PRIVACY-PRESERVING FEDERATED BIOMEDICAL ANALYTICS

Sinem Sav and Jean-Pierre Hubaux

School of Computer and Communication Sciences, EPFL

Introduction

- Biomedical analytics require a large amount of diverse data that is usually scattered across multiple healthcare institutions or hospitals.
- Data sharing among institutions is a must but often not feasible due to privacy concerns and strict regulations.
- We design a system, *PriCell*, for collaborative and privacy-preserving single-cell analysis for disease-associated cell classification with multiparty homomorphic encryption (MHE) [1].

Contributions

- We enable collaborative and privacy-preserving model training between institutions.
- Our solution does not degrade utility and preserve the data confidentiality for federated biomedical analytics.
- Our method is generalizable to various other tasks in the biomedical domain and beyond.

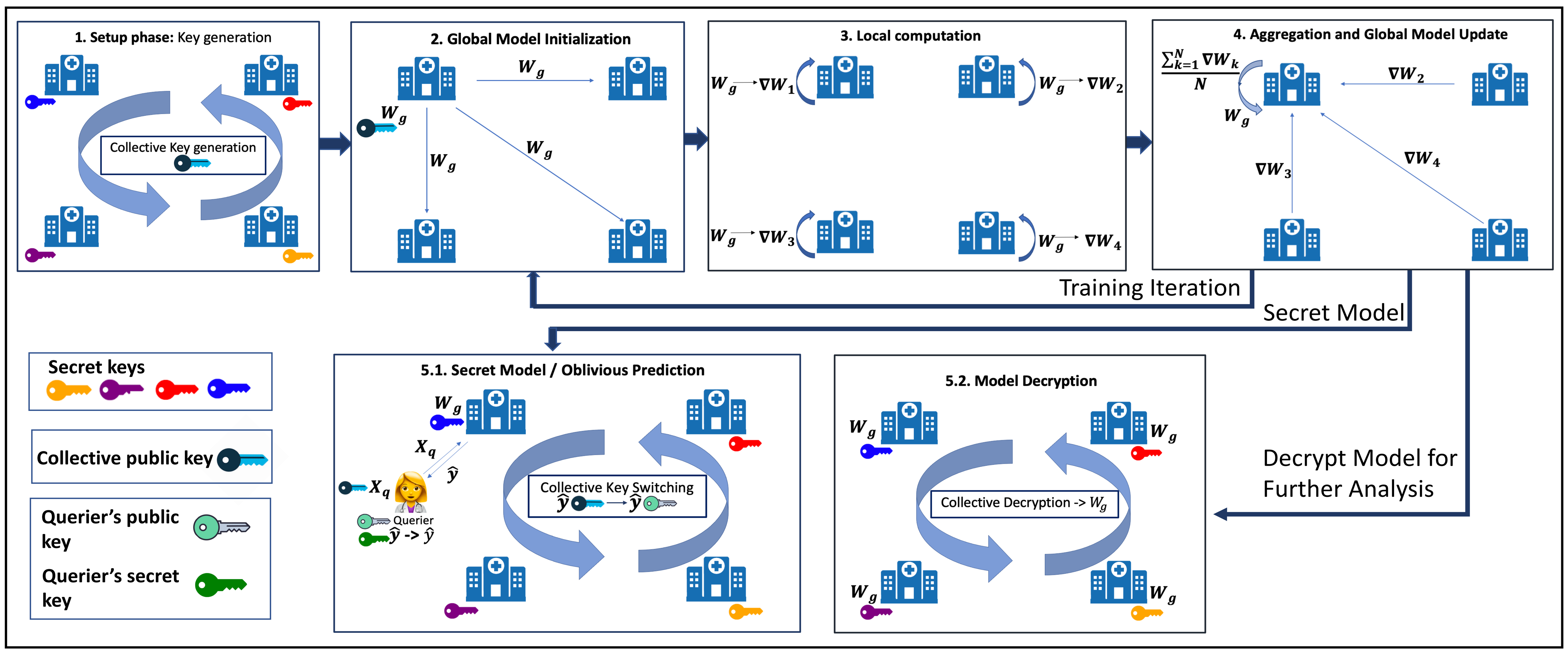
Acknowledgements

This work was partially supported by grant no. 2017-201 of the Strategic Focal Area “Personalized Health and Related Technologies (PHRT)” of the ETH Domain.

