IJITCE

ISSN 2347- 3657

International Journal of
Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686

ENHANCING IN-HOSPITAL MORTALITY PREDICTION VIA PERSONALIZED FEDERATED LEARNING IN MULTI INSTITUTIONAL ICUS

[1] *V. Ravi Kumar*, *MCA Student, Department of MCA*

[2] *Shaik Haseena, M.Tech, (Ph.D), Assistant Professor, Department of MCA*

[12] *Dr KV Subba Reddy Institute of Technology, Dupadu, Kurnool*

ABSTRACT

A fresh and promising plan to advance machine learning (ML) at many independently dispersed healthcare facilities is provided by federated learning (FL), a paradigm for resolving the difficulties of applying ML to private distributed data. Its effectiveness may be hampered, however, by the non-IID and uneven distribution of the data, which may even make the institutions less inclined to take part in its training. In order to retain the original non-IID and unbalanced data distribution, this research investigated the issue of an in-hospital mortality prediction job under a real multi-center ICU electronic health record database. Prior to proposing a personalised FL (PFL) solution called POLA to address the issue, it first examined the cause of the baseline FL's performance decrease in this data setting. POLA is a customised one-shot, two-step FL technique that may provide very effective customised models for every individual participant. In studies, the suggested POLA approach was contrasted with two other PFL techniques, and the findings show that it not only considerably lowers the number of communication cycles but also enhances FL's prediction performance. Additionally, because of its generality and flexibility, it may be expanded to other comparable cross-silo FL application contexts.

I. INTRODUCTION

An enormous quantity of EHR data has surfaced as a result of the promotion of electronic health record (EHR) systems [1]. The use of machine learning (ML) in digital health is supported by EHR databases, which provide comprehensive data including patient diagnosis and treatment. Furthermore, ML has become one of the most popular tools in secondary analysis due to its abundance of resources and useful implicit information [2]. However, the use of classical machine learning (ML), which refers to centralising or releasing this data, presents not only ethical, legal, and regulatory issues but also technological ones because of the privacy and sensitivity of EHRs [3]. Although there are some corresponding ways to circumvent these limitations, like deleting certain important information to make patient data anonymous or incorporating privacy-preserving algorithms into the transmission process to stop data leaks [4], the aforementioned issue has not been fully resolved because data migration is still required.

Federated learning (FL) [5], [6], which arose as a model to tackle the issue of machine learning on private dispersed data sources, offers encouraging opportunities to advance ML in the area of digital healthcare [7]. In compliance with user privacy protection, data security, and governmental rules, it is a distributed machine learning environment that can efficiently support several independent clients, including mobile phones, IOT devices, and organisations, to carry out isolated data consumption and ML modelling [8]. at the healthcare industry, FL may deploy machine learning (ML) at separate institutions without exchanging any raw EHR data. This allows for the exchange of useful and common information that is housed in the isolated data silos while maintaining patient privacy and

https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686

sensitive data. A general decentralised framework for predicting hospitalisation due to cardiac events [10], predicting the mortality rate and length of stay in the intensive care unit [11], including that under COVID-19 [12], and identifying clinically similar patients across institutions to support medical research and applications are all examples of how FL can be used in standard EHR applications.

Although FL has been shown to be practical and successful in EHRs from separate institutions, the unbalanced and non-independently and identically distributed (non-IID) nature of these EHR data silos might impair its efficacy. In particular, the non-IID feature may lead to a substantial decrease in the efficacy of the model in FL, such as a loss of prediction accuracy in the local ML models of the clients [13], [14], and [15]. This circumstance may be exacerbated by the data imbalance [6]. Moreover, FL performance is significantly impacted by the skewness of non-IID datasets, or the divergence of IID data. One might even argue that the degree to which the data distribution skews towards non-IID determines whether FL's validity on non-IID data can be ensured [16]. Because the FL model's performance will be impacted when this skewness reaches a particular point, leading to an accuracy-loss that rises as the skewness grows [17], [18]. All things considered, the non-IID and unbalanced data distribution nature of FL in a multi-institution EHR scenario can lower model performance, even leading to locally independent trained models performing better than the FL-trained model. This eliminates the primary motivation for these healthcare organisations to participate in FL and even renders FL meaningless.

