Proceedings of International Mathematical Sciences
ISSN: 2717-6355, URL: https://dergipark.org.tr/tr/pub/pims
Volume 5 Issue 2 (2023), Pages 91-102.
Doi: https://doi.org/10.47086/pims.1390925

DECENTRALIZED MACHINE LEARNING APPROACH ON ICU ADMISSION PREDICTION FOR ENHANCED PATIENT CARE USING COVID-19 DATA

TAKESHI MATSUDA*, TIANLONG WANG**, AND MEHMET DIK***
*BELOIT COLLEGE, BELOIT, UNITED STATES. 0009-0004-1276-9236
**BELOIT COLLEGE, BELOIT, UNITED STATES. 0009-0002-0498-6682
***ROCKFORD UNIVERSITY, ROCKFORD, UNITED STATES. 0000-0003-0643-2771

Abstract. The Intensive Care Unit (ICU) represents a constrained healthcare resource, involving invasive procedures and high costs, with significant psychological effects on patients and their families. The traditional approach to ICU admissions relies on observable behavioral indicators like breathing patterns and consciousness levels, which may lead to delayed critical care due to deteriorating conditions. Therefore, in the ever-evolving healthcare landscape, predicting whether patients will require admission to the ICU plays a pivotal role in optimizing resource allocation, improving patient outcomes, and reducing healthcare costs. Essentially, in the context of the post-COVID-19 pandemic, aside from many other diseases, this prediction not only forecasts the likelihood of ICU admission but also identifies patients at an earlier stage, allowing for timely interventions that can potentially mitigate the need for ICU care, thereby improving overall patient outcomes and healthcare resource utilization. However, this task usually requires a lot of diverse data from different healthcare institutions for a good predictive model, leading to concerns regarding sensitive data privacy. This paper aims to build a decentralized model using deep learning techniques while maintaining data privacy among different institutions to address these challenges.

1. Introduction

The COVID-19 pandemic confronted health systems worldwide with an unprecedented challenge. According to the World Health Organization (WHO), approximately 14.9 million deaths were associated with this novel coronavirus during 2020 and 2021 [1]. Surging cases overwhelmed hospitals and depleted essential resources globally, especially in intensive care units (ICUs) where shortages of beds, equipment, and staff severely constrained life-saving care [2].

2020 *Mathematics Subject Classification.* Primary: 68T05, 68T07, 68Q32.
Key words and phrases. Machine Learning; Federated Learning; Data Privacy; Biomedical.

Submitted on 14.11.2023, Accepted on 07.12.2023.
Communicated by Huseyin CAKALLİ.

The ICU is a crucial but limited healthcare resource [3]. Especially under the context of the COVID-19 era, a large number of cases have particularly stressed ICU settings with an increased need for ICU beds [4]. As cases skyrocketed in pandemic hotspots from Wuhan, Italy, to New York, ICUs were immediately overloaded with exceeding capacity [5] [6]. This emergency of ICU and other medical resources scarcity extremely affected patient outcomes and mortality throughout the pandemic [7]. Additionally, medical treatment in the ICU has the disadvantages of possible invasive procedures [8], high cost, and significant psychological effects on both patients, their families and the medical institution [9], compared to the equivalent but earlier treatment outside the ICU. Moreover, the traditional approach to determining if someone should be admitted to the ICU primarily depends on observable indicators such as the patient's breathing pattern, consciousness, and medical instability, which means decisions for sending some patients without notable into ICU are made at relatively later points waiting until the patient's health condition has already deteriorated [10]. This decision-making strategy could potentially result in delayed medical treatment, thus leading to a poor survival rate and long-term effects on the patient's physical condition[10].

