

Health Risk Assessment with Federated Learning

Ioanna Diamantoulaki*, Panagiotis D. Diamantoulakis*, Pavlos S. Bouzinis*,
Panagiotis Sarigiannidis†, and George K. Karagiannidis*

*Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

†Department of Informatics and Telecommunications Engineering, University of Western Macedonia, 501 00 Kozani, Greece
e-mails: iddiaman@ece.auth.gr, padiaman@ieee.org, mpouzinis@auth.gr, psarigiannidis@uowm.gr, geokarag@auth.gr

Abstract—The latest growth of storage capabilities has led to an accumulating volume of medical data stored locally by various healthcare entities. Given the recent progress observed in the domain of artificial intelligence, these data could be efficiently exploited, leading to improved and less expensive healthcare conditions. However, the common practice is medical data to be solely locally used, ending up poorly exploited, due to strict sharing restrictions stemming mainly from privacy limitation of their sensitive nature. Considering the centralized character of conventional machine learning approaches, it is apparent that they cannot reassure the privacy required. On the other hand, federated learning (FL) can be regarded as an upcoming and promising answer to efficient exploitation of medical data, considering its decentralized approach. In more details, FL can enable collaboration among various participating entities on the development and training of a common, central and fully shared model without need of sharing owned sensitive data. Thus, apparently FL approach not only can mitigate the privacy-preservation issues but can lead to the development of reliable and robust healthcare tools. Indicatively, FL can facilitate the development of a model capable of assessing the health risk, which can be a vital tool for medical sciences. To this end, in this work we present a tool capable of assessing the occurrence of different diseases or complications. This tool is based on FL technique utilizing deep neural network model. The FL model developed herein is indicatively applied to four different medical applications proving its generality in the healthcare domain. The FL approach discussed herein is compared with a corresponding centralized learning. According to the demonstrated results, FL can consist a useful health risk assessment tool exhibiting acceptable performance while preserving privacy in sensitive medical data.

Index Terms—Federated Learning, Health Risk Assessment, Mellitus Diabetes, Heart Disease, Maternal Mortality, Breast Cancer.

I. INTRODUCTION

It is widely recognised that health systems must put more emphasis on prevention and adopt a person-centred approach. For this purpose, the utilization of machine learning (ML), linked to a patients' database is capable of elaborating risk assessment models with improved precision. The ultimate goal of ML is to facilitate the development learning algorithms that do the learning automatically from the available data with minimal or even none human intervention [1]. In the existing literature, plenty of centralized ML techniques have been developed with the aim to contribute to medical diagnosis [2]–[5]. However, lately several privacy issues have aroused combined with the need of dealing with huge amount of data. Those two factors gradually led to the development of decentralized ML approaches, such as federated learning (FL) [6]–[11].

The main advantage of FL is that different clients do not share raw data with the server or any of the of the residual participants, since the training dataset is retained in the source of generation, e.g., specific hospital. More specifically, each client performs the model training through its local dataset individually, and forwards only the training parameters to the central server, instead of sending the overall raw data. In this manner, the central unit has no explicit access to privacy-sensitive data. Following that, the server is aggregating the received parameters, updates the global model and finally sends it to the learners, while the considered process is repeated until convergence of the global model. Therefore, the privacy-preserving mechanism constitutes an inherent characteristic of FL. The use of federated learning (FL) for medical applications has been investigated among others in [12], [13], [14]. In more details, [12] has investigated the advantages and impact of FL regarding medical applications while highlighting relevant technical considerations as well. In [13] current and emerging methods for federated as well as secure and privacy-preserving artificial intelligence are discussed with focus on medical imaging applications. Finally, in [14] the concept of clustered FL is used in order to achieve automatic diagnosis of COVID-19.