Numerous FL optimisation strategies have been developed to meet the aforementioned difficulty; D. Ting et al. [19] have compiled these techniques and separated them into two categories: local adaptation and global optimisation. Each participant might get a customised model instead of accepting a single, shared model thanks to the local adaption techniques that are specifically suggested to address the statistical difficulties in FL. Personalised federated learning (PFL), which combines an early simple "FL training + local adaptation" method with many later approaches [20], has now gained popularity as a field of study [21]. According to a number of customised FL research [20], [22], when faced with heterogeneous data settings like non-IID and imbalanced distributions, FL may recover from performance loss by customising each person's local models with their unique data.

In keeping with the idea behind PFL approaches, we contended that in the non-IID and imbalanced data environment, it is no longer feasible to create a single functional model for every FL participant. Therefore, we tweak the optimisation issue of the conventional FL and propose a Personalised One-shot Local Adaptation (POLA) FL approach to address the difficulty. The goal of the suggested approach is to enhance in-hospital mortality prediction performance in a real multiple independent intensive care unit setting. Furthermore, we naturally split the dispersed ICU datasets in two ways to create ICU centres with varying non-IID data skewness while maintaining the real data distribution in order to further confirm the efficacy of the suggested approach. In this data setting, experiments show that POLA may dramatically decrease the amount of communication cycles of FL training while also improving the model's mortality prediction ability.

This work's primary contributions are as follows: 1) We conducted experiments on baseline FL in the context of the study's data to support our research challenge. 2) We devised a

https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686

PFL approach dubbed POLA to produce highly personalised models for autonomous ICU centres by converting the basic global optimisation problem of regular FL into a problem optimised for each individual. 3) To show that the POLA not only enhances model performance but also successfully lowers the communication overhead of FL, we conducted an experimental comparison with baseline FL and two additional PFL techniques.

This is how the remainder of the paper is structured. The basic information on baseline FL, personalised FL, federated knowledge distillation, and auto ML is presented in Section II. The comprehensive designs of our suggested customised FL system are shown in Section III. Section IV presents the analysis and assessment of the experiment. Lastly, Sections V and VI discuss and wrap up the work, respectively.

II. LITERATURE SURVEY

‘‘Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis,’’

B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi,

The past decade has seen an explosion in the amount of digital information stored in electronic health records (EHR). While primarily designed for archiving patient information and performing administrative healthcare tasks like billing, many researchers have found secondary use of these records for various clinical informatics applications. Over the same period, the machine learning community has seen widespread advances in the field of deep learning. In this review, we survey the current research on applying deep learning to clinical tasks based on EHR data, where we find a variety of deep learning techniques and frameworks being applied to several types of clinical applications including information extraction, representation learning, outcome prediction, phenotyping, and de-identification. We identify several limitations of current research involving topics such as model interpretability, data heterogeneity, and lack of universal benchmarks. We conclude by summarizing the state of the field and identifying avenues of future deep EHR research.

‘‘Prediction modeling using HER data: Challenges, strategies, and a comparison of machine learning approaches,’’