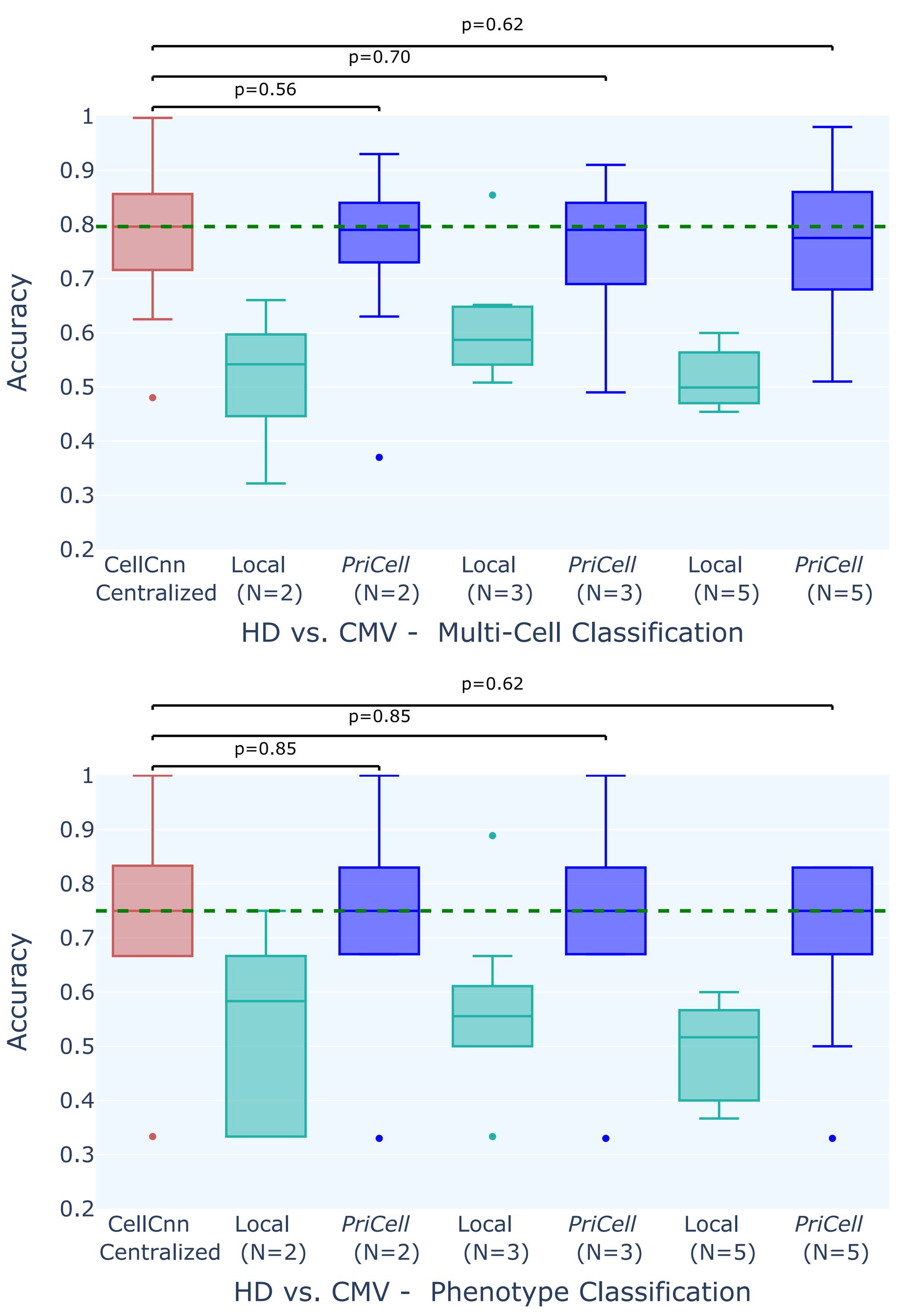
Method *

- The full analytics pipeline is performed under encryption.
- Scalable computations by relying on MHE.
- Various optimizations and approximations are introduced to enable efficient encrypted computation.

