Distributed collaborative learning for Ebola diagnosis: a real world benchmark

Idriss C. T. Tondji
African Masters in Machine Intelligence
Mbour, Senegal
itondji@aimsammi.org

Mary-Anne Hartley
EPFL
Lausanne, Switzerland
mary-anne.hartley@epfl.ch

Martin Jaggi
EPFL
Lausanne, Switzerland
martin.jaggi@epfl.ch

Abstract

BACKGROUND. The ethical and practical limitations of data sharing in the medical domain create barriers to collaborative model building that can be insurmountable in health emergencies and especially in low-resource settings. The COVID and Ebola epidemics are such examples, where insufficient data sharing created catastrophic delays in the public health response. Distributed learning techniques enable several users to collaboratively learn an algorithm without needing to directly expose personal data. Here, instead of sharing original data, only the model updates are communicated.

Motivated by such applications, we developed an open-access, open-source DIStributed COllaborative learning platform: DISCO, offering public access to federated and decentralised learning. Because not visualising the data can allow biases in the data to be learned undetected, DISCO uses iFedAvg for interpretable interoperability guarantees, that enables visualisation of biases in federated learning
AIM. In this project, we benchmark DISCO and iFedAvg on a real world use case: Building diagnostic and prognostic models on Ebola data extracted from several imperfectly interoperable sources.
METHODS/FINDINGS. We curate a previously unstudied collection of real-world tabular data sets from the 2014—2016 West African Ebola epidemic, jointly forming the largest such data set in the world. The composite data comprises the demographics and clinical information on 10019 patients treated or screened at various Ebola Treatment Centres across West Africa. We create models to predict their diagnosis (49% Ebola-positive by PCR) and outcome (53% fatal). We compare traditional 'central' model learning (on a merged data set) with federated learning on DISCO and carry out several experiments to explore the limitations of iFedAvg in detecting inter-user bias.
CONCLUSION. We find that DISCO models match the performance of their centralised counterparts and that several biases can be detected in DISCO using iFedAvg.

1 Background

Data sharing in some applications such as medicine is restricted in order to uphold the ethical and privacy rights of patients and intellectual property of data owners. Additionally, data sharing faces important practical limitations concerning interoperability, where harmonization practices are laborious, unstandardised and often unobtainable in lower resource settings or in the context of health emergencies (1). The COVID and Ebola epidemics are such examples, where insufficient data sharing created catastrophic delays in the public health response. In the absence of systematic data sharing, users are forced to work within silos, thus fragmenting the statistical power of resulting predictive models and over-representing data from higher-resource sites.

In an effort to assist this process for important data sets from low-resource settings, the Infectious Disease Data Observatory (IDDO) (2), have centralised the coordination efforts to acquire ethical approvals, harmonise, merge and store centralised data sets from neglected populations. The creation of the platform was motivated by the 2014—16 West African Ebola epidemic, where they collected and aligned the fragmented data sets into the largest central repository of Ebola data in the world to facilitate further research. The process was laborious and took several years to finalise (released in 2019). Inspired by such use cases, researchers have developed distributed learning techniques that enable several users to collaboratively learn an algorithm without needing to directly expose personal data. Here, instead of sharing original data, only the model updates are communicated.

Our group developed an open-access, open-source DIStributed COllaborative learning platform: DISCO (3), offering public access to this technology.

Because not visualising the data can allow biases in the individual data sets to be learned undetected, DISCO uses iFedAvg (4) for interpretable interoperability guarantees, that enables visualisation of biases in federated learning. While experimental benchmarks exist, there are not many real world examples of the potential of this technology. Real world examples are essential to build trust in the technology to support adoption.

In this project, we benchmark DISCO and iFedAvg on the real world use case of creating diagnostic and prognostic models on Ebola data extracted from several imperfectly interoperable sources.

