International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Privacy Preserving Machine Learning in Healthcare

[1]Venkata Raju, [2]Sirisha Balla, [3]Dr. Prasad Rayi

Submitted: 05/09/2024 **Revised:** 15/10/2024 **Accepted:** 25/10/2024

Abstract: The population of older individuals requiring care in Danish society is increasing. This trend necessitates innovative and more efficient methods of functioning within the healthcare industry. Fall prevention is an issue requiring attention. To avoid falls, it is essential to evaluate the risk of falling among senior individuals. Should the danger be elevated, the implementation of additional assistance equipment or training programs may commence. Currently, the fall risk assessment is a manual and ineffective process. The optimization of this approach via the use of machine learning techniques delineates the scope of this thesis.

Machine learning employs extensive datasets throughout the training process. Under typical circumstances, this is not an issue; nevertheless, within the context of this thesis, the training data is confidential and so has sensitive characteristics. This complicates the optimization of machine learning and is the fundamental issue addressed in this thesis. What methods may be used to train machine learning algorithms on sensitive datasets?

This thesis assesses the use of federated learning for training machine learning algorithms on decentralized datasets. Additionally, it examines the use of encryption methods and differential privacy to safeguard machine learning models against the disclosure of sensitive information. Finally, it examines how intricate machine learning models, particularly deep learning models, might be rendered explainable to facilitate clearer communication of the findings to senior citizens.
Numerous tests have shown that federated learning provides a technique for decentralized learning, although at the expense of model performance. Both encryption approaches and differential privacy may enhance the security of machine learning models against data leakage, but at the expense of model performance and complexity. Finally, it is shown how a technique known as SHAP may assist in calculating SHAP feature values, hence making machine learning models more comprehensible.

Keywords*: necessitates, nevertheless, optimization, safeguard, comprehensible.*

INTRODUCTION

This master's thesis is a component of a project titled AIR, or AI-Rehabilitation. The AIR project seeks to use artificial intelligence in the rehabilitation process for older individuals within Danish towns. The Danish society has seen a rising population of older individuals in recent years. This indicates a growing population requiring care, hence creating an increasing need for funds in the caregiving sector [33]. The heightened demand is a topic of political discourse within society [31], although the AIR project examines how to use existing money most efficiently.

[1]*Associate Professor, Department of Computer Science and Engineering, International School of Technology and Sciences for Women, A.P, India.*

[2]*Assistant Professor, Department of Computer Science and Engineering ,International School of Technology and Sciences for Women, A.P, India.*

[3]*Associate Professor , International School of Technology and Sciences for Women , A.P, India.*

By examining the use of artificial intelligence, it may be assessed which citizens are most likely to benefit from the resources within the caregiving domain. Systems receive information on training sessions with older individuals, including the assistance gadgets used. The system is then determining the optimal use of resources based on this input.

Thus far, it has been the primary emphasis of the AIR project [10]. Nevertheless, given substantial quantities of data are already being amassed, the focus is now on investigating how this data might be used for other applications within the sector. The evaluation of fall risk is a novel field of research and serves as the focus of this thesis. How might artificial intelligence and the extensive data sets now being gathered enhance fall-risk assessments conducted by caregiving staff, and what is the most effective method for communicating the results? This serves as the foundation for the thesis.

A significant concern in the AIR project is the privacy of the bulk of data acquired from Danish

residents inside the caregiving programs. The data is legally protected [38], and it is becoming difficult to gather and keep it.
Data from many towns in Danish society are accessible; however, the sharing of such data nationwide is challenging owing to GDPR laws. The primary focus of this thesis is to investigate methods for using artificial intelligence while safeguarding data privacy, hence facilitating the exchange of sensitive information.