J. Wu, J. Roy, and W. F. Stewart,

Electronic health record (EHR) databases contain vast amounts of information about patients. Machine learning techniques such as Boosting and support vector machine (SVM) can potentially identify patients at high risk for serious conditions, such as heart disease, from EHR data. However, these techniques have not yet been widely tested. To model detection of heart failure more than 6 months before the actual date of clinical diagnosis using machine learning techniques applied to EHR data. To compare the performance of logistic regression, SVM, and Boosting, along with various variable selection methods in heart failure prediction. Geisinger Clinic primary care patients with data in the EHR data from 2001 to 2006 diagnosed with heart failure between 2003 and 2006 were identified. Controls were randomly selected matched on sex, age, and clinic for this nested case-control study. Area under the curve (AUC) of receiver operator characteristic curve was computed for each method using 10-fold cross-validation. The number of variables selected by each method was compared. Logistic regression with model selection based on Bayesian information criterion provided the most parsimonious model, with about 10 variables selected on average, while maintaining a high AUC (0.77 in 10-fold cross-validation). Boosting with strict variable importance threshold provided similar performance. Heart failure was predicted more than 6 months before clinical diagnosis, with AUC of about 0.76, using logistic regression and Boosting. These results were achieved even with

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686
strict model selection criteria. SVM had the poorest performance, possibly because of imbalanced data.

‘‘A systematic review of barriers to data sharing in public health,’’

G. van Panhuis, P. Paul, C. Emerson, J. Grefenstette, R. Wilder, A. J. Herbst, D. Heymann, and D. S. Burke,

Background: In the current information age, the use of data has become essential for decision making in public health at the local, national, and global level. Despite a global commitment to the use and sharing of public health data, this can be challenging in reality. No systematic framework or global operational guidelines have been created for data sharing in public health. Barriers at different levels have limited data sharing but have only been anecdotally discussed or in the context of specific case studies. Incomplete systematic evidence on the scope and variety of these barriers has limited opportunities to maximize the value and use of public health data for science and policy. Methods: We conducted a systematic literature review of potential barriers to public health data sharing. Documents that described barriers to sharing of routinely collected public health data were eligible for inclusion and reviewed independently by a team of experts. We grouped identified barriers in a taxonomy for a focused international dialogue on solutions. Results: Twenty potential barriers were identified and classified in six categories: technical, motivational, economic, political, legal and ethical. The first three categories are deeply rooted in well-known challenges of health information systems for which structural solutions have yet to be found; the last three have solutions that lie in an international dialogue aimed at generating consensus on policies and instruments for data sharing. Conclusions: The simultaneous effect of multiple interacting barriers ranging from technical to intangible issues has greatly complicated advances in public health data sharing. A systematic framework of barriers to data sharing in public health will be essential to accelerate the use of valuable information for the global good.

‘‘Publishing data from electronic health records while preserving privacy: A survey of algorithms,’’

A. Gkoulalas-Divanis, G. Loukides, and J. Sun,

The dissemination of Electronic Health Records (EHRs) can be highly beneficial for a range of medical studies, spanning from clinical trials to epidemic control studies, but it must be performed in a way that preserves patients' privacy. This is not straightforward, because the disseminated data need to be protected against several privacy threats, while remaining useful for subsequent analysis tasks. In this work, we present a survey of algorithms that have been proposed for publishing structured patient data, in a privacy-preserving way. We review more than 45 algorithms, derive insights on their operation, and highlight their advantages and disadvantages. We also provide a discussion of some promising directions for future research in this area.

‘‘Federated learning: Strategies for improving communication efficiency,’’

J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon,

Federated Learning is a machine learning setting where the goal is to train a high-quality centralized model while training data remains distributed over a large number of clients each with unreliable and relatively slow network connections. We consider learning algorithms for this setting where on each round, each client independently computes an update to the current model based on its local data, and communicates this update to a central server, where the client-side updates are aggregated to compute a new global model. The typical clients in this setting are mobile phones, and communication efficiency is of the utmost importance.

In this paper, we propose two ways to reduce the uplink communication costs: structured updates, where we directly learn an update from a

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686

restricted space parametrized using a smaller number of variables, e.g. either low-rank or a random mask; and sketched updates, where we learn a full model update and then compress it using a combination of quantization, random rotations, and subsampling before sending it to the server. Experiments on both convolutional and recurrent networks show that the proposed methods can reduce the communication cost by two orders of magnitude

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM

PFL research has exploded in recent years. To solve the issue of the unified global model's poor generalisation in FL when dealing with a data heterogeneity problem, a variety of PFL techniques have been devised [21]. Since local adaptation is used in this work to customise FL, we also provide the following quick summary of similar techniques:

1) Adjusting the model. Simply fine-tuning all or some of the parameters of the global model derived by FL training using private data locally on the client may result in performance increases in extremely diverse data [18], [24].