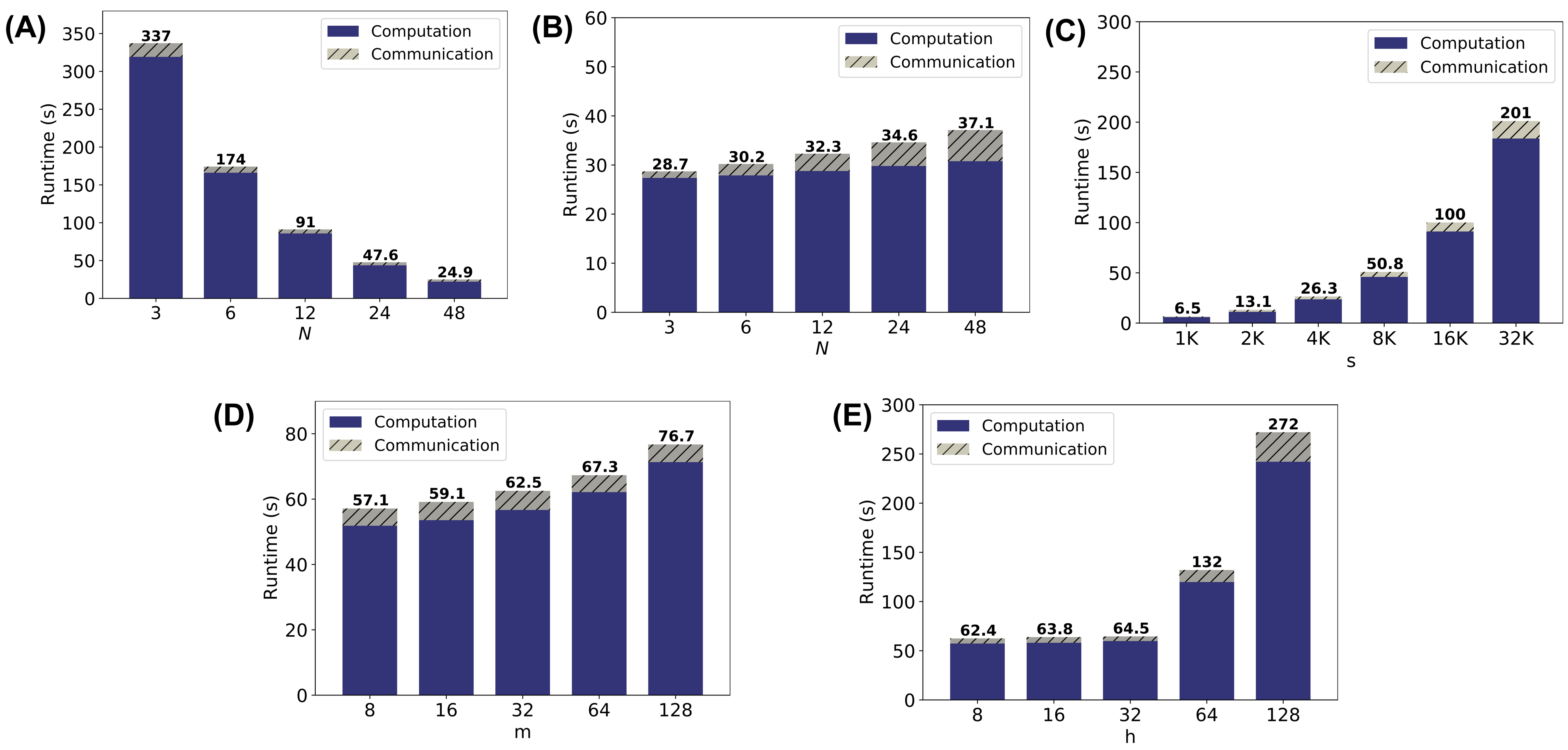
* The IP has been transferred to Tune Insight SA which provides customer care.



Results



Accuracy boxplots when classifying healthy donor (HD) vs. cytomegalovirus infection (CMV) for centralized non-secure, local, and our solution (*PriCell*).



PriCell's training execution time and communication overhead for one training epoch with increasing number of parties, data samples, features, and filters. The computation is single-threaded in a virtual network with an average network delay of 0.17 ms and 1 Gbps bandwidth on 10 Linux servers with an Intel Xeon E5-2680 v.3 CPUs running at 2.5 GHz with 24 threads on 12 cores and 256 GB RAM. (A) Increasing number of parties N when the number of global data samples s is fixed to 18,000. (B) Increasing number of parties N, each having 500 samples. (C) Increasing number of data samples s when N = 10. (D) Increasing number of features m when N = 10. (E) Increasing number of filters h when N = 10.

References

[1] Mouchet et al., Multiparty Homomorphic Encryption from Ring-Learning-with-Errors, PETS, 2021.

[2] Sav et al., Privacy-preserving federated neural network learning for disease-associated cell classification. Patterns, 2022.