To summarize, the challenges and disadvantage of critical care is: limited resources, possible late admission decisions, and significant burden on different aspects of different groups. To solve the root cause, Machine Learning (ML) methods have been proven as a robust tool to reduce the necessity for patients to be sent to ICU by facilitating earlier clinical decision-making and critical care intervention which in turn helps with better ICU resource allocation [11] [12] [13]. A more robust model can help with making more accurate and reliable decisions for an earlier intervention, which will lead to a positive change in survival rate, long-term effects, and readmission rates among patients carrying a wide range of diseases [10] [14] [15]. However, creating a robust model that can produce reliable information also necessitates access to a wide range of diverse data from different institutions [16], which challenges data privacy and integrity significantly [17]. We recognize the importance of data privacy and the distributed nature of healthcare data, which provides significant challenges to this traditional centralized approach. On the one hand, healthcare data contains highly sensitive and private information, requiring high privacy protection measures. However, healthcare data also distributed across various countries and institutions, prevents data accessibility and holds back the development of accurate predictive models. To alleviate the need for data transfer between institutions, which is a primary concern of data privacy, this paper aims to deploy a predictive model using decentralized deep learning architecture that enables model transfer among different institutions to maintain data privacy, which is commonly known as Federated Learning (FL) [18].

FL has emerged as a promising approach for training machine learning models in the biomedical field, specifically in healthcare, to address the challenges of data privacy and data accessibility. By enabling collaborative model training without the need for centralizing patient data, FL allows healthcare institutions to collectively leverage their datasets while preserving data privacy [16]. For instance, in the pioneering publication on FL in the medical domain, Sheller et al. [19] have successfully applied FL on studying brain tumor. The results showed that the deep learning model trained using FL could reach 99% of the performance of the same model trained with the traditional data-sharing method, highlighting the potential of this technique in maintaining data privacy while effectively utilizing distributed healthcare data. Overall, this study aims to explore the potential of FL techniques on deep learning models in improving the accuracy of ICU admission prediction models and addressing the challenges posed by healthcare data privacy.

2. Methods

2.1. **Data Overview.** The original data is provided by the Mexican Government [20]. We translated the attributes and chose 21 medical-related features from the data set for this research (Table 1). Irrelevant features, like registration ID, migration status, and whether the patient speak an indigenous language or not, were dropped. The data set is being updated regularly. As of the day the research began, 1,048,575 records were collected.

Table 1. Table includes the features included in the dataset, 20 features and 1 target column.

Name	Type	Description
USMR	Categorical	medical units of the first, second or third level
Medical Unit	Categorical	type of institution that provided the care
Sex	Categorical	biological gender 1 for female and 2 for male
Patient Type	Categorical	type of care. (1 = returned; 2 = hospitalization)
Date Died	Date	the date of death
Intubed	Categorical	whether the patient was connected to the ventilator
Pneumonia	Categorical	air sacs inflammation in the past
Age	Discrete	years of age
Pregnant	Categorical	whether the patient is pregnant or not.
Diabetes	Categorical	whether the patient has diabetes or not
COPD	Categorical	Chronic obstructive pulmonary disease
Asthma	Categorical	whether the patient has asthma or not
INMSUPR	Categorical	whether the patient is immunosuppressed or not.
Hypertension	Categorical	whether the patient has hypertension or not
Other Disease	Categorical	whether the patient has other disease or not
Cardiovascular	Categorical	heart or blood vessels related disease
Obesity	Categorical	whether the patient is obese or not
Renal Chronic	Categorical	chronic renal disease
Tobacco	Categorical	whether the patient is a tobacco user
Classification Final	Discrete	covid test findings. $1 \sim 3$=COVID; ≥ 4=negative
ICU	Categorical	admitted to an Intensive Care Unit