Privacy preserving machine learning

Numerous machine learning applications include data that is, to some extent, private to individuals. Browsing history, purchase history, marketing preferences, and medical heath information are examples of data used by organizations for machine learning.
Nevertheless, certain categories of data are increasingly safeguarded. On May 25, 2016, a new data protection regulation came into effect for European residents.
The General Data Protection Regulations (GDPR) restrict corporations' access to their data and provide more stringent protocols for managing personal information.
The implication for firms using personal data is that a plan for managing both present and future data must be established. As the volume of data gathered increases, the availability of data for machine learning diminishes. This presents a significant challenge for organizations that depend on their machine learning systems.
Moreover, as articulated by Roberta Kwok in [23], contemporary organizations see accumulated knowledge as very valuable assets. Consequently, it is logical that the majority of corporations choose not to disclose their information to other entities. This reiterates that, without the incorporation of privacy in machine learning, data acquisition may become more challenging.
In light of these conditions, a novel kind of machine learning, referred to as privacy-preserving machine learning, has been explored in recent years. Privacy-preserving machine learning employs techniques that use machine learning while minimizing information leakage from training data. This indicates the feasibility of uncovering patterns within datasets while maintaining data confidentiality. Utilizing privacy-preserving machine learning enables the use of the growing volume of data produced inside societies, notwithstanding its private nature. Numerous distinct methodologies for privacy-preserving machine learning will be examined further.

Machine learning in healthcare

In [15], the authors examine the contemporary use of machine learning in healthcare. They emphasize that the healthcare industry has seen a significant increase in the volume of gathered data in recent years. Innovative digital methodologies for patient engagement result in increased data collecting. As shown in chapter 1, data is fundamental to the development of machine learning models. Abundant data facilitates the identification of statistical patterns, which is the objective of machine learning.

The article [15] provides significant information on the healthcare system. Fifty percent of healthcare system costs are attributable to only five percent of the individuals inside the system. This indicates that some people incur substantial treatment costs. These individuals often suffer from chronic illnesses; nevertheless, the use of machine learning may facilitate early detection of these occurrences. Initiating therapy at an earlier stage may enhance quality of life while simultaneously reducing costs.

The research also indicates that almost 90 percent of emergency hospital visits are avoidable. Machine learning may assist in diagnosing and directing patients to appropriate treatments while minimizing expenses by reducing reliance on costly, time-consuming emergency care facilities. This demonstrates the significant potential of machine learning in healthcare.
The article "Machine Learning in Healthcare" [17] provides a comprehensive account of the use of machine learning models in contemporary healthcare. All the above examples pertain to instances when machine learning models are trained using Electronic Health Records (EHRs). These are data obtained from electronic programs used in healthcare systems. This replicates the data used in the trials for this thesis (refer to chapter 4), indicating that the methodologies are flexible.

The EHRs provide examples of the use of machine learning models in two distinct scenarios. The primary use is in forecasting the development of illnesses. Employing machine learning for this objective may facilitate early disease detection, since the models may identify patterns in typical

people exhibiting a certain illness. The alternative example pertains to the optimization of hospital operations. Machine learning models are used to identify trends in various operational tactics, therefore delineating the most effective technique. According to the findings from [17], it is evident that data obtained from healthcare systems is very beneficial for machine learning models. The impressive findings outlined in the paper suggest that additional areas, such as optimizing fall risk assessment, need further investigation.

Privacy preserving machine learning

As the volume of data in the healthcare industry rises and the emphasis on data security legislation intensifies, the need for machine learning approaches that ensure data privacy is correspondingly growing. Privacy-preserving machine learning refers to strategies that, to varying extents, safeguard the data used as input for training algorithms. It asserts a commitment that the information used remains confidential and undisclosed to external parties. A commitment, particularly significant in the healthcare industry because much data is inherently sensitive. Privacy-preserving machine learning may be accomplished via many methods. Research identifies three primary methodologies for safeguarding privacy in machine learning. Differential privacy, federated learning, and encrypted deep learning.

Differential privacy is the principle of incorporating a certain degree of noise into a query to provide plausible deniability for each input sample used. The noise contributes to the statistical uncertainty over whether the input sample contains real values or is distorted by noise. Differential privacy is used in several applications, such as in [28], where it facilitates the private publication of social network information. Chapter 3 will examine differential privacy and its application to healthcare data to provide privacy protection.