Motivated by the above, we propose the application of a robust, privacy-preserving FL based tool to disease risk assessment. To this end, the local training is based on deep neural networks (DNN), which is implemented with the capability to provide binary health risk assessment. The latter can be used as a quantitative indicator for the appearance of a specific disease. This indicator could be particularly useful for identifying a potential health problem at an early stage, assisting clinicians and medical professionals, as well as encouraging citizens to be better informed about their health, reducing avoidance of medical examinations. Thus, the developed tool can automatically provide a provisional risk assessment which can then be followed-up with further detailed and established methods and treated promptly in order to minimize their likelihood of progression to higher levels of severity. It is highlighted that the decentralized learning approach used in FL is compatible with the privacy-preservation restrictions related to the sharing of medical data. To verify the reliability and generality of the developed tool, four datasets are used, which correspond to four different health problems, namely diabetes mellitus, heart failure, maternal mortality, and breast cancer. According to results generated in this work, the developed tool exhibits a

reliable performance for all the considered health problems, as well as minor performance deviations compared to a respective centralized learning (CL) model.

II. DEVELOPED FL-BASED TOOL

As it has already been mentioned, FL enables the training of a shared model in a distributed manner, by exploiting the collected data of the multiple clients e.g., hospitals, clinics, etc., without those being intervened by the server. Hence, each client contributes to the construction of the model by performing local training on its own dataset, while the server's role is to aggregate, update, and redistribute the updated model back to the particular client. By using this approach, a model that is able to assess the risk level for different health problems can be trained in a collaborative manner, without the need to exchange sensitive medical data.

According to Fig. 1, we consider a FL network consisting of N clients indexed as $n \in \mathcal{N} = \{1, 2, \dots, N\}$ and a central server. Each client n possesses a local dataset $\mathcal{D}_n = \{\mathbf{x}_{n,k}, r_{n,k}\}_{k=1}^{D_n}$, where $D_n = |\mathcal{D}_n|$ are the data samples, $\mathbf{x}_{n,k}$ is the k -th input data vector of client n representing risk factors and $r_{n,k}$ is the corresponding output which denotes the risk level. We clarify that $\mathbf{x}_{n,k}, \forall n, k$, consists of T risk factors, i.e., the vector $\mathbf{x}_{n,k}$ has T entries. It is also noted that since in this work we focus on binary risk assessment, it holds that $r_{n,k} \in \{0, 1\}, \forall n, k$ where $r_{n,k} = 1$ indicates health risk occurrence and $r_{n,k} = 0$ the opposite case. The whole dataset is denoted as $\mathcal{D} = \bigcup_{n \in \mathcal{N}} \mathcal{D}_n$, while the size of the training data is $D = \sum_{n=1}^N D_n$. Moreover, the local loss function on the data set \mathcal{D}_n , defined as

$$F_n(\mathbf{w}) \triangleq \frac{1}{D_n} \sum_{k \in \mathcal{D}_n} f(\mathbf{w}, \mathbf{x}_{n,k}, r_{n,k}), \quad \forall n \in \mathcal{N}, \quad (1)$$

where $f(\mathbf{w}, \mathbf{x}_{n,k}, r_{n,k})$ captures the error of the model parameter \mathbf{w} for the input-output vector pair $\{\mathbf{x}_{n,k}, r_{n,k}\}$. Since we focus on binary health risk assessment, the function $f(\mathbf{w}, \mathbf{x}_{n,k}, r_{n,k})$ used in our model is the binary cross-entropy, which is highly recommended for binary classification problems [15]. The training process is capable to find the global parameter \mathbf{w} which minimizes the loss function on the whole data set, which is given by $J(\mathbf{w}) = \frac{1}{D} \sum_{n=1}^N D_n F_n(\mathbf{w})$ [11].

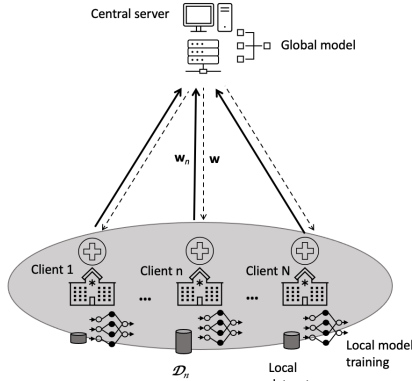


Fig. 1: Reference architecture of medical risk assessment with FL.

