


Model Complexity Reduction for ZKML Healthcare Applications: Privacy Protection and Inference Optimization for ZKML Applications—A Reference Implementation With Synthetic ICHOM Dataset

Sathya Krishnasamy, MS;¹  and Ilangovan Govindarajan, MD² 

¹President and Principal, ChainAim, Newington, Connecticut, USA, ²Chief Medical Officer, GuardianMedx, Las Vegas, Nevada

Corresponding Author: Sathya Krishnasamy, Email: sathya.krishnasamy@chainaim.com

DOI: <https://doi.org/10.30953/bhty.v7.340>

Keywords: blockchain, diabetes, distributed ledger, model complexity reduction privacy, machine learning, ICHOM, International Consortium for Health Outcomes Measurement, ZKML, zero-knowledge machine learning

Abstract

Web 3.0 represents the next significant evolution of the internet that embodies the underlying decentralized network architectures, distributed ledgers, and advanced AI capabilities. Though the technologies are maturing rapidly, considerable barriers exist to high-scale adoption. The author discusses the barriers and the mitigations through specific technologies maturing to solve those issues in an earlier paper titled Moving Beyond POCs and Pilots, published in 2023 in Blockchain in Healthcare Today. These include privacy-preserving technologies, off-chain and on-chain design optimizations, and the multi-dimensional approach needed in planning and adopting these technologies. As an extension, this paper discusses one such enabler, zero knowledge machine learning (ZKML), which merges two streams of technology in unique ways to address problems in privacy and the cost of inference. Zero-knowledge proofs (ZKP) allow one party to prove the validity of a statement to another party without revealing any additional information about the statement itself. The ZKML combines the cryptographic principle of ZKP with machine learning (ML) techniques. It is still a maturing technology and needs baselines for applications in global healthcare. In this effort, the authors conceptualize the technical and operational feasibility of using ZKML and implement a reference healthcare implementation using the synthetic International Consortium for Health Outcomes Measurement (ICHOM) in the evaluation phase in a global healthcare setting for high-volume data collection, including patient-reported outcomes. Model complexity reduction is researched and reported for the ICHOM diabetes dataset to advance the usage of ML models in global standards of healthcare data collection in network decentralized architectures for increased data protection and efficiencies.

Submitted: August 1, 2024; Accepted: August 25, 2024; Published: August 31, 2024

The sizable amount of data collection in recent years has produced unprecedented analytical capabilities. However, with the rapid increase in data ingestion and machine learning (ML), particularly in centralized systems, data privacy and security concerns have increased significantly with excessive data movements, sometimes not truly needed. As central data repositories grow larger, they become attractive targets. Healthcare data breaches have become increasingly

common. There have been many high-profile incidents, and the trend has been continuous.

A key strategy for mitigating the risks associated with healthcare data is restricting access to data at the source and reducing unnecessary data movements. These include data access using role-based access, data encryption, minimal data transfer, and regular audits. While these are paramount, it is also critical to devise collaboration systems that could communicate across decentralized networks. Zero-knowledge proofs (ZKP) can be designed for evidence and verification between systems, even before the blockchains. These

Appendix

ZKPs allow one party to prove the validity of a statement to another party without revealing any additional information about the statement itself and have been increasing in popularity as distributed ledger technology matures and has helped scale blockchain implementations.

Zero-knowledge machine learning (ZKML) represents a revolutionary fusion of cryptographic and ML technologies. It combines the cryptographic principle of ZKP with ML techniques. By integrating ZKP with ML, ZKML ensures that sensitive data remain confidential while still allowing for the development and utilization of predictive models. This integration is increasingly relevant where data privacy and security are paramount, such as in healthcare systems.

In practical terms, ZKML allows multiple entities to collaborate without compromising the confidentiality of their proprietary information. This means organizations can collaboratively train and utilize ML models on private datasets without exposing the data. For example, a medical research organization could aggregate data from various hospitals to develop a robust predictive model for disease without any hospital sharing its patient data. The use of cryptographic proofs ensures that the data remain secure and private throughout the process.