Federated learning [27] tackles a recognized issue in privacy preservation. What methods may be used to communicate data across disparate data clusters while ensuring privacy is not compromised? The GDPR imposes stringent limitations on the permissible sharing of data between companies [38].

Federated learning alters the conventional methodology of machine learning. In traditional machine learning frameworks, data is centralized, a model is trained, and subsequently, the model is deployed.

Due to the challenges posed by sensitive data and GDPR rules, federated learning trains local models and then centralizes them. When models are trained successfully, they cease to disclose information about the input samples. Consequently, it is safe to disseminate the local machine learning models. Upon sharing, a datacenter combines the models and deploys the consolidated model for future use across various organizations. This method allows for the use of local data to enhance performance in other entities while safeguarding privacy. Chapter 3 will detail the implementation of federated learning, enabling each municipality to train local models and securely disseminate the knowledge contained inside those models.

Privacy preserving needs in healthcare

Section 1.2 delineates the objectives of this thesis, which are to provide a methodology for data sharing across various municipalities while ensuring that the models generated from the shared data do not compromise the privacy of the residents whose data was used in the training process. This presents two distinct issues that need resolution via different methodologies.

Privacy in the data-sharing process necessitates that all data samples sent between municipalities have a confidential nature. Section 3.2 will examine how standard data anonymization may be insufficient to guarantee privacy. Consequently, an additional method to guarantee data privacy during data exchange is required. Federated learning, as previously discussed in section 2.3, offers a method for using data samples from other entities without the need of data sharing. Federated learning offers a mechanism for centralizing models rather than data throughout the training process. This is the selected technique to address this segment of the specified objectives. Privacy in machine learning models signifies that these models cannot be used to monitor the data employed for their training. As discussed in section 3.2, machine learning models may be undermined by adversarial attacks, potentially exposing sensitive information from the training process. To prevent this occurrence with sensitive healthcare data, the principle of differential privacy (introduced in section 2.3) will be used. Differential privacy is a technique for incorporating noise into model training to guarantee that no particular sample is too

exposed.
Information from the whole population of samples is used in the training process.
Utilizing both federated learning and differential privacy enables safe data exchange and model development. The ideas and their underlying theories will be examined and implemented in the subsequent sections.

DISCUSSION

This chapter analyzes and assesses the suggested solutions for a privacy-preserving machine learning methodology outlined in chapter 3, informed by the results from the experimental phase conducted in chapter 5. The assessment is conducted based on the efficacy of the various solutions within the domain. The first suggested solution, data central deep learning, delineates the specific issue area that led to this thesis. Data cannot be exchanged between municipalities owing to universal data protection requirements. This fact indicates that a centralized data method is infeasible for managing sensitive information, hence necessitating the emergence of privacy-preserving efforts.
The second option, traditional federated learning, relies on the principle of developing models locally and disseminating the model parameters rather than sharing the data itself. The PPIN-A project illustrated this methodology by sharing a machine learning model across municipalities. Nonetheless, PPIN-A also shown that the federated learning technique has several limits and drawbacks. The efficacy of a federated learning model results in a decline in both precision and time expenditure. In the healthcare sector, the deterioration in processing time may not be significant, since model training occurs just when a fresh dataset is accessible. This would be restricted to, at most, a few instances annually, indicating that an extended processing time would be permissible. Nonetheless, the decline in accuracy performance would be more significant. A diminished accuracy in fall risk assessment would result in increased susceptibility to errors. This may eventually result in insufficient fall prevention. The third option, federated learning with Differential Privacy Stochastic Gradient Descent (DPSGD), addresses a fundamental shortcoming of federated learning. Federated learning models may nevertheless disclose sensitive information due to the potential for model inversion attacks. The use of differential privacy mitigates this danger. Experiment PPIN-B-2 demonstrated the use of differential privacy into the machine learning model training phase to enhance model security. PPIN-B-2 demonstrated that the volume of noise used in training creates a trade-off between privacy and accuracy. This indicates that the privacy issue must be assessed prior to implementing models using differential privacy as a privacy strategy. One of the principal findings from PPIN-B-2 is that when noise levels grow, the reduction in accuracy is far less than the enhancement in privacy. If a little decrease in accuracy is not essential, a significant improvement in privacy may be achieved.