2.2. **Data Preprocessing.** For this research, the categorical feature ICU was used as the target attribute for prediction. Besides AGE and DATE DIED, all other categorical features implied if a record has the diseases or not. Among them, CLASSIFICATION FINAL indicated whether a patient tested positive for COVID-19 or not. The original data used $1 \sim 3$ for positive, ≥ 4 for negative, and we converted this attribute into binary. DATE DIED indicates when the patient deceased. A empty value in this column indicated that the person survived. The DATE DIED column was transformed into a binary attribute. Eventually, this column was dropped and converted into records in ICU column. So, the final dataset, after dropping all records with a null value (about 1% of the dataset), includes 189112 records that are hospitalized. Within those records, Column ICU has 75011 (about 39.7% of the entire dataset) entries indicating this patient will need critical medical care and should consider early intervention. And, within all those 75011 records, 16397 records were originally included in the ICU column before the data processing. The rest 58641 records came from the hospitalized records that died without being sent into the ICU. Those records was originally from DATE DIED column (Figure 1). Because a patient died under hospitalized status but not in ICU indicates that they were supposed to received early intervention medical care for a potential better outcome, we combined the ICU and DATE DIED columns.

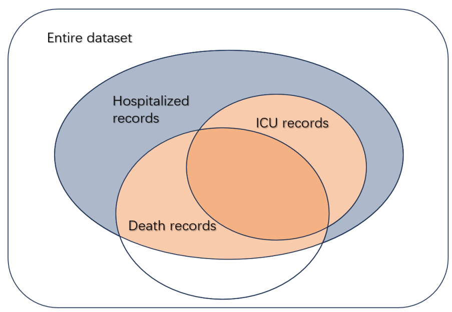

Figure 1. A Venn Diagram indicates the composition of the target column. The grey area is the records hospitalized. The orange area is the final target group, which consists of ICU records and records that died in the hospital but not in the ICU.

2.3. **Baseline Training.** To create a baseline understanding of our dataset and test data cleaning, a series of traditional machine learning techniques were performed on all of our datasets. Models like Decision Trees (DT) [21], Random Forests (RF) [22], Bayesian Classifiers (BC) [23], SVM [24], deep learning models like Convolutional Neural Networks (CNN) [25], and Recurrent Neural Networks (RNN) [26] were used. They mainly served as the comparison group that trained on the global dataset without considering data privacy.

2.4. **Federated Learning.** Federated learning (FL) is a decentralized machine learning approach that allows multiple devices or nodes to collaboratively update a shared model and hold local data samples [18] . In this research, the goal is to develop a federated learning architecture that can retain moderate accuracy, recall, and precision while not sharing the information between edges.

A global model was distributed physically to different medical institutions (Figure 2). After the local model was trained on each dataset. The updates from different medical institutions were sent back to make the update. The following Algorithm 1 explains the federated learning pipeline in more detail.

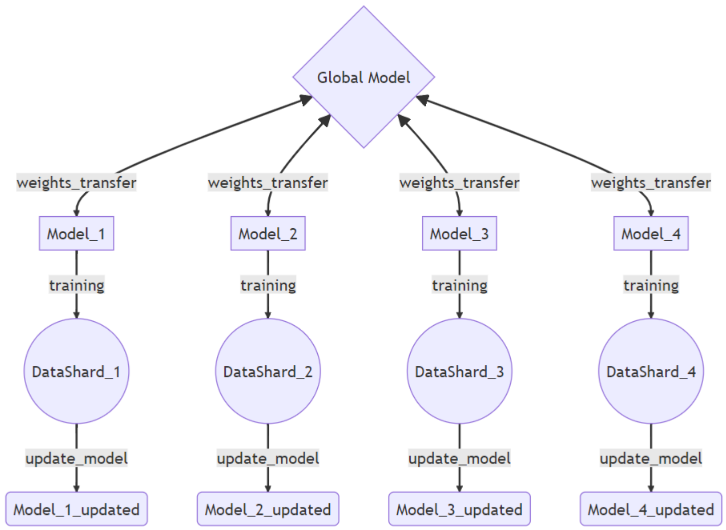

FIGURE 2. a round of Federated Learning pipeline with 4 datasets. On the top is the Global Model, the first level (top down) is the weight transfer back and forth between the global model and the models on different edges. The second level (Model to Datashard) indicates the training processes between locals models with local data. The third level (Datashard to Model updated) indicates that the updated weights of each local model are collected and ready to compile into one global model update.