2) Regularisation of local loss. In order to get better-performing personalised models, regularisation loss is added to the local training procedure to mitigate the client-drift issue brought on by data heterogeneity [25], [26].

3) Meta-learning. Its typical method in FL is to first learn a parameterised model (or meta-learner) using algorithms like as Reptile and MAML in the FL training process, after which a customised model for each client may be quickly taught with the help of the meta-learner [27], [28].

4) In line with the process of local adaptation for FL, multi-task learning seeks to learn different models for many related tasks at once. [29], [30].

5) Transfer learning helps learners perform better by facilitating the exchange of information across related areas. It assists the client models in completing the local adaptation in the FL environment in a heterogeneous data situation, resulting in customised models [31], [22].

6) Knowledge distillation (KD) may be linked to FL in order to extract information such as logit vectors [33] and classification scores [32] of the global model to help the local client models develop their customised models.

While all of these PFL techniques may enhance FL's performance on non-IID data situations, they vary in how they further customise ML models. For instance, the parameters of the global model learnt in FL are personalised via model fine-tuning, meta-learning, multi-task learning, and transfer learning. In the FL learning process, local loss regularisation allows each model's loss function to be customised. Knowledge distillation may concurrently customise hyperparameters and the parameters and structure of particular models. In order to improve FL performance as much as possible, this effort attempts to customise the models as much as feasible. As a result, the KD method with the most possibility for model customisation is used. Its associated FL applications are examined in the next section.

Disadvantages

- In order to optimise personalised models, an existing system employed a heuristic technique that used automatic machine learning (AutoML), which may be mistaken for previous similar research.
- Almost all of the federated AutoML research that has been done so far is on the NAS of DNN models, particularly convolutional neural networks (CNNs). The autonomous construction and optimisation of the DNN model may provide the most significant advantages as its structure greatly affects the communication overhead and FL performance.

PROPOSED SYSTEM

The goal of the suggested approach is to enhance in-hospital mortality prediction performance in a real multiple independent intensive care unit

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686
setting. Furthermore, we naturally split the dispersed ICU datasets in two ways to create ICU centres with varying non-IID data skewness while maintaining the real data distribution in order to further confirm the efficacy of the suggested approach. In this data setting, experiments show that POLA may dramatically decrease the amount of communication cycles of FL training while also improving the model's mortality prediction ability.

This work's primary contributions are as follows:

1) In the data context of this work, we conducted tests on baseline FL to support our research challenge.

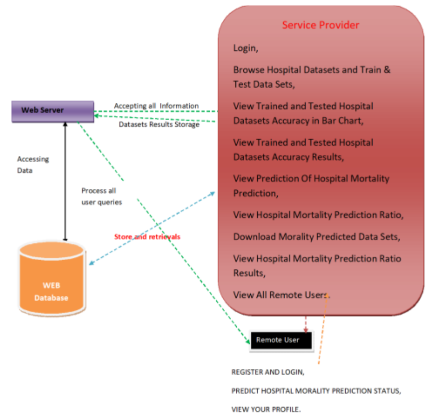
2) In order to create highly customised models for autonomous intensive care units, we first converted the initial global optimisation issue of ordinary FL into a problem optimised for each individual. Next, we suggested a PFL technique dubbed POLA.

3) To show that the POLA not only enhances model performance but also successfully lowers the communication overhead of FL, we conducted an experimental comparison with baseline FL and two additional PFL techniques.