Conclusion

This study presents and summarizes current work on PPML for model training and inference in the healthcare sector. We emphasize trends, problems, and potential future research avenues. In conclusion, we acknowledge the absence of unanimity about the definition of needs for privacy-preserving frameworks in healthcare. This necessitates cooperation among machine learning scientists, healthcare professionals, and specialists in privacy and security. We urge researchers to conduct thorough evaluations of proposed algorithms on varied medical datasets to enhance generalization, to explore the limitations of PPML in multi-modal learning contexts, to examine the potential of MLaaS in healthcare as a driver for improved service delivery, and to incorporate cutting-edge advancements in deep learning architectures to optimize model performance. Our recommendations seek to bridge research gaps and direct future inquiries in PPML to promote the forthcoming use of reliable and private machine learning in healthcare.

References

[1] Accuracy of noisy counting. https: //georgianpartners.shinyapps.io/interactive_counting/. Accessed: 2020-10-16. [2]Federated learning: Collaborative machine learning without centralized training data. https://ai.googleblog.com/2017/04/ federated-learning-collaborative.html. Accessed: 2020-10-16.

[3] Hj_lpemiddelbasen. https://hmi-basen.dk/. Accessed: 2020-10-16.

[4] Hvad er tv_rspor?? https://www.tvaerspor.dk/hvad-er-tvarspor/. Accessed: 2020-10-16.

[5] Interpreting complex models with shap values. https://medium.com/@gabrieltseng/ interpreting-complex-models-with-shap-values-1c187db6ec83. Accessed: 2020-10-16.

[6] Local vs. global di_erential privacy. https://desfontain.es/privacy/ local-global-differential-privacy.html. Accessed: 2020-10-16.

[7] Udacity course on secure and private ai. https: //www.udacity.com/course/secure-and-private-ai--ud185? irclickid=VzXTSiQ0NxyORSgwUx0Mo3ERUkiQpKVtnzkY080&irgwc=1& utm_source=affiliate&utm_medium=ads_n&aff=259799. Accessed: 2020-10-16.

[8] What is secure multi-party computation? https://medium.com/pytorch/ what-is-secure-multi-party-computation-8c875fb36ca5. Accessed: 2020-10-16.

[9] Why one-hot encode data in machine learning? https://machinelearningmastery.com/ why-one-hot-encode-data-in-machine-learning/. Accessed: 2020-10-16.

[10] Aarhus University. AIR (AI Rehabilitation). https://projekter.au.dk/air/, 2020. Online; accessed 24 September 2020.

[11] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with di_erential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pages 308{318, 2016.

[12] Mohammad Al-Rubaie and J Morris Chang. Privacy-preserving machine learning: Threats and solutions. IEEE Security & Privacy, 17(2):49{58, 2019.

[13] Eugene Bagdasaryan, Omid Poursaeed, and Vitaly Shmatikov. Di_erential privacy has disparate impact on model accuracy. In Advances in Neural Information Processing Systems, pages 15479{15488, 2019.

[14] Aur_elien Bellet, Amaury Habrard, and Marc Sebban. A survey on metric learning for feature vectors and structured data. arXiv preprint arXiv:1306.6709, 2013.

[15] R. Bhardwaj, A. R. Nambiar, and D. Dutta. A study of machine learning in healthcare. In 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), volume 2, pages 236{241, 2017.

[16] Peter Bjerregaard and K Juel. Middellevetid og d_delighed i danmark. Ugeskrift for Laeger, 155(50):4097{100, 1993.

[17] A. Callahan and N. Shah. Machine learning in healthcare. 2017.

[18] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of di_erential privacy. Foundations and Trends in Theoretical Computer Science, 9(3-4):211{407, 2014.

[19] Jianjiang Feng and Anil K Jain. Fingerprint reconstruction: from minutiae to phase. IEEE transactions on pattern analysis and machine intelligence, 33(2):209{223, 2010.

[20] J Franken_eld. Arti_cial intelligence (ai). Investopedia udgivet, 13(6):19, 2019.