Algorithm 1 Basic Federated Learning Architecture

$W_{0,0} \sim F_w$
$\alpha_1, \alpha_2, \alpha_3, \alpha_4$
$N = 100$
while $n < N$ **do**
  //initialize models on edges base on the global model
  $W_{n,1}, W_{n,2}, W_{n,3}, W_{n,4} = W_{n,0}$

  //Training each model on their dataset for one epoch
  $W'_{n,1}, W'_{n,2}, W'_{n,3}, W'_{n,4} \sim W_{n,1}, W_{n,2}, W_{n,3}, W_{n,4}$

  //update the global model weight based on the weighted average
  $W_{n+1,0} = \alpha_1 W'_{n,1} + \alpha_2 W'_{n,2} + \alpha_3 W'_{n,3} + \alpha_4 W'_{n,4}$

end while

The training ran 100 rounds in total. The local model trained on each edge for 1 epoch, which refers to local training. The updates of the global model were calculated based on the weighted average of the models from different shards. During the training, global accuracy and loss were monitored.

3. Results

3.1. **Baseline Training's performance.** Accuracy, recall, and precision are monitored for baseline training. Accuracy is the overall accuracy. Due to the nature of binary classification, the recall and the precision are measured on the class indicating the need for critical care and early intervention. Overall accuracy ranges from 70% (DT) to 76% (DNN). Precision ranges from 65 % (DT) to 78% (RNN). Recall ranges from 38% (SVM), to 59% (DNN) (Table 2).

Table 2. Baseline Training models' prediction performance, need add precision also

	Decision Tree	Random Forest	SVM	Bayesian Classifier	DNN	CNN	RNN
Acc	70.82%	72.68%	73.49%	76.30%	76.39%	76.25%	76.26%
Precision	65.74%	67.95%	88.22%	81.16%	76.14%	77.56%	78.25%
Recall	55.34%	59.0%	38.35%	52.50%	59.02%	56.53%	55.68%

3.2. **Federated Learning's performance.**

3.2.1. *Accuracy.* Accuracy serves as a general measurement of the architectures predicting power.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Deep learning models like CNN, RNN, and DNN are used separately as the base models for FL. The accuracy ranges from 76.18% to 76.28% on the global dataset and performs equivalently well on each data shard (Table. 3).

Table 3. Federated Learning model Accuracy

Base Models	Global	Shard1	Shard2	Shard3	Shard4
DNN	76.22%	76.14%	76.61%	76.59%	75.97%
CNN	76.18%	76.14%	76.40%	76.28%	75.90%
RNN	76.28%	76.17%	76.57%	76.40%	76.00%

3.2.2. *Precision.* Precision is measured by True Positive rate over True Positive rate and False Positive rate. A way to interpret this is how many correct predictions the model made about a class are correct among all predictions made for this class.

$$Precision = \frac{TP}{TP + FP}$$

The precision of three FL architectures ranges from 74.75% by the CNN-based model to 75.20% by the RNN-based model (Table. 3).

Table 4. Federated Learning architecture precision

Base Models	Global	Shard1	Shard2	Shard3	Shard4
DNN	74.83%	74.59%	74.54%	75.43%	74.76%
CNN	74.75%	74.82%	74.20%	75.13%	74.83%
RNN	75.20%	75.05%	74.70%	75.70%	75.34%

3.2.3. *Recall.* Recall is measured by the True Positive record number divided by the sum of the True Positive record number and the False Negative record number. Recall can be interpreted as among all records that need early intervention, how many of them are successfully detected.

$$Recall = \frac{TP}{TP + FN}$$

The recall rate for FL with DL models ranges from 60.07% by the RNN-based model, to 60.84% by the DNN-based model (Table. 5).

