Performance Report of UMSAEP grant (2022-2023)

Title: Federated Machine Learning

PIs: Sajal K. Das (CS, Missouri S&T) and Antoine Bagula (CS, UWC)

I. Executive Summary

The UMSAEP project between Missouri S&T and UWC is at the forefront of Federated Machine Learning (FML), a paradigm shift in machine learning, designed to tackle the complexities of secure and efficient analysis of vast datasets. FML ensures data privacy by decentralizing the learning process, facilitating collaborative model training across diverse sites without the need for direct dataset sharing. FML models find applications in critical real-life scenarios, including healthcare, agriculture, waterways, energy, smart cities, and smart villages. The joint project places particular emphasis on the deployment of FML models on lightweight IoT devices, offering relevance to developing countries. In the healthcare sector, FML successfully addresses such challenges as resource constraints, data scarcity, and regulatory compliance aligned with the United Nations Sustainable Development Goals (SDGs). Building upon the foundations of key studies, the project incorporates techniques from related works in the literature, and innovatively addresses issues of heterogeneity and labelled data scarcity within the FML framework. The project delves into the evolution of Federated Learning (FL), pinpointing its limitations, especially in handling unlabelled data. Extensions to FL, such as domain adaptation and self-supervised learning, have been explored, opening avenues for enhanced learning in decentralized settings. Outlined within three strategic work packages, the project aimed to advance FML:

1. **Federated Learning Security:** Focuses on implementing privacy-preserving techniques, including differential privacy and homomorphic encryption, ensuring computational efficiency and safeguarding sensitive data.
2. **Federated Learning Communication:** Investigates communication-efficient strategies such as local updating, model compression, and decentralized training to address scalability challenges in large-scale machine learning deployments.
3. **Federated Learning Deployment:** Tackles challenges related to deploying FML in diverse environments, placing emphasis on statistical heterogeneity, asynchronous communication, and fault tolerance to ensure robust and adaptable deployment.

The collaborative efforts between UWC and MST in 2022-2023 yielded two significant FML models: (a) FedFaSt, a selective federated learning algorithm (published in the flagship IEEE Globecom 2023 conference); and (b) FedDAFL, a federated transfer learning approach (to be submitted to IEEE Globecom 2024 by the deadline of April 15, 2024). These models demonstrated superior performance across key metrics, including accuracy, convergence, time complexity, and resilience against attacks and frugal labelling challenges. Ultimately, the project aspires to offer innovative solutions at the intersection of FML, sustainability, and technology. By addressing complex challenges and promoting responsible and efficient machine learning practices, it contributes to the global endeavour for progress and positive impact. The UMASEP also facilitated visit exchanges between Missouri S&T and UWC, leading to further collaboration opportunities, as summarized below.

II. Outcomes

The collaboration between Sajal K. Das (MST) and Antoine Bagula (UWC) resulted in several fruitful and tangible outcomes.

Visit Exchanges:

(i) Bagula visited Das at MST for two weeks in October 2022. He delivered a seminar talk and interacted with Das' colleagues, postdocs, and Ph.D. students.

(ii) Das visited UWC in March 2023 as part of the UM System's delegation team to present an invited talk at the Precision Medicine Workshop held at UWC. Additionally, Das spent a week at UWC to further collaboration with Bagula and his Ph.D. student Ferdinand Kahenga. Das also explored collaboration with Dr. Jacques Joubert in the School of Pharmacy.

(iii) Das paid another visit to UWC in August 2023 and worked intensely with Bagula and his PhD student Ferdinand Kahenga that led to a research publication in IEEE Globecom 2023 (see below). He further interacted with Dr. Jacques Joubert in the School of Pharmacy at UWC. Additionally, Das visited the University of Cape Town (UCT), delivered a seminar talk, and explored collaboration with Dr. Josiah Chavula and his Ph.D. student Emmanuel Ackerson.

Publications:

The collaboration between Das and Bagula focused on two aspects of federated learning modeling: Horizontal Federated Learning and Federated Transfer Learning. Two significant outputs emerged from the collaboration:

1. F. Kahenga, A. Bagula, and S. K. Das, "FedFaSt: Selective Federated Learning using Fittest Parameters Aggregation and Slotted Clients Training," *Proceedings of the IEEE Conference on Global Communications* (GlobeCom), Kuala Lumpur, Malaysia, Dec 2023. (**DOI:** 10.1109/GLOBECOM54140.2023.10437003)

   This paper proposes a novel selective federate learning (FL) algorithm, called fittest aggregation and slotted training (FedFaSt) that relies on a "free-for-all" client training process to score the clients' efficiency while applying the "natural selection" principle to elect the fittest clients to be used in the FL training and aggregation processes. Utilizing a combined data quality and client performance metric for scoring clients, FedFaSt implements a slotted training model enabling teams of fittest clients to participate in the training and aggregation processes for a fixed number of successive rounds, called slots. Performance validation using X-rays datasets reveal that FedFaSt outperforms other selective federated learning algorithms (e.g., FedAvg, FedRand, and FedPow) in terms of accuracy, convergence to the global optimum, time complexity, and robustness against attacks.

2. F. Kahenga, A. Bagula, and S. K. Das, “FedDAFL: Federated Transfer Learning Using Domain Adaptation with Frugal Labelling” (To be submitted to IEEE Globecom 2024 by April 15, 2024)

   This paper introduces FedDAFL, a novel approach to label-scarce federated settings focusing on scenarios with frugal labelling. Employing a teacher-learner paradigm, FedDAFL enables learners with unlabeled datasets to leverage fully labelled teachers. The method incorporates domain adaptation techniques within a multi-teacher to multi-learner framework, accommodating totally, partially, or not labeled client nodes. Combining self-learning and semi-supervised methods, FedDAFL focuses on enhancing model accuracy with each client’s unlabeled data. In comparison to existing approaches, FedDAFL superior performance across identically and non-identically distributed datasets (as seen in the Pneumonia Chest X-rays dataset), thus showcasing the effectiveness in addressing frugal labelling challenges in federated transfer learning.

Additional Outcomes:

(i) Das has been appointed by UWC in 2023 as an Extraordinary Professor for a period of 3 years.

(ii) Das has been formally invited as co-supervisor of Bagula’s Ph.D. student Ferdinand Kahenga

(iii) Recently Dr. Jacques Joubert received a funding from the Bill and Melinda Gates Foundation. This exciting project aims at developed AI-based solutions for advancing healthcare equity through language access in South Africa. He invited Das to collaborate on this effort.

(iv) Emanuel Ackerson at UCT has expressed interest in having Das as his Ph.D. co-supervisor.