Advantages

- An overview of the suggested two-step, one-shot PFL strategy is provided in the system. The term "two-step" here refers to FL training and local adaptation, where the former is used to create a common model with sufficient global generalisation experiment and the latter is used to create high-performance customised models for autonomous people.
- A traditional heuristic method called Genetic Algorithm (GA) is presented in order to automate and simplify it. GA is a traditional and efficient evolutionary method that uses crossover, mutation, and selection to get the best answer. It can concurrently provide a broad search space and ideal answers for the model structures and hyperparameters in this research that need automated design.

SYSTEM ARCHITECTURE



IV. IMPLEMENTATIONS

Modules

Service Provider

The Service Provider must use a working user name and password to log in to this module. After successfully logging in, he may do certain tasks such Examine Train & Test Data Sets and Hospital Datasets. View the accuracy of trained and tested hospital datasets in a bar chart, view the accuracy of trained and tested hospital datasets, view the hospital mortality prediction ratio, and view the prediction of hospital mortality. View Hospital Mortality Prediction Ratio Results, Download Morality Predicted Data Sets, and View All Remote Users.

View and Authorize Users

The administrator may see a list of all registered users in this module. Here, the administrator may see the user's information, like name, email, and address, and they can also grant the user permissions.

Remote User

A total of n users are present in this module. Before beginning any actions, the user needs register. Following registration, the user's

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686

information will be entered into the database. Following a successful registration, he must use his password and authorised user name to log in. Following a successful login, the user will be able to see their profile, predict hospital morality, register and log in, and more.

ALGORITHMS

Logistic regression Classifiers

The relationship between a collection of independent (explanatory) factors and a categorical dependent variable is examined using logistic regression analysis. When the dependent variable simply has two values, like 0 and 1 or Yes and No, the term logistic regression is used. When the dependent variable contains three or more distinct values, such as married, single, divorced, or widowed, the technique is sometimes referred to as multinomial logistic regression. While the dependent variable's data type differs from multiple regression's, the procedure's practical application is comparable.

When it comes to categorical-response variable analysis, logistic regression and discriminant analysis are competitors. Compared to discriminant analysis, many statisticians believe that logistic regression is more flexible and appropriate for modelling the majority of scenarios. This is due to the fact that, unlike discriminant analysis, logistic regression does not presume that the independent variables are regularly distributed.

Both binary and multinomial logistic regression are calculated by this software for both category and numerical independent variables. Along with the regression equation, it provides information on likelihood, deviance, odds ratios, confidence limits, and quality of fit. It does a thorough residual analysis that includes diagnostic residual plots and reports. In order to find the optimal regression model with the fewest independent variables, it might conduct an independent variable subset selection search. It offers ROC curves and confidence intervals on expected values to assist in identifying the optimal classification cutoff point. By automatically identifying rows that are not utilised throughout the study, it enables you to confirm your findings.

Naïve Bayes

The supervised learning technique known as the "naive bayes approach" is predicated on the straightforward premise that the existence or lack of a certain class characteristic has no bearing on the existence or nonexistence of any other feature. However, it seems sturdy and effective in spite of this. It performs similarly to other methods of guided learning. Numerous explanations have been put forward in the literature. We emphasise a representation bias-based explanation in this lesson. Along with logistic regression, linear discriminant analysis, and linear SVM (support vector machine), the naive bayes classifier is a linear classifier. The technique used to estimate the classifier's parameters (the learning bias) makes a difference.

Although the Naive Bayes classifier is commonly used in research, practitioners who want to get findings that are useful do not utilise it as often. On the one hand, the researchers discovered that it is very simple to build and apply, that estimating its parameters is simple, that learning occurs quickly even on extremely big datasets, and that, when compared to other methods, its accuracy is rather excellent. The end users, however, do not comprehend the value of such a strategy and do not get a model that is simple to read and implement.