Table 5. Federated Learning architecture recall

Base Models	Global	Shard1	Shard2	Shard3	Shard4
DNN	60.84%	60.81%	60.9%	61.05%	60.61%
CNN	60.40%	60.44%	60.61%	60.34%	60.20%
RNN	60.07%	60.17%	60.43%	59.90%	59.78%

4. Discussion

Our FL architecture with Deep Learning models reached 99.8% accuracy of the baseline modeling, where data privacy is not well preserved (Figure 3). FL with DNN models achieved 76.3% accuracy, which surpass all machine learning models and some deep learning models that are trained on congregated dataset.

In the research, the focus is on the prediction of records that actually need early medical care to prevent ICU entrance. So, the higher the measurement of the precision, the more records that are predicted as needing ICU-level treatment are correct. The precision of our FL architecture reached 85% (75.2% by FL with RNN models compared to 88.22% by Bayesian Classifier) of maximum precision from baseline training (Figure 4). However, the recall of Bayesian Classifier is considered extremely low, only 38.35%. Therefore, Bayesian Classifier should be considered as an outlier and not considered for the comparison. Then, the precision of the highest FL architecture reached 92.7% of the highest precision by the SVM from baseline training.

Recall plays a vital role in real-world applications, too. Based on the focus of this research, the measuring for recall indicates how many actual records that need ICU entrance have been successfully detected. The recall of our FL architecture outperforms all the other models. The lowest recall by the CNN-based model in FL architecture obtained a recall of 60.4%, compared to the highest recall by DNN

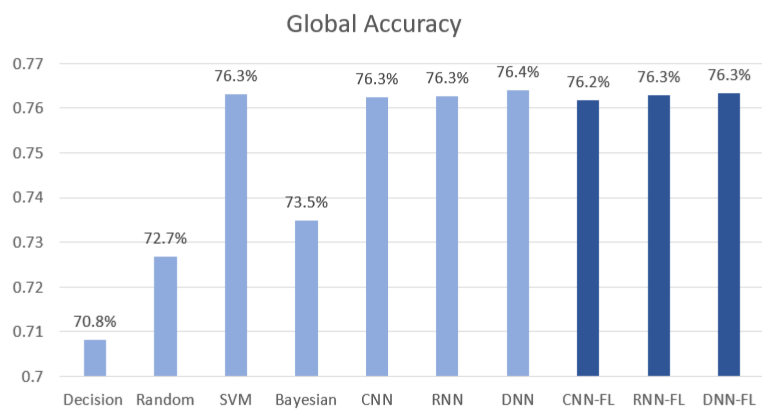

Figure 3. Global Accuracy for the FL. X-axis indicating with model is trained and tested, the y-axis indicating the accuracy. Dark blue indicating the federated learning design, and light blue indicating the baseline traditional machine learning models' results.

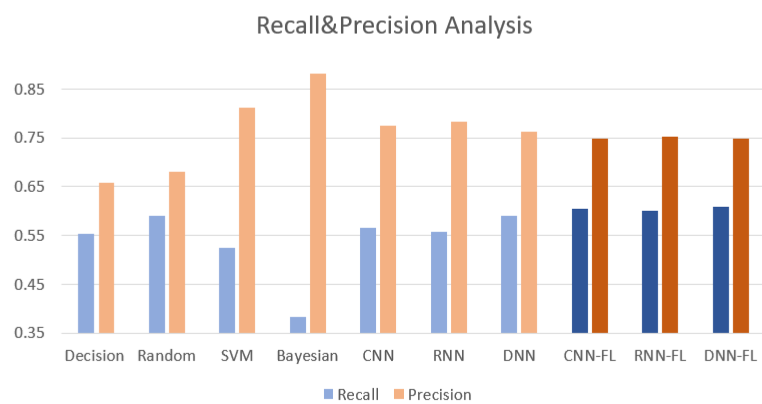

Figure 4. Global Recall and Precision Analysis. Each pair of columns consists of Recall (left, blue) and Precision (right, organce). The darker pairs indicating the federated learning models and its corresponding precision and recall. X-axis is the models used, y-axis is the percentage for recall and precision.

alone with a recall of 59.02% in baseline training. This indicates our FL architecture can successfully predict more records that need earlier ICU-level treatment. The result also implies FL architecture can improve the recall of target attributes in general.

