filippo Mignosi (DISIM) at the University of L'Aquila.

I would like to express my deepest gratitude to both Prof. Placidi and Prof. Mignosi for their unwavering support and mentorship throughout my doctoral journey. Their expertise, insightful feedback, and dedication have been instrumental in shaping this work and my development as a researcher. I am profoundly grateful for the rigor they instilled in me and for the trust they placed in my work over the years. My sincere thanks also go to my colleagues at the A<sup>2</sup>VI-Lab; the vibrant and intellectually stimulating environment we shared was essential to the completion of this project.

A significant part of this work was carried out during my Erasmus+ stay at the Empenn Laboratory (Unité U1228, INSERM, Inria, Université de Rennes, France). I wish to extend my heartfelt thanks to Prof. Francesca Galassi for her exceptional supervision and for the warm welcome I received from the entire Empenn team. I am immensely grateful for their patience and for the significant investment they made in my training, which allowed me to greatly enhance my practical expertise in deep learning. This experience was profoundly enriching, both professionally and personally, opening my mind to new perspectives and cultures.

I wish to extend my sincere gratitude to the reviewers for their insightful feedback and rigorous critique. Their expertise was instrumental in refining the argumentative structure and enhancing the overall clarity of this thesis. The final document is more precise and accessible, ensuring a more effective communication of the research findings.

I am deeply grateful to my family for their unconditional support, patience, and understanding. Their encouragement has been a constant source of strength, especially during the most challenging moments of this journey. This achievement would not have been possible without their unwavering belief in me.

Last but not least, I would like to thank my partner Giulia, who has supported me throughout these years with love, patience, and understanding, and our daughter Matilde, the greatest gift life has given us.

Alessandro Di Matteo

---