2 Related work

2.1 Distributed learning

Distributed learning comprises several approaches of communicating and aggregating model updates between users. It can be divided into 2 major categories: federated (FL) and decentralised learning (DeL) which differ according to where the model updates are coordinated (on a central server for FL, vs by the users themselves in DeL)(5). In FL, there is an aggregation server that coordinates the training iterations and collects, aggregates and distributes the models to and from the training Nodes based on predefined criteria. A disadvantage of centralised FL is the dependence on a central server where failure would disrupt the training process. In DeL, there is no central node, each training node is connected to one or more peers and aggregation occurs on each node in parallel (6). The figure 1 illustrates the two network architectures.

DISCO provides both modalities, but for the purposes of this thesis, only FL is used.

While these techniques may seem to resolve all the issues of data sharing, they still have vulnerabilities to privacy concerns (e.g. reidentification attacks) (7) and are under particular threat from learning biases in unseen data (e.g. data poisoning) (8; 9).

2.2 Interoperability

Interoperability has been defined as the ability of different systems and applications to communicate, exchange data, and use the information that has been exchanged (10).

In reality, data is heterogenous and imperfectly interoperable. Multiple extensions have been developed in order to address this issue (11) but mostly focus on how to detect and ignore the "poisoned" data rather than interpreting the interoperability bias for potential correction. Interoperability is fundamental to ensure unbiased interpretation of the features. It is very challenging to harmonise

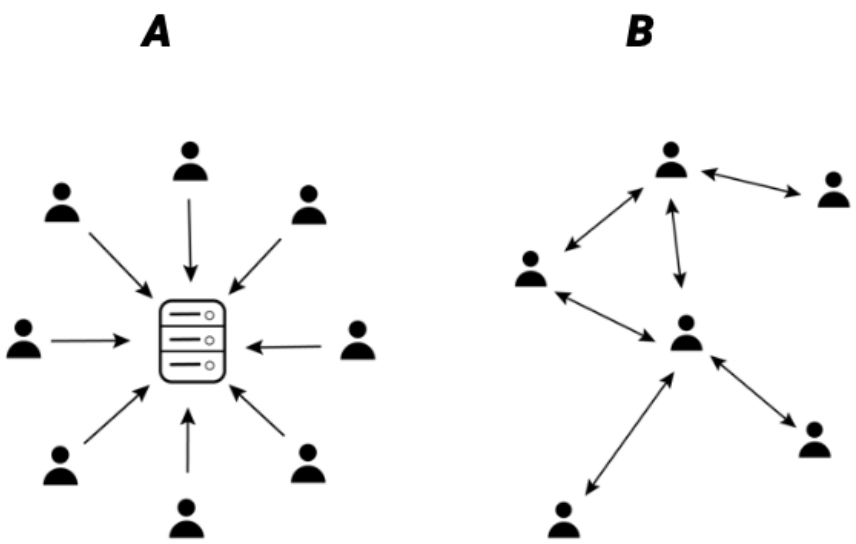


Figure 1: **Architectures of (A) Federated (FL) and (B) Decentralised learning**.

the sources of data in an FL set up as the data is unseen and thus cannot be directly examined to detect inconsistencies, misalignments, or skewed distributions. Such issues could render collaborative learning not only useless, but harmful. In an effort to learn and adapt models to such biases, the iFedAvg model (Interpretable interoperability for Federated Average) was proposed by our group (4). It extracts interoperability metrics from the weights and biases shared during federated learning via local affine layers before and after the model f_{shared}.

iFedAvg has been incorporated into DISCO. This work builds predictive models on a tabular data set of Ebola virus disease (EVD). It comprises a series of poorly interoperable data sets originating from multiple actors across several countries with heterogeneous data collection practices (1).