The whole training process is divided in an arbitrary number of communication rounds, denoted by i . Thus, during the first round the server initializes \mathbf{w}^0 , while the i -th round is described by the following steps [16]–[18]:

- 1) The central server reports the global parameter \mathbf{w}^i to all clients during the considered round.
- 2) After receiving the global model parameter, each client $n \in \mathcal{N}$, train its local model by applying a few steps of the stochastic gradient descent (SGD) method using appropriately tuned value of learning rate. SGD method has been chosen since it has been found that models trained by it appear to have vanishing generalization errors with few iterations [19]. However, it needs to be mentioned that regarding local training, alternative gradient descent-based methods can be used as well [17].
- 3) After receiving all the local parameters, the server aggregates them, in order to update the global model parameter, by applying

$$\mathbf{w}^{i+1} = \frac{1}{D} \sum_{n \in \mathcal{N}} D_n \mathbf{w}_n^{i+1}. \quad (2)$$

As far as the training of local model is concerned, we use classic feedforward fully connected DNN of architecture appropriately tuned for each dataset utilized. More specifically, the input layer consist of neurons equal to the number of risk factors included in the dataset examined. The number of both hidden layers and respective neurons vary for each examined dataset in order to accommodate the different size of input layer introduced. Regarding the output layer, it consists of one neuron, which is activated via sigmoid function, given the fact that here we focus on provisional health risk assessment using binary classification to denote if health problem risk occurs or not.

III. DESCRIPTIONS OF CASE STUDIES

By focusing on binary health risk assessment, we assess the occurrence risk level for four different cases, namely diabetes mellitus, typical heart disease, maternal mortality, and breast cancer. Statistics of datasets utilized are presented in Table I, where ‘‘C’’ and ‘‘N’’ denote categorical and numerical type of risk factor. Here, the DNN architecture is described by vector $\mathbf{l} = l_1, \dots, l_n$, where each of l_k , for $k = 1, \dots, n$ represents the neurons number for k layer. For D1, D2, D3 and D4, \mathbf{l} is given by (6,4,2,1), (10,8,4,2,1), (5,3,2,1) and (9,4,2,1) respectively.

A. Dataset D1 - Diabetes Mellitus

Dataset D1 is related to type 2 Diabetes Mellitus, which is a metabolic disorder associated with chronic hyperglycaemia leading to complications of various organs [20]. Thus, diabetes represents a major public health concern. D1, available at Kaggle [21], is originally from the National Institute of Diabetes and Digestive and Kidney Diseases [22]. It consists of $D=768$ sample instances, each of them corresponding to female patients of equal number. Aforementioned dataset includes eight risk factors, namely ‘‘Pregnancies’’ for total number of pregnancies, ‘‘Glucose’’ for plasma glucose concentration, ‘‘DiabetesPedigreeFunction’’ for diabetes pedigree function, ‘‘SkinThickness’’ for triceps skin fold thickness in

TABLE I: Risk factors

Risk factor	Type	Mean	Median	Dispersion	Min.	Max.
Pregnancies	C	-	1	2.41	-	-
Glucose	N	120.89	117.0	0.26	0.0	199.0
Blood Pressure	N	69.11	72.0	0.28	0.0	122.0
Skin Thickness	N	20.54	23.0	0.78	0.0	99.0
Insulin	N	79.80	30.50	1.44	0.0	846.0
BMI	N	31.993	32.0	0.246	0.0	67.1
Diabetes Pedigree Function	N	0.472	0.373	0.702	0.078	2.420
Age	N	33.24	29.0	0.35	21.0	81.0

(a) Diabetes Mellitus (D1)

Risk factor	Type	Mean	Median	Dispersion	Min.	Max.
Cholesterol	N	198.80	223	0.55	0	603
MaxHR	N	136.81	138	0.19	60	202
RestingBP	N	132.40	130	0.14	0	200
Age	N	53.51	54	0.18	28	77
Oldpeak	N	0.887	0.6	1.201	-2.6	6.2
ST _{slope}	C	-	2	0.893	-	-
RestingECG	C	-	0	0.949	-	-
Gender	C	-	1	0.514	-	-
FastingBS	C	-	0	0.543	-	-
Exercise Angina	C	-	0	0.675	-	-
ChestPain Type	C	-	3	1.13	-	-

(b) Heart Failure Dataset (D2)

Risk factor	Type	Mean	Median	Dispersion	Min.	Max.
Age	N	29.87	26.0	0.45	10.0	70.0
SystolicBP	N	113.20	120.0	0.16	70.0	160.0
DiastolicBP	N	76.46	80.0	0.18	49.0	100.0
BS	N	8.726	7.5	0.377	6.0	19.0
BodyTemp	N	98.665	98.0	0.014	98.0	103.0
Heart Rate	N	74.3	76.0	0.11	7.0	90.0

