Innovations

A Survey of Recent Enhancement in Deep Learning for Electronic Health Records (EHR)

Mr. M. Priyadharshan, Mr. Dhanaselvan J U*, Mr. Agash A, Mr. Mohammed Afsal A, Mr. Deepak Raj M.
Assistant Professor, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India
Department of Computer Science and Engineering, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India

Abstract: *Nowadays, the usage of the Electronic Health Record (EHR) system in healthcare organizations has increased. The electronic health record consists of a vast amount of sensitive patient information, like clinical notes, laboratory test results, and procedures critical for healthcare and medical research. By the Electronic Health Record, we can perform operations like information extraction, representation learning, outcome prediction, and deidentification. Electronic medical records are more convenient to store and utilize than paper-based records. But in this, ensuring the privacy and security of this data is paramount to maintaining patient trust and complying with the regulatory requirements. In this, we discussed the recent developments in the deep learning techniques to ensure data privacy in the electronic health record and concluded with the challenges of the techniques.*

Keywords:*Data Privacy, Electronic Health Record, Deep Learning, Machine Learning, Health care, Data preservation.*

1. Introduction:

An Electronic Health Record (EHR) collects patient information in both structures and unstructured digital formats. The main goal of EHRs is to increase the effectiveness of healthcare systems so that valuable insights can be drawn from the data they collect.

Most of techniques for analyzing rich EHR data were based on traditional machine learning such as logistic regression, support vector machines (SVM), and random forest [3]. But while using these techniques obtaining data from other institutions is complicated due to the privacy of data.

To Overcome these problems, we can go with the decentralized deep learning techniques such as Federated Learning, split learning and Hybrid Learning to improve the data privacy while developing a model to make useful insights [2].

In this paper we discussed the various implementation and security methods in deep learning for Electronic Health Record to ensurerobust security.

Abbreviations	Explanation
EHR	Electronic Health Record
SVM	Support Vector Machines
DL	Deep Learning
ML	Machine Learning
LR	Linear Regression
FL	Federated Learning
SL	Split Learning
SFL	Hybrid Split Federated Learning

Table -1. *Abbreviations table*

2. Traditional Machine Learning Approaches:

We covered the many machine-learning approaches and their drawbacks in this part. Traditional centralized ML algorithms such as Support Vector Machines (SVM), Random Forest, Linear Regression [3], Decision Tree etc...

2.1 Linear Regression

Linear Regression is the widely used statistical technique for predicting a continuous outcome based on one or more input features. The basic premise of the linear regression is to establish a linear relationship between these variables

In context of the EHR, Linear regression is used to predict the outcome like probability of disease based on the lab results and the previous medical history.

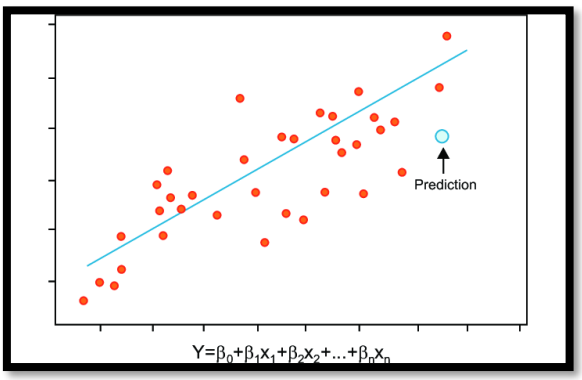

Fig. 1 Linear Regression

Disadvantages of Linear Regression

It's limited to the inability to model nonlinear regression and its susceptibility to multicollinearity that makes it less suitable for the complex EHR data.

2.2 Support Vector Machine

Support vector machines (SVM) are a popular machine learning approach for applications including regression and classification [4]. In healthcare, it is used for various tasks such as diagnosis, prognosis and disease predictions [4].

The goal is to choose the best hyperplane with the largest margin that divides the data points into distinct groups.

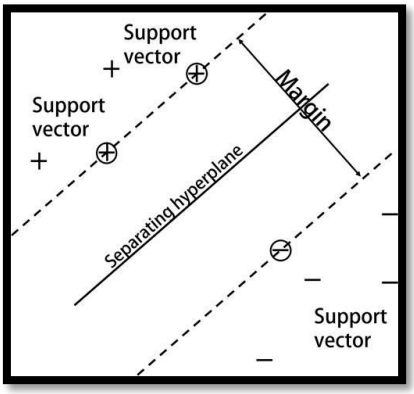

Fig. 2 Support Vector Machine

Disadvantages of SVM

One of the main disadvantages of this technique is it becomes very complex when we are working on very large datasets. The training time increases as the number of data points increases.