As a consequence, we display the learning process's outcomes in a fresh way. Both the deployment and comprehension of the classifier are simplified. We discuss several theoretical facets of the naive bayes classifier in the first

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686

section of this lesson. Next, we use Tanagra to apply the method on a dataset. We contrast the outcomes (the model's parameters) with those from other linear techniques including logistic regression, linear discriminant analysis, and linear support vector machines. We see that the outcomes are quite reliable. This helps to explain why the strategy performs well when compared to others. We employ a variety of tools (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b, and RapidMiner 4.6.0) on the same dataset in the second section. Above all, we make an effort to comprehend the outcomes.

Random Forest

Random forests, also known as random decision forests, are ensemble learning techniques that build a large number of decision trees during training for tasks like regression and classification. The class chosen by the majority of trees is the random forest's output for classification problems. The mean or average forecast of each individual tree is given back for regression tasks. The tendency of decision trees to overfit to their training set is compensated for by random decision forests. Although random forests are less accurate than gradient enhanced trees, they often perform better than choice trees. However, their performance may be impacted by data peculiarities.

Tin Kam Ho[1] developed the first algorithm for random decision forests in 1995 by using the random subspace technique, which in Ho's definition is a means of putting Eugene Kleinberg's "stochastic discrimination" approach to classification into practice.

Leo Breiman and Adele Cutler created an algorithm extension and filed for a trademark in 2006 for "Random Forests" (owned by Minitab, Inc. as of 2019). The extension builds a set of decision trees with controlled variance by combining Breiman's "bagging" concept with random feature selection, which was initially proposed by Ho[1] and then separately by Amit and Geman[13].

Businesses often employ random forests as "blackbox" models since they need minimal setup and provide accurate forecasts across a variety of inputs.

SVM

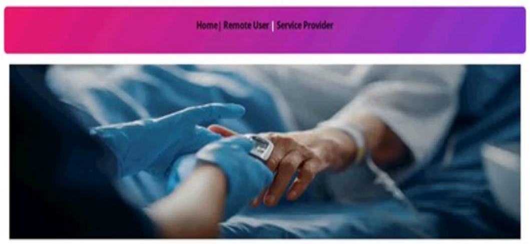
The goal of a discriminant machine learning approach in classification problems is to identify a discriminant function that can accurately predict labels for newly acquired instances based on an independent and identically distributed (iid) training dataset. A discriminant classification function takes a data point x and assigns it to one of the several classes that are part of the classification job, in contrast to generative machine learning techniques that call for calculations of conditional probability distributions. Discriminant techniques are less effective than generative approaches, which are mostly used when prediction entails the identification of outliers. However, they need less training data and processing resources, particularly when dealing with a multidimensional feature space and when just posterior probabilities are required. Finding the equation for a multidimensional surface that optimally divides the various classes in the feature space is the geometric equivalent of learning a classifier.

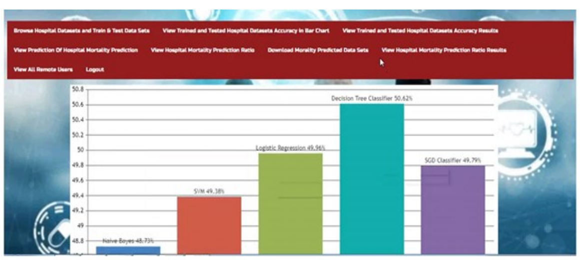
SVM is a discriminant approach that, unlike genetic algorithms (GAs) or perceptrons, which are both often used for classification in machine learning, always returns the same optimum hyperplane value since it solves the convex optimisation issue analytically. The initialisation and termination criteria have a significant impact on the solutions for perceptrons. While the perceptron and GA classifier models are distinct every time training is started, training yields uniquely specified SVM model parameters for a given training set for a certain kernel that converts

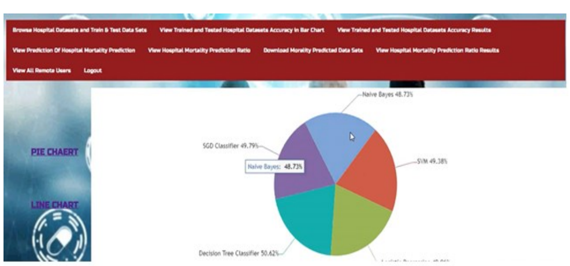
ISSN 2347-3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686