2.3 Ebola Virus Disease (EVD)

Ebola virus disease (EVD) is an epidemic viral infection that appeared for the first time in 1976 in Central Africa in multiple isolated epidemics, and very little systemic data was collected on the disease. Thus, when an Ebola virus variant was first detected in Guinea in march 2014 (12), very little was known about the disease and there was a great need to learn collaboratively from the many actors responding to the crisis. This 3-year-long epidemic was marked by chaos and misinformation that was mostly attributed to the lack of a coordinated response and central source of data-driven information. It affected tens of thousands of people in several east African countries, collapsing the healthcare systems (13). The Ebola virus is a rapidly transmissible hemorrhagic virus with a death rate of 90% in low and middle income countries (LMIC) and 30% in high income countries (HIC), showing the importance of data heterogeneity between sites. This inequitable death rate was mostly due to the lack of resources and raised awareness of the need to create predictive tools to better allocate scarce resources to those who need it most.

Research on such data is often focused on making diagnostic and prognostic models on a subset of patients data to allocate limited resources to the most critical patients and improve early case identification. Colubri and al. (14) introduce machine learning techniques to EVD dataset by building a prognosis model using only 106 patients, with an area under the receiver operator characteristic curve of 0.8. Multivariate logistic regression have been applied on 470 patients admitted to five Ebola treatment units in order to build a prognosis model, with Area Under the Curve ranging from 0.7 (15). Hartley et al. (16) proposes a highly predictive prognostic tools, which stratify the risk of EVD mortality

3 Aims and objectives

3.1 Aim

The overall aim of this project is to explore the potential and limitations of collaborative learning on a real world clinical data set, the case of the Ebola virus disease (EVD) data set.

3.2 Objectives

The aim of this project is achieved through the sub-objectives outlined below:

- **Perform a relevant literature review**
  Determine the literary landscape concerning retrospective cohort studies performed on clinical tabular EVD data sets, across different regional healthcare facilities.
- **Data Pre-processing**
  Harmonize and merge data sources into a central data set.
- **Create central models for diagnosis and prognosis Ebola**
  i.e. working from single merged central data set.
- **Create federated versions of the above models**
  i.e. working from data sets fragmented into their original geographic sources
- **Implement and test iFedAvg**
  Generate several common biases in the Ebola data set and test whether they are detected

4 Method

4.1 Study design

This is a retrospective cohort study investigating the utility of distributed learning on real world data generated in the health emergency of the 2014—16 West African Ebola epidemic. We train predictive models for diagnosis and prognostication either centrally (i.e. on a merged data set) or federated (i.e. on fragmented data). We simulate the original data partitions, splitting the data by geography where each of user is the Ebola Treatment Center (ETC) that independently collected the data. We deviate from the real world scenario by harmonising the feature names (i.e. assuming that all ETCs collected data using a single nomenclature, which was not the case in reality). We then try to detect natural biases in the data using iFedAvg and also perform several data poisoning interventions in one of five users (i.e. label flip, feature flip).

4.2 Data poisoning experiment

- **Baseline**
  The data sets were split according to their 13 collection source sites as seen in Table 1.
- **Label flip**
  This scenario will have one of the labels of the users flipped, namely all patients labelled as Ebola positive will be labelled as Ebola negative and vice versa. Here, We expect that the model is going to invert the predicted value.
- **Feature Shift**
  This scenario will have a single feature of one of the users shifted in value.

4.3 data set Description

This work uses a composite data set curated by the Infectious Disease Data Observatory (IDDO) (2) which was established to coordinate the centralisation of fragmented data sets on neglected diseases. Their first major project focused on data collected during the 2014—16 Ebola epidemic which they aggregated into the largest central repository of Ebola data in the world (i.e., the Ebola Data Platform, EDP) to facilitate coordinated research.

The Ebola Data Platform (EDP) comprises tabular clinical data on 10019 anonymized patients treated at 12 Ebola Treatment Centres (ETCs) between 2014 and 2016. The data includes both categorical and continuous features such as demographic details (e.g., age, sex, location), clinical signs and symptoms (e.g., fever, headache), laboratory values (e.g., Ebola qPCR for quantitation of viral load) and outcomes for each patient in terms of diagnosis (EVD+ vs EVD-), and for EVD+ patients, their final disposition (death vs recovery).

