

Deliverable D1.1

SETTING UP SL CAPABLE COMPUTE INFRASTRUCTURE

Lead beneficiary	TUD
Author(s)	Oliver Saldanha
Dissemination level	PU
Type	R
Delivery date	28/4/23

*ODELIA is funded by the European Union's Horizon Europe Framework
under Grant Agreement 101057091*



**Funded by
the European Union**

TABLE OF CONTENTS

Introduction	3
VPN Setup	3
Initial Dataset for testing.....	4
Github repositories	4
Initial training success.....	4

DISCLAIMER

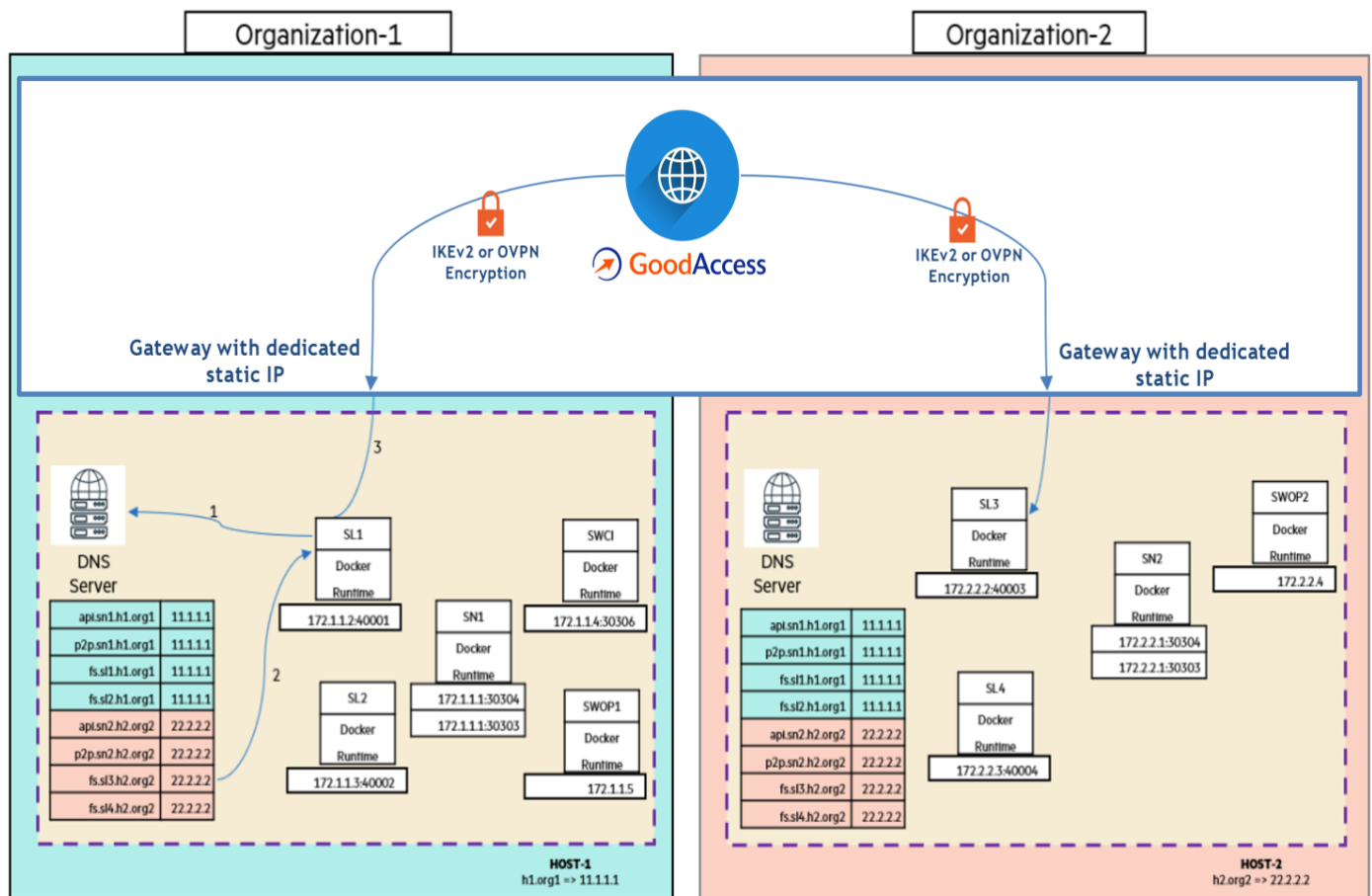
Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

INTRODUCTION

This progress report outlines several significant achievements in the ODELIA D1.1 phase.