Currently, ZKML is still in nascent research and development. While ZKPs have been a topic of cryptographic research since the 1980s, their application to ML is new and complex. This technology faces challenges related to computational efficiency and resource demands. Implementing ZKPs can be resource-intensive, increasing processing times and costs.

Model complexity reduction is a recent technique to reduce model complexity, hence reducing computation times. It has been attempted using simpler models from Kaggle. The International Consortium for Health Outcomes Measurement (ICHOM) is dedicated to developing standard sets of outcome measures that can be used globally to assess the quality of care for various medical conditions. This research aims to evaluate the usage of a global standard healthcare dataset (ICHOM) and apply model complexity reduction for a real-world, non-trivial example to a synthetic dataset in the ICHOM schema. Hence, this research will be a reference for developing further models and optimizations that will make it conducive to optimizing the complexity and use of ZKML for many use cases in healthcare that need privacy and collaborative decision-making.

A Paradigm Shift: Decentralized Systems, AI Models at Source, and Collaborative Decision-Making

As we approach emerging architectures built upon decentralized systems, restricting data at source and understanding ways to work with the data at source via privacy-protected ML becomes essential. Data are

federated anyway in systems such as U.S. healthcare, and effective mechanisms are needed to enable federated learning, address privacy and security issues, and reduce computational overheads. By sharing model updates rather than data, federated learning enhances privacy. Similarly, with distributed ledgers, an increasingly popular design concept is to provide proof to verification systems through ZKP. Though not a new concept, this method has successfully scaled blockchain systems and is becoming increasingly popular for next-generation distributed ledgers. These mechanisms also reduce data transmission risk and align with data protection regulations and collaborative decision-making. The ZKML is a promising emerging technology that integrates AI/ML (artificial intelligence/machine learning) technologies into distributed ledgers.

Impact of ZKML on Healthcare Privacy

Healthcare data are sensitive; therefore, privacy is the first design principle in healthcare systems. With ZKML, healthcare providers can share insights derived from ML models without exposing the underlying patient data. For example, a hospital could use a ZKML model to predict patient outcomes. The model processes the data and generates predictions, while the ZKP ensures that these predictions are accurate without revealing specific patient information. ZKML allows different entities to collaboratively compute results without disclosing their individual data. This is particularly useful in healthcare research, where multiple institutions may wish to combine their data to improve disease predictions without compromising patient confidentiality. As data privacy concerns intensify and the need for secure ML models grows, integrating advanced privacy-preserving technologies becomes crucial.

Literature Review

The ZKPs are cryptographic methods that enable one party (the prover) to convince another party (the verifier) that a statement is true without revealing any additional information beyond the validity of the statement itself. Introduced by Goldwasser, Micali, and Rackoff (1985),¹ they established the theoretical framework for interactive proofs and ZKPs, demonstrating their feasibility and foundational importance.

Further work by Fiat and Shamir² extended it to non-interactive ZKP, which does not require multiple rounds of communication between the prover and verifier. Recent advancements in ZKPs, such as Zero-Knowledge Succinct Non-Interactive Arguments of Knowledge (zk-SNARKs) by Ben-Sasson and colleagues^{3,4} have enhanced their efficiency and scalability and opened possibilities for real-world applications.

Recently, ML showed success in modeling [healthcare prediction tasks, ranging from disease diagnosis and

prognosis to patient treatment. Guerra and colleagues⁵ reviewed privacy-preserving ML literature for training and inference and concluded healthcare datasets are diverse and a fraction of them considered validating with independent standard datasets. They indicated the risks of centralized training for federated learning as a risk and also called out the need for collaboration between different entities across multiple roles of ML scientists, healthcare practitioners, and privacy and security experts, which need privacy-preserving mechanisms to work together over distributed ledgers.