We build 2 predictive tasks: Diagnosis and Prognosis. For diagnosis, only patients where an EVD test was performed are included. For prognosis, only patients where the outcome is known and a patient was confirmed EVD positive are considered (6382 patients). Thus, the same ETCs are not represented for both tasks. The reason being that some ETCs did not monitor mortality, or others only treated EVD+ cases. The following tables show the number of samples at each treatment centre.

Table 1: **Diagnosis Ebola data set**

Centres	Number of samples	Positive rate
Donka	2350	37.87
Gueckedou	2260	66
Kailahun	1186	50.13
Bo	852	62.20
Makeni	141	56.81
Port-loko	851	45.74
Foya	539	79.79
Bong	523	31
Kerry-Town	487	95
Kambia	450	21
Magburaka	224	29
Nzerekore	156	57.66

Table 2: **Prognosis Ebola data set**

Centres	Number of samples	Mortality rate
Gueckedou	1366	66.58
Foya	450	66
Donka	748	50.13
Nzerekore	166	62.20
Makeni	176	56.81
Port loko	181	65.74
Kambia	1154	50
Bong	168	50
Kailahun	852	55.63
Bo	440	61.36
Magburaka	418	57.41
Kerry-Town	263	56.84

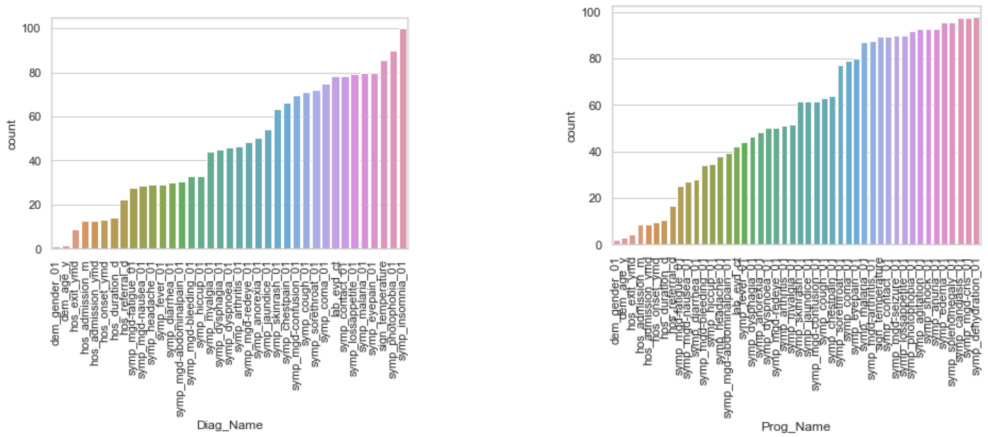
A limitation of the platform on which we run the experiments has led to only a subset of users being able to participate in the federation.

4.4 Data preprocessing

During the epidemic, different ETCs collected non-intersecting subsets of symptoms. This illustrates an important challenge in terms of interoperability. That is, we cannot produce effective predictive models if a data set from one centre has a feature available, while a data set from another centre has the same feature completely missing in the extreme case.

The figure 2 below, demonstrates the count or percentage of missing values for each column-wise feature for each row-wise data set(Diagnosis and Prognosis), where high count denotes large incompleteness.

The main goal in the step is to deal with the incompleteness of data. To do so, we followed two approaches:



(a) Diagnosis data set (b) Prognosis data set.

Figure 2: **Percentage of missing values per feature for each data set.**

- **Missingness mechanism analysis**
  Determine the categories for missing data mechanisms
- **Imputation of incomplete data**
  Replacement of missing values by some specific values

There are several ways that missing values can manifest themselves: missing at random (MAR), missing completely at random (MCAR) and missing not at random (MNAR). We realize that this problem of incompleteness must be carefully investigated, followed by a systematic approach in dealing with it. Variables are missing at random (MAR) mean that the missingness mechanism is not associated (independant) with the target variable and missing not at random (MNAR) means that the missingness mechanism is associated with the target variable.

We set the percentage rate threshold to 70% at which we remove features with missing values.