(c) Maternal Mortality (D3)

Risk factor	Type	Mean	Median	Dispersion	Min.	Max.
Clump Thickness	N	3.961	3.63	0.714	0.0	10.0
Unif. Cell Size	N	2.652	0.91	1.163	0.0	9.99
Unif. Cell Shape	N	2.722	0.99	1.106	0.0	10.0
Marginal Adhesion	N	2.320	0.85	1.245	0.0	9.98
Single Cell Size	N	2.739	1.79	0.815	0.02	9.96
Bare Nuclei	N	3.063	0.86	1.194	0.0	9.99
Bland Chromatine	N	2.942	2.2	0.831	0.01	9.97
Normal Nucleoli	N	2.371	0.80	1.29	0.0	10.0
Mitoses	N	1.115	0.60	1.587	0.0	9.92

(d) Breast Cancer (D4)

mm, “BloodPressure” for diastolic blood pressure in mmHg, “Insulin” for 2-h serum insulin in mU/mL, “BMI” for Body Mass Index in kg/m² and “Age” for age in years and corresponding risk label value, equal to 0 if the occurrence of diabetes is of high risk and 1 for the opposite case. All risk factors, which are either of numerical or categorical type (Table I-a) are discretized using the Entropy-Minimum

Description Length (MDL) discretization method. It needs to be mentioned that, according to Fig. 2a, only six risk factors ($T=6$) are finally considered to the risk assessment, since blood pressure and skin thickness have been found to be rarely dependent on the risk level according to the importance of features, in terms of risk factors, using ExtraTreesClassifier as proposed in [23].

B. Dataset D2 - Heart Failure

Heart failure consists a significant public health problem related to increased mortality, morbidity and healthcare expenditures [24]. Here, we use dataset D2 downloaded from [25]. In more details, D2 consists of $D=918$ sample instances with eleven risk factors and a risk level value equal to 1 or 0, indicating the occurrence or no occurrence of heart failure risk, respectively. The consisting risk factors are: “Age” in years, “Gender”, “RestingBP” for resting blood pressure in mmHg, “Cholesterol” for serum cholesterol in mm/dl, “FastingBS” for fasting blood sugar, “RestingECG” for resting electrocardiogram results, “MaxHR” for maximum heart rate achieved, “ExerciseAngina” for exerciseinduced angina, “Oldpeak” for old peak measured in depression, “ST_{slope}” for the slope of the peak exercise ST segment. Risk factors shown in Table I-b are discretized using the Entropy-MDL discretization method. As it is shown in Fig. 2b, it needs to be mentioned that only ten risk factors ($T=10$) are considered to the risk assessment since “RestingBP” has been found to be rarely dependent on the risk level according to the importance of features, in terms of risk factors, using ExtraTreesClassifier.

C. Dataset D3 - Maternal Mortality

Maternal mortality discussed here is associated with complications of pregnancy and childbirth. Corresponding dataset is D3, which has been downloaded from [26] and has been collected from different hospitals, community clinics, maternal health cares from the rural areas of Bangladesh [27] and [28]. D3 consists of $D=1014$ sample instances including six risk factors and a risk level value equal to 0 and 1 corresponding to the cases of existence or no existence of high risk for maternal mortality, respectively. Included risk factors are: “Age” for the patient’s age during pregnancy in years, “SystolicBP” for the upper value of blood pressure in mmHg, “DiastolicBP” for the lower value of blood pressure in mmHg, “BS” for the blood glucose levels in terms of a molar concentration in mmol/L, “BodyTemp” for body temperature and “Heart Rate” for normal resting heart rate in beats/min. Risk factors depicted in Table I-c are discretized using the Entropy-MDL discretization method. Also, as it is shown in Fig. 2c, only five risk factors ($T=5$) are finally considered to the risk assessment, since Body Temperature has been found to be rarely dependent on the risk level according to the importance of features, in terms of risk factors, using ExtraTreesClassifier.