A newly emerging technology, ZKML applies ZKP to ML for privacy, ensuring that sensitive data remain confidential while allowing for developing and utilizing predictive models for privacy and collaboration. The ZKP, as such, are computationally expensive and get even more computationally expensive for ML inference proofs. Recent work from Alejandro Martinez Gator⁶ produced a model complexity reducer (MCR) library and illustrated its reference implementations on sample datasets. However, from a global healthcare perspective, there is a need to validate and benchmark this effort with a high-scale standardized global healthcare schema.

Methods

Research goals

Guardian Medx is a comprehensive care plan offering personalized medical care with continuous monitoring and assistance. The goal is to improve the wellness of seniors and reduce hospitalizations in the southern region of India. The principals aimed to follow an evaluation process first to learn from earlier work done in diabetic care in South India⁷, and identify the cultural elements involved in the disease dynamics, from diagnosis to treatment to adherence to continuous maintenance, and to find the suitable standardized format to capture both clinical and patient-reported outcomes for a holistic basis. The intent is to find the data collection schema conducive to privacy-preserving collaboration in emerging Web 3 technologies, including ML for data analysis and privacy-preserving consent and cooperation.

The starting point for this research is to find a high-scale standardized global healthcare schema that addresses the challenges indicated above, allowing for adequate data collection and identifying the privacy-preserved constructs needed for collaborative learning purposes at a global level.

The specific research goals include the following:

- Learn from earlier research and create a data collection approach that captures the cultural elements of diabetic care.

- Identify a specific data schema that is standardized and can be used for data collection and privacy-preserved learning at a global scale.
- Explore ZKML as the model for that global data schema.
- Identify the parameters, limitations, and mitigations using MCR.
- Baseline the data schema with synthetic data.

After reviewing the literature on diabetic care delivery in India⁷, the big data collection and ML effort⁸ and standardization formats and adaptability reports^{9,10} the data schema was decided to be the ICHOM schema, with the adaption needed.

The Ichom Datasets and Impact on Global Healthcare for Diabetes

The ICHOM is dedicated to developing standard sets of outcome measures that can be used globally to assess the quality of care for various medical conditions. For diabetes, a chronic condition with significant variation in presentation and management across different cultures and healthcare systems, this standardization is crucial.

Diabetes care can vary significantly due to cultural differences, socioeconomic factors, and healthcare infrastructure. For instance, the management strategies and patient outcomes in a high-income country might not directly apply to a low-income setting with different resources and cultural attitudes toward health. ICHOM's dataset addresses this challenge by allowing healthcare systems to adapt standardized measures to local contexts. This cultural tailoring ensures that the outcome measures are relevant and practical in diverse settings, thereby improving the dataset's utility and impact on global health outcomes.

By providing a standard set of metrics, ICHOM helps identify gaps in care and outcomes across different regions and populations. By adopting a uniform approach to measuring outcomes such as blood glucose levels, quality of life, and complication rates, the ICHOM diabetes dataset enables healthcare providers to benchmark their performance against global standards, identify best practices, and improve patient care and population health cohorts. Given the nature of the data and the standardization effort, the principals initiated the evaluation with synthetic datasets. They used the ICHOM older population and diabetic datasets to provide insights into the complex nature of diabetes care and its impact on patient outcomes.

This research offers a comprehensive data-oriented approach to diabetes management by collecting and analyzing data on demographics, diagnosis, lifestyle and social factors, treatment methods, diabetes control, acute events, chronic complications, and patient-reported

outcomes. It seeks to illuminate the intricate interplay of factors affecting diabetes management and optimize diabetes care cost-effectively with early diagnosis and remote and continuous monitoring at scale.

As part of the data evaluation exercise for the research, synthetic datasets with hypothetical values are produced from earlier literature and the researchers' experiences. Data adequacy, baseline, and meaningful data mappings using the ICHOM diabetes datasets V5.0 are established for practical deployments in the Indian cultural setting. These datasets were selected from the ICHOM V5 diabetic dataset. Synthetic data was set up for 100 patients, with several iterations of data and clinical validations. Exploratory data analysis with univariate and bivariate analysis and correlations was developed and analyzed.