Considering all three valuation factors, the FL architecture has proven to be a robust tool for better overall accuracy and recall, and tantamount precision than traditional machine learning. Most importantly, FL reaches the predicting power under the circumstance of keeping the privacy of data under no risk of leaking or retrieving. The scalability of the architecture from a data perspective, the main concern about ML in the medical field [17], can be improved.

5. Limitations & Future Work

Developing robust, trustworthy AI tools that preserve fairness is critically important, especially in high-stakes applications like healthcare. Achieving trustworthiness encompasses attributes like explainability, fairness, privacy preservation, and robustness. However, the overall prediction performance of health AI models is often prioritized over potential biases they may have [27]. In FL, there are potential risks of under-representing minority groups, if the contribution to the global model from different edges is guided by the training size, which is statistical heterogeneity [28]. In the future, we would like to explore the potential of leveraging fairness through multiple methods, like local debiasing [29] and fairer aggregation strategies[28].

Although the FL system aims to address privacy concerns by keeping patients' private data in local storage during training, potential security issues persist, particularly in the transmission of gradients and partial parameters, leading to indirect privacy leakage [30]. Three main attack categories in FL are identified: Data poisoning attacks, involving the embedding of tainted data to compromise data integrity [31]; Model poisoning, which manipulate machine learning models to produce incorrect results [31]; and Inferring attacks, focused on detecting privacy records or restoring training data [32]. Existing defense methodologies have some potential in more research, and the need for stronger protection measures, such as anomaly detection and data encryption, is emphasized to mitigate these attacks in the federated setting [31][32]. Future work on this should explore and develop more robust protection methods.

There is rich literature discussing whether FL overfits or underfits under different data quality, parameters' sizes, and extents of the local updates [33] [34] [35]. Evaluation of overfitting and underfitting usually requires a validation dataset during the training phase. However, traditionally collecting a validation dataset violates the main data privacy protection schema provided by FL. A representative and effective validation set needs to combine a certain amount of data from each dataset on the edge, but a congregated dataset is what FL trying to avoid due to privacy concerns. Moreover, validation and testing datasets are usually not directly accessible to the FL server [36], and the global model is tested on selected clients or data shards separately. In the future, our team will investigate more about the necessity of evaluating the global model of FL and its corresponding metrics.

An ICU decision will potentially put pressure on both medical institutions as well as the patients themselves [9], both mentally and physically. Traditional Machine Learning, even Federated Learning, produces a one-number confident prediction that might worsen this situation. Ethically, using probability estimation instead of one-point prediction could be a challenging but effective improvement to this situation. Plus, due to the nature of most probability predictions that produce a distribution of predictions as the output, incorporating differential privacy can add an extra layer of data protection as well as help balance the trade-off between privacy and accuracy [37]. Inspired by this idea, we would like to further investigate the feasibility of incorporating probability estimation and differential privacy in FL architectures.

6. Conclusion

In conclusion, Federated Learning demonstrated to be an effective tool to help clinical decision-making without losing data privacy. Particularly, our design of FL outperformed other traditional machine learning and deep learning techniques on the ICU admission data set. This design and architecture imply that, with the help of FL, medical institutions can potentially make more effective decisions regarding early interventions on patients to improve the treatment outcome, critical medical resource allocation, and alleviation of avoidable burdens on both sides. Besides that, this paper also tried to raise public's awareness of data privacy and ethics to encourage us to rethink our machine learning pipeline when building models for supporting clinical decision-making.

References

[1] World Health Organisation, "14.9 million excess deaths associated with the covid-19 pandemic in 2020 and 2021," 2022.