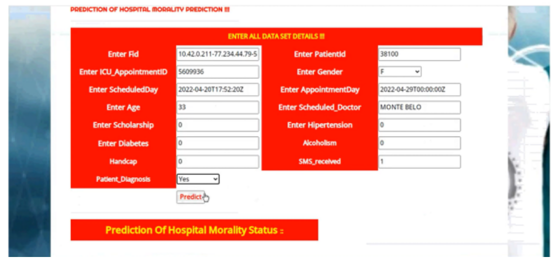
the data from the input space to the feature space. The only goal of GAs and perceptrons is to reduce training error, which will result in several hyperplanes satisfying this criterion.

V. SCREEN SHOTS









VI. CONCLUSION

The goal of this project is to allow FL to create highly customised machine learning models for every participant in order to address the decline in predicting performance in a real multi-center intensive care unit situation. It is more important for real-world healthcare applications since it preserves the separate ICU centres' natural, full non-IID and imbalanced data distribution. In order to determine the cause of the baseline FL's performance decline, we first examined its features in this data setting. Then, in order to help FL recover from non-IID and imbalanced data, we suggested POLA, a one-shot and two-step customised technique. By creating a one-shot adaption for FL, POLA creates a customised local model for every autonomous ICU centre, rebalancing local data knowledge with global expertise. Through the creation of highly customised and high-performing models, we empirically show that it not only enhances FL performance but also drastically lowers the number of FL training communication cycles.

REFERENCES

ISSN 2347–3657

Volume 13, Issue 2, 2025

https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686

[1] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, ''Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis,'' *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1589–1604, Sep. 2018, doi: 10.1109/JBHI.2017.2767063.

[2] J. Wu, J. Roy, and W. F. Stewart, ''Prediction modeling using HER data: Challenges, strategies, and a comparison of machine learning approaches,'' *Med. Care*, vol. 48, no. 6, pp. S106–S113, Jun. 2010, doi:10.1097/MLR.0b013e3181de9e17.

[3] W. G. van Panhuis, P. Paul, C. Emerson, J. Grefenstette, R. Wilder, A. J. Herbst, D. Heymann, and D. S. Burke, ''A systematic review of barriers to data sharing in public health,'' *BMC Public Health*, vol. 14, no. 1, pp. 1–9, Dec. 2014, doi: 10.1186/1471-2458-14-1144.

[4] A. Gkoulalas-Divanis, G. Loukides, and J. Sun, ''Publishing data from electronic health records while preserving privacy: A survey of algorithms,'' *J. Biomed. Informat.*, vol. 50, pp. 4–19, Aug. 2014, doi: 10.1016/j.jbi.2014.06.002.

[5] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, ''Federated learning: Strategies for improving communication efficiency,'' 2016, *arXiv:1610.05492*.

[6] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Y. A. Arcas, ''Communication-efficient learning of deep networks from decentralized data,'' in *Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2017,pp. 1273–1282.

[7] 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 Digit. Med.*, vol. 3, no. 1, pp. 1–7, 2020, doi: 10.1038/s41746-020-00323-1.

[8] P. Kairouz. (Dec. 2019). *Advances and Open Problems in Federated Learning*. [Online]. Available: https://hal.inria.fr/hal-02406503

[9] J. Lee, J. Sun, F. Wang, S. Wang, C.-H. Jun, and X. Jiang, ''Privacypreserving patient similarity learning in a federated environment: Development and analysis,'' *JMIR Med. Informat.*, vol. 6, no. 2, p. e20, Apr. 2018, doi: 10.2196/medinform.7744.

[10] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, ''Federated learning of predictive models from federated electronic health records,'' *Int. J. Med. Informat.*, vol. 112, pp. 59–67, Apr. 2018, doi: 10.1016/j.ijmedinf.2018.01.007.