The model was built using a Light Gradient Boosting Machine (LightGBM) regressor—an open-source, high-performance gradient boosting framework designed for efficient and scalable ML tasks. It is specially tailored for speed and accuracy, making it a popular choice for both structured and unstructured data in diverse domains. Key characteristics of LightGBM include its ability to handle large datasets with millions of rows and columns, support for parallel and distributed computing, and optimized gradient-boosting algorithms using histogram-based techniques and leaf-wise tree growth.

A crucial aspect of ZKML is model complexity reduction, which is critical with current inference costs and distributed ledger technology scalability to make these models more efficient and practical for real-world applications. This article explores the concept of model complexity reduction in the context of ZKML, with specific examples and a focus on its critical role in healthcare. Model complexity reduction uses concepts of pruning—removing unnecessary parts of the model that contribute minimally to the final decision, quantization—reducing the precision of weights and activations for efficient computing, and knowledge distillation—where the knowledge is transferred to a simpler model that still retains predictive capabilities. This reduced model can then be used within a ZKML framework to perform computations efficiently while providing privacy guarantees through ZKPs.

Furthermore, a stretch goal was to determine the technical feasibility of using the synthetic data generated for ZKML applications effectively in proof and verification systems, given the potential for cross-learning insights without losing privacy.

Model Complexity Reduction for the ICHOM Diabetes Synthetic Dataset

Model complexity reduction in ZKML is essential to mitigate overfitting, improve interpretability, and increase computational efficiencies, reducing the computational resources for training and inference. The ZKML software

and its model reduction library used for this research was GIZA ZK Cook.

The complexity reduction algorithm executes the following steps.

1. Correlation analysis and feature importance for feature selection and reduction: Features with high correlation are candidates that may contribute to redundancy and reduction. Using techniques like recursive feature elimination, features of low importance are eliminated.
2. L1 (Lasso) regularization drives less important feature coefficients to zero, and L2 (Ridge) regularization penalizes large coefficients to reduce complexity without eliminating features. L1 and L2 regularization are combined to balance the benefits of both. For tree-based models, pruning removes branches that contribute minimally to the predictions and consolidates and splits nodes.
3. Principal component analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) are applied for dimensionality reduction.
4. Multi-pass cross-validation and hyperparameter tuning for balancing model complexity and accuracy.

The MCR was used from the Giza library. The analysis model runs in a Python environment.

Results and Discussion

The data calibration from the evaluation phase shows the feasibility of practical and substantial use of ICHOM v5 datasets in periodic and continuous monitoring to prevent the progression of conditions, reducing the quality of life (Figure 1), how it relates to the quality-of-life scores (Figure 2), and how adherence can be mapped granularly (Figure 3).

From the Guardian Medx clinical evaluation of synthetic data produced based on anonymized extraction from earlier aggregated results for the South Indian population, the model is deemed to help significantly in advancing outcome-based and cost-effective remote

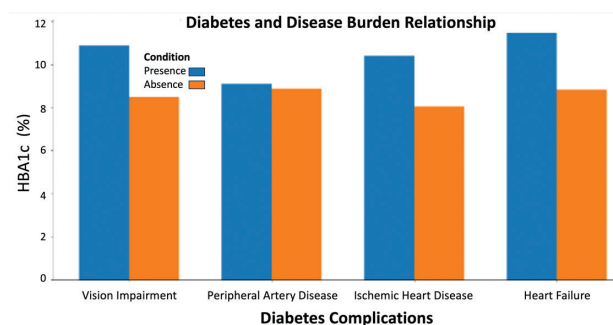


Figure 1. Diabetes quality-of-life complications. HbA1c: glycosylated hemoglobin.