[2] Y. M. Arabi, S. N. Myatra, and S. M. Lobo, "Surging icu during covid-19 pandemic: an overview," *Current Opinion in Critical Care*, vol. 28, no. 6, p. 638, 2022.

[3] G. R. Gristina and M. Piccinni, "Covid-19 pandemic in icu. limited resources for many patients: approaches and criteria for triaging," *Minerva Anestesiologica*, vol. 87, 2021.

[4] G. L. Anesi and M. P. Kerlin, "The impact of resource limitations on care delivery and outcomes: routine variation, the coronavirus disease 2019 pandemic, and persistent shortage," 2021.

[5] R.-H. Du, L.-M. Liu, W. Yin, W. Wang, L.-L. Guan, M.-L. Yuan, Y.-L. Li, Y. Hu, X.-Y. Li, B. Sun, *et al.*, "Hospitalization and critical care of 109 decedents with covid-19 pneumonia in wuhan, china," *Annals of the American Thoracic Society*, vol. 17, no. 7, pp. 839–846, 2020.

[6] A. Olivas-Martinez, J. L. Cárdenas-Fragoso, J. V. Jiménez, O. A. Lozano-Cruz, E. Ortiz-Brizuela, V. H. Tovar-Méndez, C. Medrano-Borromeo, A. Martinez-Valenzuela, C. M. Román-Montes, B. Martinez-Guerra, *et al.*, "In-hospital mortality from severe covid-19 in a tertiary care center in mexico city; causes of death, risk factors and the impact of hospital saturation," *PloS one*, vol. 16, no. 2, p. e0245772, 2021.

[7] G. French, M. Hulse, D. Nguyen, K. Sobotka, K. Webster, J. Corman, B. Aboagye-Nyame, M. Dion, M. Johnson, B. Zalinger, *et al.*, "Impact of hospital strain on excess deaths during the covid-19 pandemic—united states, july 2020–july 2021," *Morbidity and Mortality Weekly Report*, vol. 70, no. 46, p. 1613, 2021.

[8] P. Eggimann, J. Bille, and O. Marchetti, "Diagnosis of invasive candidiasis in the icu," *Annals of Intensive Care*, vol. 1, 2011.

[9] S. p. Kafantaris and O. Kadda, "Advantages and disadvantages of patients' hospitalization in intensive care units," *Health amp; Research Journal*, vol. 7, p. 155–159, Oct. 2021.

[10] S. A. Candan, N. Elibol, and A. Abdullahi, "Consideration of prevention and management of long-term consequences of post-acute respiratory distress syndrome in patients with covid-19," *Physiotherapy Theory and Practice*, vol. 36, 2020.

[11] A. Dauvin, C. Donado, P. Bachtiger, K. C. Huang, C. M. Sauer, D. Ramazzotti, M. Bonvini, L. A. Celi, and M. J. Douglas, "Machine learning can accurately predict pre-admission baseline hemoglobin and creatinine in intensive care patients," *npj Digital Medicine*, vol. 2, 2019.

[12] B. Lee, K. Kim, H. Hwang, Y. S. Kim, E. H. Chung, J. S. Yoon, H. J. Cho, and J. D. Park, "Development of a machine learning model for predicting pediatric mortality in the early stages of intensive care unit admission," *Scientific Reports*, vol. 11, 2021.

[13] S. Subudhi, A. Verma, A. B. Patel, C. C. Hardin, M. J. Khandekar, H. Lee, D. McEvoy, T. Stylianopoulos, L. L. Munn, S. Dutta, and R. K. Jain, "Comparing machine learning algorithms for predicting icu admission and mortality in covid-19," *npj Digital Medicine*, vol. 4, 2021.

[14] S. Schmidt, J. K. Dieks, M. Quintel, and O. Moerer, "Critical care echocardiography as a routine procedure for the detection and early treatment of cardiac pathologies," *Diagnostics*, vol. 10, 2020.