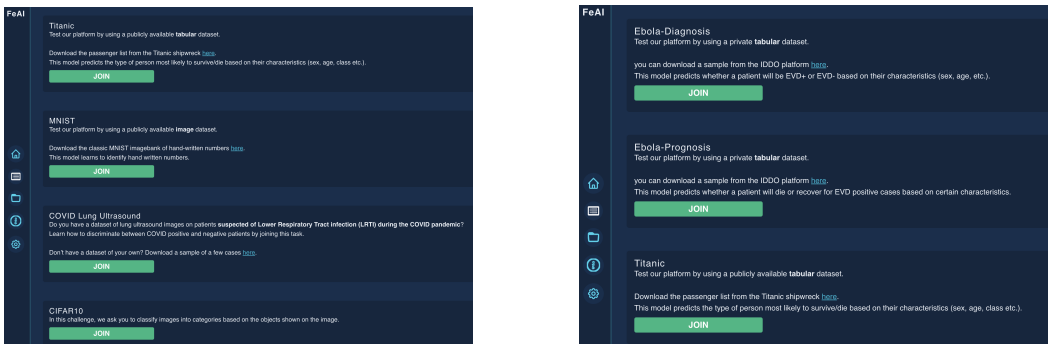
We analyse missingness separately for each task (diagnosis and prognosis) NB: the denominator changes with the reduced cohort of Ebola positive (prognosis task). At the features level, we are analysing variables independently of each others. The principle here is very simple, what we do is the following. For each variable that contains missing values, we create a new encoded variable containing binary values 0 and 1. the value 1 to indicate the row with a missing value and 0 otherwise. The main goal is to study the relationship between the missingness of a particular variable and being Ebola positive. In others words, we are checking if there is an association between having a missing value for a variable and the target or having Ebola. We are going to the logistic regression model with a binary predictor variable to compute the Odds-Ratios and the p-value. The model describes the following linear relationship:

$$logit(p) = \beta_0 + \beta_1 * X \tag{1}$$

Where p is the probability of the target variable y to be 1 or having Ebola, X binary predictor variable and β_0,β_1 the parameter values. The null hypothesis states that there is no relationship between the two variables being studied (one variable does not affect the other) and the alternative hypothesis states that having a missing value for a variable did affect the target variable. The $p - value$ or probability value express the level of statistical significance between 0 and 1. The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.

4.5 Centralised model

During the preprocessing step, All the individual studies on each local data set are inputted one after the other. After that, these local data sets have been then combined into a single centralised data set consisting of all the studies. At the end, we end up with two central data sets, the one for Diagnosis task and another one for the prognosis task.



(a) Before (b) After

Figure 3: **Disco interface before and after the tasks creation.**

We build a neural network model to perform the classification task on each data set. The architecture of the model consists of three hidden layers with a number of units 124, 64 and 32 for each layer respectively. The activation function used is the Relu non-linear function in the hidden layers and the Sigmoid non-linear function in the output layer.

4.6 Interpretable, interoperable distributed learning (iFedAvg)

4.6.1 DISCO

The DISCO platform (3) is a web application that offers infrastructures for Distributed Collaborative Machine Learning. This platform was developed by the Machine Learning Optimization (MLO) Laboratory at EPFL and it offers support for both Decentralised Learning (peer2peer communication) and Federated Learning (uses a central server to share and aggregate weights). It contains several popular tasks such as **Titanic**, **MNIST** or **CIFAR-10**. For this study, we created two new tasks for Ebola diagnosis and prognosis. There are very few well known way to actually take part in collaborative learning without having to build your own system. For that reason, developing platforms such as DISCO is crucial in order to make Decentralised Learning and Federated Learning appealing and accessible to users worldwide. The goal of DISCO is to become an intuitive platform that allows users to take part in Collaborative Learning even without requiring expert knowledge in the Machine Learning field.
We can see the resulting DISCO interface before and after the tasks creation in figure 3.