Source: Copyright by the authors

monitoring care, based on data analysis and regressor

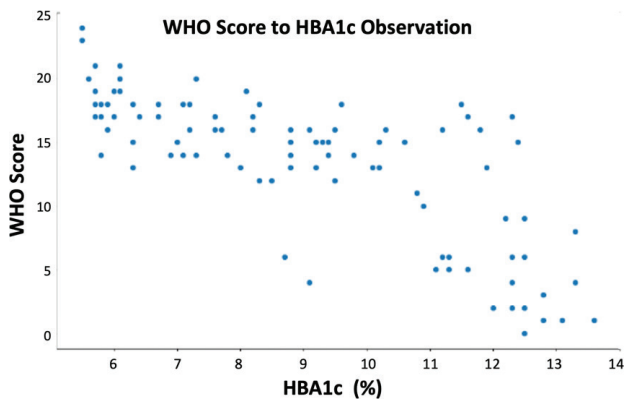


Figure 2. International Consortium for Health Outcomes Measurement: WHO (World Health Organization) scores versus HBA1c (glycated hemoglobin). Source: Copyright by the authors

models for the data analyzed in phase 1. This is particularly the case in HBA1c (glycated hemoglobin) control, frequency of tracking, and relationship to complications and quality of life scores. The co-existence of other chronic diseases was evident from the data. The ICHOM data captures patient-reported outcomes with WHO scores, which showed a negative trend in quality-of-life scores with increasing HBA1C values reported. All adherence metrics also showed the expected relationship with the HBA1c results.

The study opens the possibility of alerts for action based on escalation predictions as the remote monitoring feeds come in for at-scale patient management for early interventions.

Model Complexity Before Model Reduction

The key aspect is to evaluate the model complexity reduction baselines for this synthetic dataset and compare the