D. Dataset D4 - Breast Cancer

Dataset D4, consisting of $D=683$ instances, has been downloaded from [29]. Risk factors, included in D4, are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei

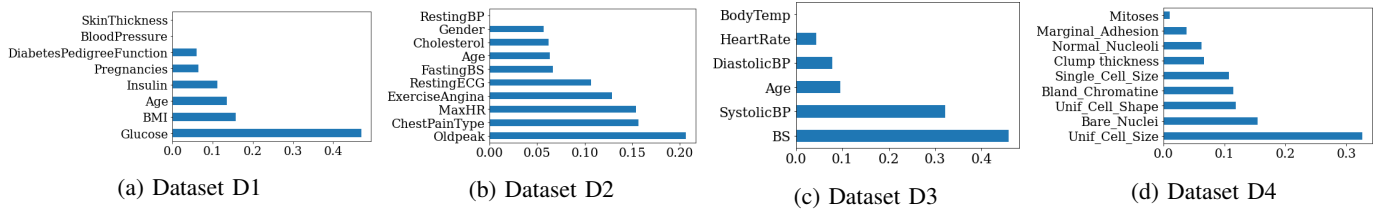


Fig. 2: Feature importance for the considered diseases

TABLE II: Federated learning (FL)/centralized learning(CL) - Metrics for all datasets

Dataset	Binary Accuracy $\left(\frac{TP+TN}{TP+TN+FP+FN}\right)$	Precision $\left(\frac{TP}{TP+FP}\right)$	Recall $\left(\frac{TP}{TP+FN}\right)$	AUC	Specificity $\left(\frac{TN}{TN+FP}\right)$	F1 $\left(\frac{TP}{TP+\frac{1}{2}(FP+FN)}\right)$
Diabetes mellitus (D1)	0.791/ 0.799	0.695 /0.694	0.746/ 0.782	0.837 /0.836	0.808/0.808	0.719/ 0.735
Typical heart disease (D2)	0.832 /0.788	0.904/ 0.925	0.794 /0.692	0.926/ 0.927	0.883/0.883	0.846 /0.791
Maternal mortality (D3)	0.925 /0.902	0.960 /0.915	0.942/ 0.962	0.927 /0.925	0.870 /0.702	0.951 /0.938
Breast cancer (D4)	0.963/ 0.978	0.965/ 0.966	0.948/ 0.983	0.998/0.998	0.974/ 0.975	0.957/ 0.974

present in the image [30].

Dataset D4 includes nine risk factors (“Clump thickness” for clump thickness, “Unif Cell Size” for uniformity of cell size, “Unif Cell Shape” for uniformity of cell shape, “Marginal Adhesion” for marginal adhesion, “Single Cell Size” for single epithelial cell size, “Bare Nuclei” for bare nuclei, “Bland ChromaLine” for bland chromaline, “Normal Nucleoli” for normal nucleoli and “Mitoses” for mitoses, Table I-d). Entropy-MDL discretization method is applied in order to accommodate appropriately the amount of the categorical values for all risk factors, which end up being nine ($T=9$), as it is shown in Fig. 2d. Risk level value equals 0 or 1, indicating the case of benign and malignant respectively.

IV. PERFORMANCE EVALUATION AND DISCUSSION

The purpose of this work is to provide a reliable health risk assessment tool capable of successful medical provisional diagnosis irrespectively of disease or complication type without violating privacy issues corresponding either to patients or medical institutions. The number of clients vary for each examined case ($N=3, 2, 5$ and 8 for D1, D2, D3 and D4 respectively), proving our tool is adaptable to varying clients number, and it is assumed that the total number of data samples D is equally distributed to participating clients. In all datasets examined, 80% of data samples are used for training keeping the rest of total ones for testing. The efficiency of our tool is measured in terms of appropriate statistic metrics for medical applications. It is highlighted that the accuracy may not be a reliable measure in the case of imbalanced datasets [31], while an indicative example is depicted in Fig. 3. Thus, more delicate measures are also utilized, namely Precision, Recall, Specificity, F1 (Table II) and AUC, where AUC is the area under the Receiver Operating Characteristics (ROC) curve, TP, TN, FP and FN correspond to True Positives, True Negatives, False Positives and False Negatives, respectively.