2.3 Random Forest

Random Forest (RF) has the capability to manage complicated and high-dimensional data while producing accurate predictions, they are popular techniques in EHR.

Because of their versatility and robustness, RF models are utilized in HER analysis for a variety of tasks such as patient readmissions, disease prediction, and mortality prediction.

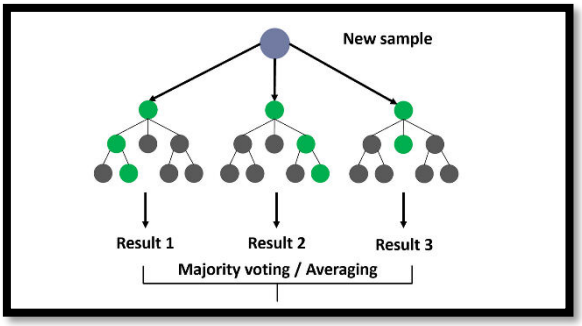

Fig. 3 Random Forest

Disadvantages of RF

The EHR data is shared among various systems or organizations. Encryption and other safe data-sharing techniques are required because RF models require raw data for training and prediction stages. In this training or inferring models, data breaches could occur if appropriate security standards aren't followed.

2.4 Decision Tree

Decision trees are simple and powerful ML technique, this is highly interpretable and useful for clinical decision support systems [5].

Decision trees can model complex, non-linear relationship between variables, which are common in medical data.

Disadvantages of Decision trees

It is vulnerable to model inversion attacks, where adversaries can infer sensitive patient data by probing the model with carefully crafted inputs.

3. Decentralized Deep Learning Approach:

In this section we have discussed the various decentralized deep learning approaches and their advantages such as Split Learning (SL), Federated Learning (FL), Split Hybrid Learning (SFL) [2].

3.1 Federated Learning (FL)

In this approach, it enables multiple organizations to train the model to predict the disease or medicine recommendation without sharing their raw data [6]. In this different organization, they train their own model, and finally all the models are collected and processed by the one centralized system, as you see in Fig. 4.

This approach is particularly useful in healthcare, where data privacy and security are critical concerns.

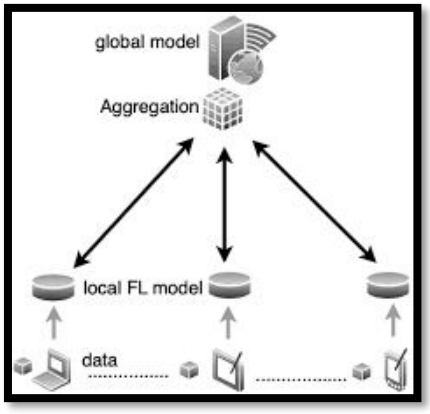

Fig. 4 Federated Learning

3.2 Split Learning (SL)

Split Learning is [2] an important technique designed for data privacy and data security particularly suitable for electronic health records. In this approach, it is splitting the neural network model into multiple segments and distributed across different entities.

Each entity trains its segment on local data, and only the intermediate output is shared between the multiple organizations to prevent the raw data.

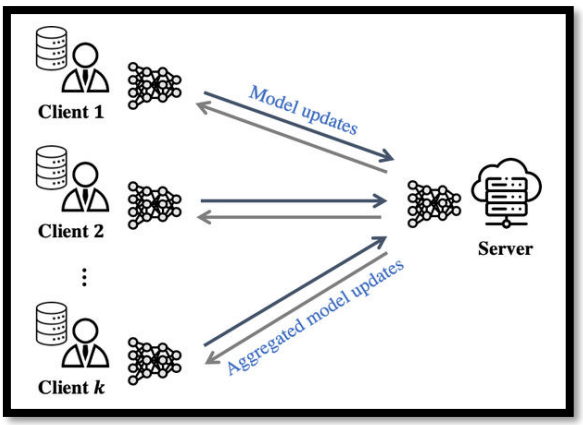

Fig 5. Split Learning

Advantages of Split Learning

- **Reduced risk of data leakage:** Here the intermediate models are shared instead of the entire dataset. So, it reduces the risk of data leakage or breaches.

- **Protection against the attacks:** Split Learning makes it more difficult for adversaries to undertake inference attacks to reverse-engineer the original data derived from shared model updates or activations.

3.3 Hybrid Split Learning (SFL)

It is the result of combining Split Learning (SL) with Federated Learning (FL). A neural model is divided between the client and the server in split learning (SL) [2]. Federated Learning (FL), on the other hand, enables several clients to work together to train the model without sharing the raw data [7].