4.6.2 iFedAvg

In an effort to learn and adapt models to such biases, the iFedAvg model was proposed to extract interoperability metrics from the weights and biases shared during federated learning.
Interpretable Federated Average (iFedAvg) proposed by Rochewitz et al. (4) adds local affine layers before and after the model f_{shared}. This creates a combined model that can be specified as $f_{in} \circ f_{shared} \circ f_{out}$ where $\circ$ indicates a composition.
The personal interoperability input layer f_{in} define a function $f_{in}(x) : R^D \longrightarrow R^D$ such that :

$$f_{in} = (x + b_{in}) \circ W_{in} \tag{2}$$

Where the bias and weight vectors $b_{in}, w_{in} \in R^D$and refers to the element-wise multiplication. We can define the output layer f_{out} in an analogous manner $f_{out}(x) : R^K \longrightarrow R^K$ such that :

$$f_{out} = (x + b_{out}) \circ W_{out} \tag{3}$$

Where the bias and weight vectors b_{out} , $w_{out} \in R^K$.
It is important to note that this method is devised to work on tabular data and in a context of Federated Learning. During training, each user will train its local model $f_{in} \circ f_{shared} \circ f_{out}$ on its own data, then at the end of each round, the users will only send the shared part of their model with the federation, namely f_{shared}.

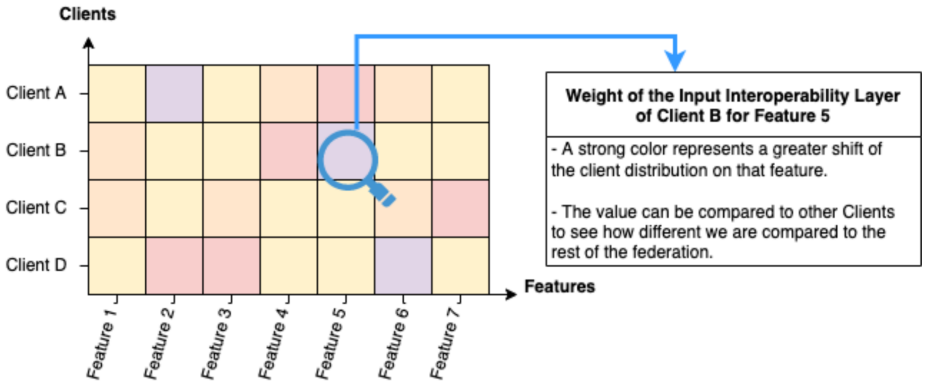


Figure 4: **Heatmap interpretation for Weights of Input Interoperability layers**.

In order to allow for interpretability of the feature-wise Interoperability of the users, the iFedAvg framework of Rochewitz et al. that have been integrate into the DISCO platform generate visualizations of the features distributions in each users. These are in the form of heatmaps in order to display interoperability weights and biases values. This allows the user to easily compare the weights and explain the differences in data distributions.
An example of how the heatmap display looks and how to do the interpretation of each cell can be illustrate in the figure 4.

To be able to read and understand the heatmap, a good start is to know the meaning of each the following terms:

- **Weight** of a feature shows the influence of that feature on the model.
- **Bias** of a feature represents how our local distribution for that feature is different compared to the federation. More the value is large more the shift in the distribution is important and less there is a compatibility between the data.