At first, our FL based risk assessment tool is compared with the conventional centralized learning approach. According to the results derived herein, the proposed FL risk assessment tool is capable of achieving high efficiency medical prognosis, since all the examined metrics exhibit, in general, minor

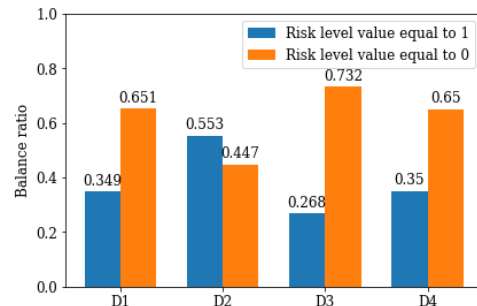


Fig. 3: Balance ratio for examined datasets

modifications, compared to the ones provided by the CL corresponding approach (Table II). Efficiency levels have been found to strongly depend upon the quality of dataset, in terms of number of different instances, amount and relevance of risk factors, etc. Although all examined datasets succeed in exhibiting high levels of Binary Accuracy (at least over 0.791 according to Table II), we can not rely exclusively on this metric since it is not a reliable one in the case of not well-balanced datasets, alike the ones used here (Fig. 3). Our FL tool exhibits a high level distinguishment between binary risk level 0 and 1, since minimum AUC value observed equals 0.837 (Table II). In the case of medical applications, Precision is considered as an important metric, given the fact that it is important a patient who is negative not to be faulty diagnosed as positive and undergo through unnecessary and often expensive further medical tests. In the aforesaid table, it can be observed that our FL tool provides excellent levels of precision for the D2, D3 and D4 (precision has been found to exceed 0.9) and an acceptable level for D1 (0.70 approximately). For medical application, it is also crucial to achieve high level of Recall since it is unacceptable a patient who is positive to be classified as negative. Thus, it is promising that excellent Recall values are observed by our FL tool for D3 and D4 (above 0.94) and satisfactory ones for D1 and D2. F1 values observed are indicative of the satisfactory balance we achieve between Precision and Recall. Specificity of the datasets, which is in general associated with successfully diagnosing healthy patients as negatives, also exhibits high levels for all of investigated datasets.

V. CONCLUSIONS

In this work, we investigated the application of a FL based model to health risk assessment irrespectively to disease or complication type. The developed tool is capable of providing an automatic provisional health risk assessment which can then be followed-up with further elaborated methods and treatments. By using four datasets that correspond to different health problems, it was shown that the developed tool exhibits reliable performance in terms of observed metrics values whereas it has a satisfactory generality potential in the healthcare. Although the main motivation for using FL in healthcare is the protection of privacy-sensitive data, it has been found that our FL model achieves similar performance to a respective CL model. Based on the aforementioned observations, the prediction of future health risk assessment would be of high importance to be investigated. Also, it deserves to be noted that although FL can help to reduce the transmission of sensitive data to a cloud server, it has been shown that exchanging model updated during the training process can still reveal sensitive information. To this end, a privacy-by-design federated learning mechanism for health risk assessment needs to be developed.