[15] R. Sangani, E. Mokaya, H. Mujahid, S. Hadique, S. Culp, L. Constantine, and A. Moss, "Early palliative care intervention reduces icu readmissions in high-risk patients," *Chest*, vol. 158, 2020.

[16] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," *npj Digital Medicine*, vol. 3, 2020.

[17] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: A survey and outlook," 2021.

[18] T. Yang, G. Andrew, H. Eichner, H. Sun, W. Li, N. Kong, D. Ramage, and F. Beaufays, "Applied federated learning: Improving google keyboard query suggestions," *arXiv preprint arXiv:1812.02903*, 2018.

[19] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*, pp. 92–104, Springer, 2019.
[20] SALUD, "Información referente a casos covid-19 en méxico," 2023.
[21] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 18, no. 6, pp. 275–285, 2004.
[22] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, pp. 197–227, 2016.
[23] P. Langley, W. Iba, K. Thompson, *et al.*, "An analysis of bayesian classifiers," in *Aaai*, vol. 90, pp. 223–228, Citeseer, 1992.
[24] S. Suthaharan and S. Suthaharan, "Support vector machine," *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, pp. 207–235, 2016.
[25] Y. LeCun, Y. Bengio, *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
[26] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities.," *Proceedings of the national academy of sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.
[27] E. Röösli, S. Bozkurt, and T. Hernandez-Boussard, "Peeking into a black box, the fairness and generalizability of a mimic-iii benchmarking model," *Scientific Data*, vol. 9, 2022.
[28] R. Poulain, M. F. B. Tarek, and R. Beheshti, "Improving fairness in ai models on electronic health records: The case for federated learning methods," 2023.
[29] J. Hong, Z. Zhu, S. Yu, Z. Wang, H. H. Dodge, and J. Zhou, "Federated adversarial debiasing for fair and transferable representations," 2021.
[30] J. W. Bos, K. Lauter, and M. Naehrig, "Private predictive analysis on encrypted medical data," *Journal of Biomedical Informatics*, vol. 50, 2014.
[31] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," vol. 108, 2020.
[32] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, and H. Qi, "Beyond inferring class representatives: User-level privacy leakage from federated learning," vol. 2019-April, 2019.
[33] A. Barros, D. Rosário, E. Cerqueira, and N. L. da Fonseca, "A strategy to the reduction of communication overhead and overfitting in federated learning," in *Anais do XXVI Workshop de Gerência e Operação de Redes e Serviços*, pp. 1–13, SBC, 2021.
[34] L. Lyu, X. Xu, Q. Wang, and H. Yu, "Collaborative fairness in federated learning," *Federated Learning: Privacy and Incentive*, pp. 189–204, 2020.
[35] X. Zhang, Y. Li, W. Li, K. Guo, and Y. Shao, "Personalized federated learning via variational bayesian inference," in *International Conference on Machine Learning*, pp. 26293–26310, PMLR, 2022.
[36] A. Li, L. Zhang, J. Tan, Y. Qin, J. Wang, and X. Y. Li, "Sample-level data selection for federated learning," vol. 2021-May, 2021.
[37] Q. Zhang, Z. Bu, K. Chen, and Q. Long, "Differentially private bayesian neural networks on accuracy, privacy and reliability," vol. 13716 LNAI, 2023.

Takeshi MATSUDA,
700 College Street, Beloit, WI, USA, 53511 Phone: +1 (608) 473-9572, ORCID: 0009-0004-1276-9236
Email address: `takeshimatsuda27@gmail.com`

Tianlong WANG,
700 College Street, Beloit, WI, USA, 53511, Phone: +1 (608) 718-6769, ORCID: 0009-0002-0498-6682
Email address: `wangtianlong1207@gmail.com`

Mehmet DIK,
5050 E State St, Rockford, IL, USA, 61108, Phone: +1 (815) 986-9524, ORCID: 0000-0003-0643-2771
Email address: `mdik@rockford.edu`