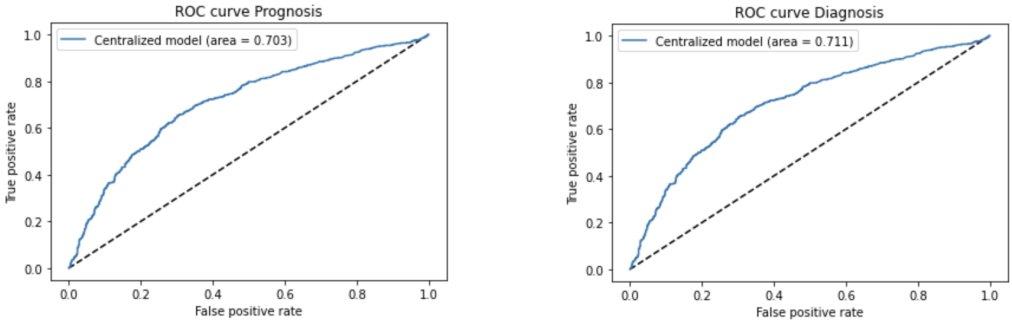
5 Results

5.1 Preprocessing

After applying the missingness threshold, we end up with 23 variables for the Ebola diagnosis data set and 25 variables for the Ebola prognosis data set. The result of the missingness mechanism analysis showed us two different categories for missing data mechanisms into our data set which are MNAR and MAR.
For the Ebola diagnosis data set, we detected only two variables (age and gender) that are MAR and the rest of variables are MNAR ($p < 0.05$). The missingness of gender (OR=0.73,$p = 0.139$) and age (OR=0.78,$p = 0.137$) are not significantly associated with being Ebola positive.
For the Ebola prognosis data set, we were able to detect only two variables (fever and abdominal pain) that are MAR and the rest of 23 variables are missing MNAR. The missingness of having the symptom of fever(OR=1.001,$p = 0.97$) and abdominalpain (OR=1.1,$p = 0.53$) are not significantly associated with being Ebola positive.

We replace the missing value in both data sets by the median in the case of quantitative variables and the mode in the case of qualitative variables.

Importantly, we seek to maintain the real world scenario of pre-processing without ever sharing data. Thus the pre-processing techniques used in this thesis are all processes that could be performed by the data owners without sharing data. For example, the variable names could be aligned by each data user, and highly missing features could also be removed.



(a) Prognosis model (b) Diagnosis model

Figure 5: **ROC Curve for each centralized model on test sanples**.

5.2 Centralised model

ROC is a probability curve and the area under the curve is a scoring of the model's discriminative performance between two classes. The higher the value, the better the model is at distinguishing between classes.
In our experiment, we used 2003 patients and 1276 patients which represent 20% of the diagnosis and prognosis data as test examples. We obtained an area under the receiver operator characteristic curve of 71.1% for the diagnosis task and 70.3% for the prognosis task. The figure 5 illustrate sthe ROC curve for each model on the test samples.

5.3 Federated model

The model defined on DISCO during the creation of each diagnostic and prognostic task is the same as defined on the centralized setting. As we used a neural network model, we maintained the same architecture in order to be able to compare the results. We used the DISCO platform to train the models with the federated setting, we obtained an area under the receiver operator characteristic curve of 69% for the diagnosis task and 68% for the prognosis task.

5.4 iFedAvg

iFedAvg produces heatmaps of weights and biases that can be used to inspect the unseen data sets for interoperability issues and biases.

We simulate several scenarios of extreme biases (label flip and feature shift) as examples. In all of these experiments, we will assume that we are the user with data differing from the federation. The interoperability parameter values of the user out of distribution can thus be read in the line of the heat map labelled "You".

- **Diagnosis data set:**

  - **Baseline**
    This section shows the heatmap resulting from the training with users who have data split by ETC (i.e. as it was in the real world scenario.
    We can see the resulting heatmap in figure 6 We observe relatively even spread out variations in the bias heatmap, reflecting the real world differences in distributions. While the weights heatmap is quite calm. This means that features have almost the influence on the model. The weight values stay really close from their initial value of 1 and don't bring much information. For that reason we decided not to include them in our results.
    We will use this scenario as a baseline to compare with latter scenarios.
    the weight values stay really close from their initial value of 1 and don't bring much information. For that reason we decided not to include them in our results section.
  - **Label flip:**
    This shows the heatmap resulting from the training with one user with its label flipped.
    We can see the resulting heatmap in figure 7. The expected result shows YOU with

Figure 6: **Ebola diagnosis** Heatmap interpretation for Weights and bias with data splitted by ETC. Both input and output layers

systematically different input biases and an opposite trend for output bias. We can see that the interoperability parameters of the clients with its labels flipped contains lot of variations than the other clients. We can also see that the output bias of the modified client is highlighted. in this case the difference between the shifted client and the other is much more noticeable and the output bias values being 0.46 for the modified client and around 0.1 for the other clients.