VI. ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant 957406.

REFERENCES

- [1] A. C. Dimopoulos, M. Nikolaidou, F. F. Caballero, W. Engchuan, A. Sanchez-Niubo, H. Arndt, J. L. Ayuso-Mateos, J. M. Haro, S. Chat-terji, E. N. Georgousopoulou *et al.*, "Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk," *BMC medical research methodology*, vol. 18, no. 1, pp. 1–11, 2018.
- [2] I. Kononenko, I. Bratko, and M. Kukar, "Application of machine learning to medical diagnosis," *Machine learning and data mining: Methods and applications*, vol. 389, p. 408, 1997.
- [3] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in medicine*, vol. 23, no. 1, pp. 89–109, 2001.
- [4] K. R. Foster, R. Koprowski, and J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research-commentary," *Biomedical engineering online*, vol. 13, no. 1, pp. 1–9, 2014.
- [5] J. G. Richens, C. M. Lee, and S. Johri, "Improving the accuracy of medical diagnosis with causal machine learning," *Nature communications*, vol. 11, no. 1, pp. 1–9, 2020.
- [6] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.
- [7] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečný, S. Mazzocchi, B. McMahan, T. Van Overveldt, D. Petrou, D. Ramage, and J. Roselander, "Towards federated learning at scale: System design," in *Proc. Machine Learning and Systems*, A. Talwalkar, V. Smith, and M. Zaharia, Eds., vol. 1, 2019, pp. 374–388.
- [8] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020.
- [9] M. Aledhari, R. Razzak, R. M. Parizi, and F. Saeed, "Federated learning: A survey on enabling technologies, protocols, and applications," *IEEE Access*, vol. 8, pp. 140 699–140 725, 2020.
- [10] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.
- [11] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of fedavg on non-IID data," *arXiv preprint arXiv:1907.02189*, 2019.
- [12] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein *et al.*, "The future of digital health with federated learning," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–7, 2020.
- [13] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.
- [14] A. Qayyum, K. Ahmad, M. A. Ahsan, A. Al-Fuqaha, and J. Qadir, "Collaborative federated learning for healthcare: Multi-modal covid-19 diagnosis at the edge," *arXiv preprint arXiv:2101.07511*, 2021.
- [15] C. A. Kulkarni, M. R., and S. S. Sharath, "Custom binary cross entropy based anomaly detection in bank transactions using deep convolutional neural network," in *iiWAS2021: The 23rd International Conference on Information Integration and Web Intelligence, Linz, Austria, 29 November 2021 - 1 December 2021*, E. Pardede, M. Indrawan-Santiago, P. D. Haghghi, M. Steinbauer, I. Khalil, and G. Kotsis, Eds. ACM, 2021, pp. 319–323. [Online]. Available: <https://doi.org/10.1145/3487664.3487708>
- [16] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.
- [17] P. S. Bouzinis, P. D. Diamantoulakis, and G. K. Karagiannidis, "Wireless federated learning (WFL) for 6G networks-Part I: Research challenges and future trends," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 3–7, 2022.
- [18] —, "Wireless federated learning (WFL) for 6G networks—Part II: The compute-then-transmit NOMA paradigm," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 8–12, 2022.
- [19] M. Hardt, B. Recht, and Y. Singer, "Train faster, generalize better: Stability of stochastic gradient descent," in *Proc. International conference on machine learning*. PMLR, 2016, pp. 1225–1234.
- [20] K. G. M. M. Alberti and P. Z. Zimmet, "Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. provisional report of a who consultation," *Diabetic medicine*, vol. 15, no. 7, pp. 539–553, 1998.
- [21] "Pima indians diabetes database," <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>.
- [22] J. W. Smith, J. E. Everhart, W. Dickson, W. C. Knowler, and R. S. Johannes, "Using the adap learning algorithm to forecast the onset of diabetes mellitus," in *Proc. of the annual symposium on computer application in medical care*. American Medical Informatics Association, 1988, p. 261.
- [23] J. Fernandes, N. Simoes, D. Vaz, S. Tiwari, A. Naik, and D. R. Edla, "Prediction of malignant lung nodules in CT scan images using cnn and feature selection algorithms," in *Proc. International Conference on Applied Artificial Intelligence and Computing (ICAIC)*, 2022, pp. 218–224.
- [24] V. L. Roger, "Epidemiology of heart failure," *Circulation research*, vol. 113, no. 6, pp. 646–659, 2013.
- [25] "Heart failure prediction dataset," <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>.
- [26] "Maternal health risk data set data set," <https://archive.ics.uci.edu/ml/datasets/Maternal+Health+Risk+Data+Set>.
- [27] M. Ahmed, M. A. Kashem, M. Rahman, and S. Khatun, "Review and analysis of risk factor of maternal health in remote area using the internet of things (iot)," *InECCE2019*, pp. 357–365, 2020.
- [28] M. Ahmed and M. A. Kashem, "Iot based risk level prediction model for maternal health care in the context of Bangladesh," in *Proc. 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*. IEEE, 2020, pp. 1–6.
- [29] "Breast cancer wisconsin (original) data set," [https://archive.ics.uci.edu/ml/datasets/breast-cancer-wisconsin+\(original\)](https://archive.ics.uci.edu/ml/datasets/breast-cancer-wisconsin+(original)).
- [30] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," *Proceedings of the national academy of sciences*, vol. 87, no. 23, pp. 9193–9196, 1990.
- [31] J. Tohka and M. Van Gils, "Evaluation of machine learning algorithms for health and wellness applications: A tutorial," *Computers in Biology and Medicine*, vol. 132, p. 104324, 2021.