In these models, a split version of the model is given to each client who trains the model locally [9]. None of the clients sent the complete model or the raw for these.

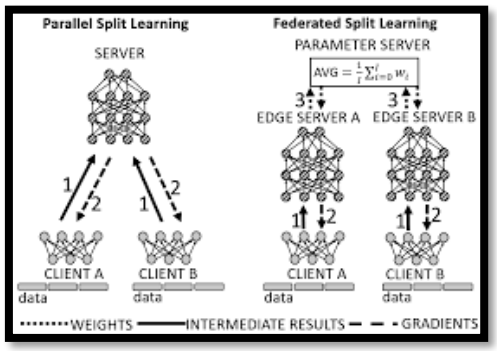

Fig 6. Hybrid split Federated Learning

Advantages of Hybrid Split Federated Learning (SFL)

- **Increase Model Generalization:** SFL enables multiple institutions to contribute their data while maintaining privacy as well. This leads to a more model generalization and robust.

- **Reduces client infrastructure cost:** Since only a fraction of the model is trained locally, SFL reduces the need for high-performance hardware on the client-side and reduces the infrastructure and operational cost.

Aspect	**Traditional Machine Learning Techniques**	**Decentralized Deep Learning Techniques**
Data Storage	Central server aggregates all data.	Row data remains local, it only updates are shared.
Privacy	High risk of breaches due to data centralization.	The data privacy was improved while the data was kept locally.
Security	Single point of failure for attacks may cause high risk.	Risk is low with the distributed data storage.
Personalization	Limited Personalization	It allows multiple organizations to develop or train the model.
Scalability	Infrastructure and data transfer are limited.	Highly scalable with the distributed computation.
Ownership	Central authority owns the data.	Local institutions have their ownership.

Table – 2.*Difference between Traditional ML and Decentralized Deep Learning.*

4. EHR Security Techniques in Deep Learning:

There are several security mechanisms to enhance the security of the Electronic Health Record (EHR) such as Data Encryption, Differential Privacy, Homomorphic Encryption, Anomaly Detection [10].

4.1 Data Encryption

Data encryption is a key component of data security for electronic health records (EHR) because it protects the data from unauthorized access and cyber threats.

4.2 Differential Privacy

To preserve individual privacy, this can be achieved by adding noise to the data or model output using differential privacy.

4.3 Homomorphic Encryption

One kind of encryption approach is homomorphic encryption, which enables processing on encrypted data without requiring its first decryption.

4.4 Anomaly Detection

Anomaly detection is electronic health records (EHR) is the process of identifying unusual patterns or outliers in the healthcare data.

5. Conclusion

In this article, we discussed the various traditional machine learning and the decentralized deep learning approach to protect the data in the Electronic Health Records (EHR) and discussed the limitations of traditional machine learning approaches.

References

1. *Xu, J.; Xi, X.; Chen, J. Sheng, V.S.; Ma, J.; Cui, Z.: A Survey of Deep Learning for Electronic Health Records. Appl. Sci. 2022, 12, 11709.*
2. *C. Shiranthika et al: Decentralized Learning in Healthcare: A Review of Emerging Techniques. Vol. 11 June – 2023.*
3. *Benjamin Shickel, Patrick James Tighe, Azra Bihorac, and Parisa Rashidi: Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis.*
4. *Rosita Guido, Stefania Ferrisi, Danilo Lofaro, Domenico Conforti: An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review.*
5. *A. Sheik Abdullah et al. "Disseminating the Risk Factors with Enhancement in Precision Medicine" [Mar-2024].*
6. *B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arca's, ''Communication-efficient learning of deep networks from decentralized data,'' in Proc. AISTATS, 2017, pp. 1273–1282.*
7. *A. Rauniyar, D. H. Hagos, D. Jha, J. E. Håkegård, U. Bagci, D. B. Rawat, and V. Vlassov, ''Federated learning for medical applications: A taxonomy, current trends, challenges, and future research directions,''2022, arXiv:2208.033*
8. *Y. J. Ha et al.: Spatio-Temporal Split Learning for Privacy-Preserving Medical Platforms. Vol- 9, 2021.*
9. *V. Turina, Z. Zhang, F. Esposito, and I. Matta, ''Combining split and federated architectures for efficiency and privacy in deeplearning,'' in Proc. 16th Int. Conf. Emerg. Netw. EXperiments Technol., Nov. 2020, pp. 562–563.*
10. *William Hurst, Bedir Tekinerdogan, Tarek Alskaif,Aaron Boddy: Securing electronic health records against insider-threats: A Supervised Machine learning approach. Oct. 2022, pp. 2352-6483.*