[11] L. Huang, A. L. Shea, H. Qian, A. Masurkar, H. Deng, and D. Liu, ''Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records,'' *J. Biomed. Informat.*, vol. 99, Nov. 2019, Art. no. 103291, doi: 10.1016/j.jbi.2019.103291.

[12] A. Vaid et al., ''Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: Machine learning approach,'' *JMIR Med. Informat.*, vol. 9, no. 1, Jan. 2021, Art. no. e24207, doi: 10.2196/24207.

[13] K. Chandiramani, D. Garg, and N. Maheswari, ''Performance analysis of distributed and federated learning models on private data,'' *Proc. Comput. Sci.*, vol. 165, pp. 349–355, 2019, doi: 10.1016/j.procs.2020.01.039.

[14] A. Nilsson, S. Smith, G. Ulm, E. Gustavsson, and M. Jirstrand, ''A performance evaluation of federated learning algorithms,'' in *Proc. 2nd Workshop Distrib. Infrastructures Deep Learn.*, Dec. 2018, pp. 1–8, doi: 10.1145/3286490.3286559.

[15] H. Zhu, J. Xu, S. Liu, and Y. Jin, ''Federated learning on non-IID data: A survey,'' *Neurocomputing*, vol. 465, pp. 371–390, Nov. 2021, doi: 10.1016/j.neucom.2021.07.098.

ISSN 2347–3657

Volume 13, Issue 2, 2025

https://doi.org/10.62647/ijitce.2025.v13.i2.pp676-686

[16] K. Hsieh, A. Phanishayee, O. Mutlu, and P. Gibbons, ''The non-IID data quagmire of decentralized machine learning,'' in *Proc. Int. Conf. Mach. Learn.*, vol. 119, Nov. 2020, pp. 4387–4398. [Online]. Available: http://proceedings.mlr.press/v119/hsieh20a/hsieh20a.pdf

[17] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, ''Federated learning with non-IID data,'' 2018, *arXiv:1806.00582*.

[18] T. Yu, E. Bagdasaryan, and V. Shmatikov, ''Salvaging federated learning by local adaptation,'' 2020, *arXiv:2002.04758*.

[19] D. Ting, H. Hamdan, K. A. Kasmiran, and R. Yaakob, ''Federated learning optimization techniques for non-IID data: A review,'' *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 12, pp. 1315–1329, 2020. [Online]. Available: https://iaeme.com/MasterAdmin/Journal_uploads/IJARET/VOLUME_11_ISSUE_12/IJARET_11_12_125.pdf, doi: 10.34218/IJARET.11.12.2020.125.

[20] V. Kulkarni, M. Kulkarni, and A. Pant, ''Survey of personalization techniques for federated learning,'' in *Proc. 4th World Conf. Smart Trends Syst., Secur. Sustainability (WorldS)*, Jul. 2020, pp. 794–797, doi: 10.1109/WorldS450073.2020.9210355.

[21] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, ''Towards personalized federated learning,'' *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 18, 2022, doi: 10.1109/TNNLS.2022.3160699.

[22] Q. Wu, K. He, and X. Chen, ''Personalized federated learning for intelligent IoT applications: A cloud-edge based framework,'' *IEEE Open J. Comput. Soc.*, vol. 1, pp. 35–44, Feb. 2020, doi: 10.1109/OJCS.2020.2993259.

[23] S. Zhang, A. Choromanska, and Y. Lecun, ''Deep learning with elastic averaging SGD,'' in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2015, pp. 685–693.

[24] S. K. Pye and H. Yu. (Aug. 2021). *Personalised Federated Learning: A Combinational Approach*. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2021arXiv210809618K/abstract

[25] S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh, ''SCAFFOLD: Stochastic controlled averaging for ondevice federated learning,'' in *Proc. 37th Int. Conf. Mach. Learn.*, 2019, pp. 5132–5143.