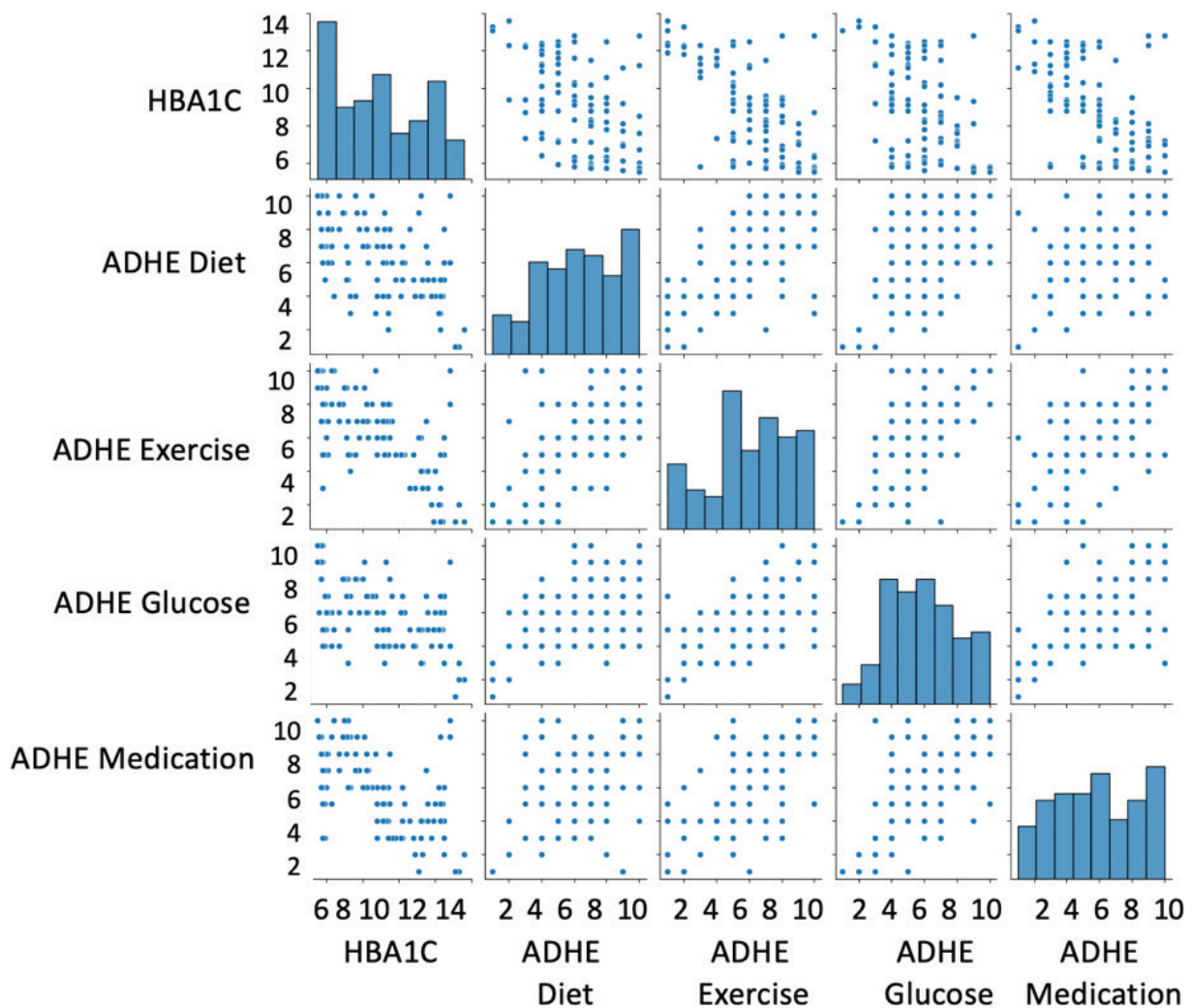


Figure 3. Effect of adherence (ADHE) on HBA1c (glycated hemoglobin) control. Source: Copyright by the authors

```

lgbm_reg.get_params()
{'boasting_type': 'gbdt',
 'class_weight': None,
 'colsample_bytree': 1.0,
 'importance_type': 'split',
 'learning_rate': 0.1,
 'max_depth': 8,
 'min_child_samples': 20,
 'min_child_weight': 0.001,
 'min_split_gain': 0.0,
 'n_estimators': 1200,
 'n_jobs': None,
 'num_leaves': 31,
 'objective': None,
 'random_state': None,
 'reg_alpha': 0.0,
 'reg_lambda': 0.0,
 'subsample': 1.0,
 'subsample_for_bin': 200000,
 'subsample_freq': 0}

[12] model, transformer = mcr(model = lgbm_reg,
                             X_train = X_train,
                             y_train = y_train,
                             X_eval = X_test,
                             y_eval = y_test,
                             ...)
  
```

Figure 4 Model complexity before model complexity reduction. ICHOM: International Consortium for Health Outcomes Measurement; MCR: model complexity reducer. Source: Copyright by the authors

before-and-after model parameter complexity. This is crucial to optimizing the run-time cost and seeing how this model fares for use in distributed ML and collaborative systems over distributed ledgers.

The LGBM regressor was configured with the parameters n-estimators: 1,200 and max-depth: 8 (Figure 4). In contrast, the model complexity is reduced to n-estimators: 150 and max-depth: 4, after the model complexity reduction after running through the ZKCook library (Figure 5).

Representing this in terms of nodes:

- Number of Nodes = Number of trees * ($2^{\text{depth}} - 1$)
- The before complexity reduction number of nodes evaluate to $1,200 * (2^8 - 1) = 306,000$
- The before complexity reduction number of nodes evaluate to $150 * (2^4 - 1) = 2,250$

Model Complexity After Model Complexity Reduction

The difference accounts for a 99.26% reduction, and it is also in line with some of the other reference examples. The choice of the regressor for the problem and further reduction using MCR based on the evaluation data shows that the ICHOM V5 diabetes dataset can be used to capture data so that the interoperability can be enhanced to study and report results in collaborative settings to be used in ZKML applications.

These results imply that the needed adaptability for the diabetes study could be captured in the global format of ICHOM and show promise based on the evaluation data. In addition, these results show promise in generating privacy-protected proofs for verification systems and collaborative sharing of proof of diagnosis based on patient consent in a privacy-protected way with other collaborating parties. Given these results in the evaluation phase, the proofs can be generated optimized for computational efficiency once the study progresses.