VPN SETUP

Firstly, the Virtual Private Networks (VPN) setup with swarm learning has been completed, which is a critical step in enabling secure and efficient communication between the different nodes. Good Access has been selected as the VPN provider due to its reliable Gateway with a static IP and support for Internet Key Exchange version 2 (IKEv2) and Open Source Virtual Private Network (OpenVPN) encryption. A static IP address is important in VPNs because it provides a fixed, unique address that can be used to identify and connect to a specific device or network in our setup every centre will have its own IP. This makes it easier to manage and configure the VPN, and ensures that you can maintain a consistent connection and access resources on the network without interruption. This VPN setup is ideal as it provides a mesh VPN architecture that aligns with the swarm architecture used, ensuring that traffic is not routed through a central host, which is important for security and efficiency. Encryption is important in VPNs because it provides a secure and private connection between the user's device and the VPN server, protecting their data from interception or unauthorized access. A detailed network diagram shown below has been prepared and submitted to each center's IT department for approval.



INITIAL DATASET FOR TESTING

Completing the data pre-processing of the DUKE dataset is essential to ensure accurate and reliable results in the training process. The dataset is openly available and was collected between 2000 and 2014 at Duke University Hospital (Durham, North Carolina, USA), and includes 922 cases of biopsy-confirmed invasive breast cancer. The MRI protocol involved a T1-weighted fat-suppressed sequence (one pre-contrast and four post-contrast scans) and a non-fat-suppressed T1-weighted sequence.

Data pre-processing involves stratifying the data, cropping out the left and right breasts and saving them separately, as tumours mostly appear in only half of the breast. The transformation of data from Digital Imaging and Communications in Medicine (DICOM) files to Neuroimaging Informatics Technology Initiative (NIFTI) files and splitting it into the left and right sides of the breast is important in medical imaging, particularly in breast cancer screening. DICOM files are not always compatible with all software tools and analysis methods. Therefore, transforming DICOM files into NIFTI files, which are widely used in machine learning methods. Breast cancer is often asymmetric, meaning it can occur in one breast but not the other, or can be more pronounced in one breast than the other. Splitting the image into left and right provides us with more precise data. Overall, transforming DICOM files to NIFTI files and splitting the data into the left and right sides of the breast is an important step in medical imaging analysis and can help with the detection and diagnosis of breast cancer, ultimately leading to better patient outcomes.

To standardize the data, it is resampled to a uniform shape.

The processed sequence is utilized to compute the pre-contrast and post-contrast subtraction images along with the resampled T1-weighted image.

Github repositories

A Github repository for the pre-processing pipeline has been created by UKA (https://github.com/mueller-franzes/odelia_breast_mri/tree/main/scripts/preprocessing).

The pre-processed data from the DUKE dataset for MRI breast cancer is trained to obtain a model for tumour detection. Multiple techniques such as 3 Dimensional and 2 Dimensional Convolutional Neural Network (CNN) and Attention-based Multiple Instance Learning (AttMIL) were tested, and a Resnet-based 3D (CNN) outperformed the other techniques.

A Github repository for model training and swarm learning integration of the pipeline has been created by UKA/TUD (<https://github.com/KatherLab/swarm-learning-hpe>).

INITIAL TRAINING SUCCESS

Involving four nodes in joint training on the DUKE dataset, including VHIO, TUD, Radboud, and UKA, is a major accomplishment that demonstrates successful collaboration and coordination. The project has been split between five centres, including TUD, and successfully completed swarm training for tumour detection.

Finally, the other centre's are ordering hardware and waiting for delivery is a promising sign that the project will continue to expand and involve more collaborators.

Overall, this progress report indicates excellent progress in the project, and it is exciting to see what the future holds for the team.