ODELIA

Open Consortium for Decentralized Medical Artificial Intelligence

HORIZON-HLTH-2021-CARE-05-02

Deliverable D1.1

GUIDANCE DOCUMENT ON RUNNING SL IN CLINICAL ENVIRONMENTS

Lead beneficiary	TUD
Author(s)	Oliver Saldanha
Dissemination level	PU
Type	R
Delivery date	21/09/2023

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

Funded by the European Union

ODELIA

TABLE OF CONTENTS

DISCLAIMER

SUMMARY

This progress report outlines several significant achievements in the ODELIA D1.2 phase.

INTRODUCTION

Swarm Learning, an innovative approach to machine learning in healthcare, expands upon the foundational principles of federated learning by incorporating a decentralizing element, thereby eliminating the requirement for a singular, authoritative entity. This strategy leverages the power of Artificial Intelligence (AI), the immediacy of edge computing, and the robust security framework provided by blockchain technology. In summary, Swarm Learning represents a forward-thinking, privacy-centric Machine Learning (ML) model that decentralizes the learning process. At its core, the Swarm Learning framework utilizes the computational resources located at or in close proximity to the distributed data points to execute the Machine Learning algorithms that are responsible for model training. The security paradigm of blockchain technology is employed to facilitate safe and secure peer learning and knowledge sharing. In the Swarm Learning paradigm, the training process of the ML models takes place at the edge of the network, where the data is most fresh and where timely, data-driven decisions are of paramount importance. This decentralization facilitates a unique information-sharing environment where only the distilled insights from the ML model training, not the raw data, are shared amongst collaborating ML peers. This not only bolsters data security but also enhances the privacy of the data involved, making Swarm Learning an optimal choice for sensitive environments like clinical settings.

HARDWARE REQUIREMENTS OF THE MACHINE

The following hardware requirement is recommended for the MRI Brest cancer tumor prediction:

- **RAM**: At least 32 GB of RAM is required, but ideally, 64 GB of RAM should be used for optimal performance.
- **CPU:** A minimum of 8 CPU cores is required, but it is recommended to have 16 CPU cores for better efficiency.
- **GPU:** An NVIDIA GPU with a minimum of 24 GB of RAM is required, but for improved performance, a GPU with 48 GB of RAM is recommended.
- **Storage:** The absolute minimum storage requirement is 4 TB, but for better data management and storage capacity, it is recommended to have 8 TB of storage.

By demonstrating that medical report generation can be accomplished with these lightweight hardware specifications, it highlights the feasibility and accessibility of the task.

SOFTWARE REQUIREMENTS OF THE MACHINE

Operating system: To successfully run the Swarm Learning Environment on the Linux-qualified on Ubuntu 20.04 Operating system, the following recommendations and compatibility information should be considered:

- Supported Ubuntu Versions: The Swarm Learning Environment has been tested and confirmed to work on the following Ubuntu versions: Ubuntu 20.04 LTS, Ubuntu 22.04.2 LTS, and Ubuntu 20.04.5 LTS.
- Avoid Experimental Releases: It is advised to avoid using experimental releases of Ubuntu beyond the LTS 20.04 version, as they may lead to unsuccessful operation of the swop node.

Container hosting platform: HPE Swarm Learning is optimized for Docker 20.10.5, ensuring compatibility with IPv4. It's advisable to run Docker as a non-root user for security, and configuring network proxy settings is made straightforward. These steps enhance the functionality and security of HPE Swarm Learning in collaborative machine learning setups.

Machine Learning Framework: We used PyTorch 1.5-based Machine Learning models implemented but also qualify with Keras 2.9.0 (TensorFlow 2 backend) and using Python3.

NETWORK REQUIREMENTS OF THE MACHINE

To establish a secure swarm learning environment in a hospital, several critical factors must be considered, particularly network and infrastructure settings. First and foremost, it is imperative that hospitals maintain a distinct network environment for swarm learning to thwart any unauthorized access. This separation is pivotal in reducing security risks, ensuring that the swarm learning infrastructure remains isolated from the hospital's clinical network. To fortify the security posture, the implementation of robust firewalls and intrusion detection systems is paramount, offering protection against potential cyber threats. Moreover, adopting a mesh-based VPN architecture guarantees the encryption of all data, reinforcing its safety and privacy.

- A minimum of one or a maximum four open TCP/IP ports in each node. All swarm nodes must be able to access the ports of every other node. For more information on port details that must be opened.
- Stable internet connectivity to download Swarm Learning package and Docker images.

DATA PRIVACY AND SECURITY

Data Anonymization: Patient data used for swarm learning should be anonymized and de-identified to comply with healthcare privacy regulations, such as HIPAA (in the United States) or GDPR (in Europe).
Access Control: Implement strict access controls and authentication mechanisms to ensure that only authorized personnel can access and use patient data for swarm learning.

CONTACT AND ROLES

When setting up a swarm learning environment in a hospital, it's essential to identify and establish contacts for each site involved in the collaboration. Here's a breakdown of relevant contacts and their roles:

Research Contact: This contact is typically a healthcare professional or researcher responsible for coordinating research efforts at the hospital. They oversee data collection, data quality, and research objectives.
IT/Technical Contact: The IT/Technical contact is responsible for managing the technical infrastructure, network, and software required for swarm learning. They ensure that the IT systems are secure, scalable, and compliant with privacy regulations.
Security or Chief Information Security Officer (CISO): The CISO or security contact is responsible for ensuring the security of the hospital's digital assets, including patient data. They oversee cybersecurity measures, risk assessments, and incident response planning.
Data Protection Officer (DPO): In compliance with regulations like GDPR, a Data Protection Officer is responsible for ensuring that the hospital's data processing activities, including those related to swarm learning, comply with data protection laws. They oversee data privacy and compliance efforts.

CONCLUSION

These are the guidelines that must be adhered to when implementing swarm learning within a clinical setup.