This specific adaptation included a subset of the complete dictionary of the ICHOM dataset, as it was adapted to this clinical setting, which is a limitation. So, we see this as a starting point for other work to use standardized datasets such as the ICHOM datasets in different cohorts and other diseases from the ICHOM datasets. Some of those situations will have an increased number of data columns and data analysis requirements, which will give us additional reference points for complexity before and after baselines and their impact on proof and verification systems. Also, an important point to note is that this is a preliminary baseline, as the technology matures very fast from all angles—standards development, model reduction techniques, reduction of proof, and verification systems acceleration both at the software and the hardware level. Hence, it will become

```

[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf

[13] model.get_params()

{'boosting_type': 'gbdt',
 'class_weight': None,
 'colsample_bytree': 1.0,
 'importance_type': 'split',
 'learning_rate': 0.1,
 'max_depth': 4,
 'min_child_samples': 20,
 'min_child_weight': 0.001,
 'min_split_gain': 0.0,
 'n_estimators': 150,
 'n_jobs': None,
 'num_leaves': 25,
 'objective': None,
 'random_state': None,
 'reg_alpha': 0.0,
 'reg_lambda': 0.0,
 'subsample': 1.0,
 'subsample_for_bin': 200000,
 'subsample_freq': 0,
 'min_data_in_leaf': 35,
 'feature_fraction': 0.39476727085158336,
 'bagging_fraction': 0.18662014176974878,
 'verbose': -1,
 'early_stopping_rounds': 10}

```

Figure 5. Model complexity before model complexity reduction. ICHOM: International Consortium for Health Outcomes Measurement; MCR: model complexity reducer. Source: Copyright by the authors

important to have a registry of ZKML developments across these parameters.

Conclusions and Future Work

Based on data analysis and regressor models for the data analyzed in the evaluation phase, GuardianMedx clinical evaluation deems the model to help significantly advance outcome-based and cost-effective remote monitoring care. Applying the Giza ZKcook model complexity reduction algorithm to ICHOM diabetes data resulted in more interpretable and computationally efficient models. The computing times were significantly reduced on a standardized ICHOM dataset to use ML and privacy-protected settings to retain data at the source. This increases security and provides verifiable proofs for any prediction models driving agents and to use ML models in conjunction with decentralized distributed ledgers to open collaboration possibilities without giving out the internal details of the data. Given these results in the evaluation phase, once the study progresses, the proofs can be generated optimized for computational efficiency. Further work can be extended to other adaptations of the ICHOM framework for diabetes in another cohort in another setting to compare results, as well as using them in other disease datasets from ICHOM. Shortly, the team intends to extend the ZKML functionalities to feed agents for further processing and advancing privacy-protected multi-party insights.

Funding

None.

Conflicts of Interest

None.

Contributors

Sathya Krishnasamy is the President and Principal of ChainAim Technologies. His 25 years of background span extensive experience in managed care payor settings in leading U.S. healthcare firms, including Aetna and Anthem. He focuses on emerging technologies, including AI/ML systems and distributed ledger technologies. He also serves as an advisor in many industry efforts in payor-provider collaboration, standards organizations, and efforts such as Account Aggregators in India advancing Fintech, Healthcare, and Skills sectors. He currently serves as President and Principal at ChainAim, offering technical strategy consulting and application and development services.

Sathya Krishnasamy helped conceptualize the use of ICHOM for data, evaluate the synthetic dataset, establish a model complexity baseline, and evaluate the ZKML healthcare use case.

Dr Govindarajan is the Chief Medical Officer of GuardianMedX. He is a healthcare executive with a Strong Medical Background and technological knowledge. He has 35 years of extensive experience in internal

medicine and geriatrics in India and serves as an advisor for geriatric and palliative care for many government entities in India. He has managed and administered patient-centric quality care following a unique continuum of care in clinics, hospitals, nursing homes, hospices, and homes.

Dr Govindarajan started the initiative and performed the research for data collection needs, design and evaluation of the ICHOM for diabetes data, and clinical evaluation of the synthetic datasets.

Data Availability Statement (DAS), Data Sharing, Reproducibility, and Data Repositories

The data dictionary for the ICHOM V5 diabetes data dictionary is available at <https://www.ichom.org/patient-centered-outcome-measure/diabetes/>

Application of AI-Generated Text or Related Technology

None.

Acknowledgments

Yugesh Panta, a Master of Science student at the Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, helped the principals with research data collection, validation, analysis, and model-tuning aspects.

References

1. Goldwasser S, Micali S, Rackoff C. The knowledge complexity of interactive proof systems. Proceedings of the seventeenth annual ACM symposium on Theory of computing—STOC '85. 1985.
2. Fiat A, Shamir A. How to prove yourself: practical solutions to identification and signature problems. *Advances in Cryptology—CRYPTO' 86* [Internet]. 2019;186–94. Available from: https://link.springer.com/chapter/10.1007%2F3-540-47721-7_12 [cited 2024 July 31].
3. Ben-Sasson E, Chiesa A, Tromer E, Virza M. Succinct non-interactive zero knowledge for a von neumann architecture [Internet]. 2019. Available from: <https://eprint.iacr.org/2013/879.pdf> [cited 2024 July 31]
4. Ben-Sasson E, Bentov I, Horesh Y, Riabzev M. Scalable, transparent, and post-quantum secure computational integrity [Internet]. ePrint IACR. 2018. Available from: <https://eprint.iacr.org/2018/046> [cited 2024 July 31]
5. Guerra-Manzanares A, Lechuga J, Maniatakos M, Shamout FE. Privacy-preserving machine learning for healthcare: open challenges and future perspectives. ICLR 2023 Workshop on Trustworthy Machine Learning for Healthcare. arXiv. 2023; 1-13. <https://arxiv.org/abs/2303.15563>
6. Gotor AM. Maximizing model efficiency with model-complexity-reducer (MCR). [zkcook/docs/mcr.pdf](https://github.com/gizatechxyz/zkcook/blob/main/docs/mcr.pdf) at main gizatechxyz/zkcook [Internet]. GitHub. [cited 2024 Aug 1]. Available from: <https://github.com/gizatechxyz/zkcook/blob/main/docs/mcr.pdf>
7. Das AK, Saboo B, Maheshwari A, Nair VM, Banerjee S, Jayakumar C, et al. Health care delivery model in India with relevance to diabetes care. *Heliyon*. 2022 Oct;8(10):e10904. <https://doi.org/10.1016/j.heliyon.2022.e10904>
8. Musacchio N, Giancaterini A, Guaita G, Ozzello A, Pellegrini MA, Ponzani P, et al. Artificial intelligence and big data in diabetes care: a position statement of the Italian Association of Medical Diabetologists. *J Med Internet Res*. 2020 Jun 22;22(6):e16922. <https://doi.org/10.2196/16922>
9. Diabetes [Internet]. ICHOM. [cited 2024 Aug 1]. Available from: <https://www.ichom.org/patient-centered-outcome-measure/diabetes/>
10. Benning L, Das-Gupta Z, Fialho LS, Wissig S, Tapela N, Gaunt S. Balancing adaptability and standardisation: insights from 27 routinely implemented ICHOM standard sets. *BMC Health Serv Res*. 2022 Nov 28;22(1):1424. <https://doi.org/10.1186/s12913-022-08694-9>

APPENDIX

Acronyms defined

AI/ML: Artificial Intelligence / Machine Learning

HbA1c: Glycosylated Hemoglobin

ICHOM: International Consortium for Health Outcomes Measurement

LightGBM: Light Gradient Boosting Machine

MCR: Model Complexity Reducer

ML: Machine Learning

zk-SNARKs: Zero-Knowledge Succinct Non-Interactive Arguments of Knowledge

ZKML: Zero knowledge machine learning

ZKP: Zero-knowledge proofs

Copyright Ownership: This is an open-access article distributed in accordance with the Creative Commons Attribution Non-Commercial (CC BY-NC 4.0) license, which permits others to distribute, adapt, enhance this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0>.