- **Feature Shift:**
  This section shows the heatmap resulting from the training with one user with one feature shifted. In our case we decided to add 100 to the age variable. We can see the resulting heatmap in figure 8. The weights heatmap is quite calm, which means that features have almost the influence on the model. Here, we want to focus on the age cell as this is the feature we shifted for the modified user. We find out that our expected result is achieved as we can see that the age cell of the modified client is different from the other clients. We observe that the modified client have an bias of -0.18 whereas the other clients have a value of the bias around 0.2. We can thus see that the shifted client is going in a different direction from the other users which is what we expected. Since the features are only directly linked to the input layer, it makes sense to consider using only the input layer.

- **Prognosis data set:**

- **Baseline**
  This section shows the heatmap resulting from the training with users split by ETC.
  We can see the resulting heatmap in figure 9. We also observe relatively even spread out variations in the bias heatmap, reflecting the real world differences in distributions. We can see that features have the influence on the model as the weights heatmap is quite calm.

- **Label flip:**
  This shows the heatmap resulting from the training with one user with its label flipped.
  We can see the resulting heatmap in figure 10.
  If we remove the output layer, the difference between clients is way less significant and we are left with a heatmap similar to the ones splitted by ETC. It makes sense to keep the output layer in cases where we are interested by the targets distributions.

Figure 7: **Ebola diagnosis** Heatmap interpretation for weights and biais with Label flip. Both input and output layers enabled.

Figure 8: **Ebola diagnosis** Heatmap interpretation prognosis for weight and biais with feature shift.

Figure 9: **Ebola Prognosis** Heatmap interpretation prognosis for weight and biais.Spit by ETC.

Figure 10: **Ebola Prognosis** Heatmap interpretation prognosis for weights and biais with Label flipped.

Figure 11: **Ebola Prognosis** Heatmap interpretation prognosis for weight and biais with feature shift

- **Feature Shift:**
  This shows the heatmap resulting from the training with one user with one feature shifted. In our case we decided to add 150 to the duration of hospitalisation variable. Here, we want to focus on the duration cell as this is the feature we shifted for the modified user.
  We can see the resulting heatmap in figure 11. We find out that our expected result is achieved as we can see that the "duration" cell of the modified client is different from the other clients as compared to the baseline where we split by ETC. Let's Take a deeper look into the parameters value. We observe that the modified client have a bias of duration cell of 0.65 whereas the other clients have a bias of the duration cell close to zero. Since the features are only directly linked to the input layer, it makes sense to consider using only the input layer.

6 Discussion

Over the course of this project we managed to create two new tasks on DISCO. These personalized tasks allows us to easily run experiments with the real world data set for diagnosis and prognosis Ebola.

The results presented above show that, in a real scenario, it is indeed possible to observe insights or to identify changes in the distribution of data between users with the help of the heatmap. But, the heatmaps are still noisy and it is not always that easy to know exactly what is happening in the data. So, The iFedAvg framework is not necessarily an perfect tool to exactly know what is happening inside each clients' dataset but rather informs the user that something might be deviant.

For that reason, we don't have to consider the data of a certain user useless based only on the heatmap interpretation of iFedAvg, but rather guide us to where investigate in order to find where that difference comes from and what can be the cause.

The results presented in the previous section show that the use of both input and output layer might affect the interpretability of the data. Even if the framework was presented with both an input and an output layer, it turns out that it might not be always necessary or useful to include both layers in the model construction

7 Conclusion

We benchmark DISCO and an interpretable, interoperable distributed learning framework on a previously unstudied collection of real-world data sets from the 2014 - 2016 West African Ebola epidemic. We find that DISCO models match the performance of their centralised counterparts and that several biases can be detected in DISCO using iFedAvg.

There are several limitations with our project that needs to be addressed. First of all, we have the limited communication between users. It is important to note that only five users have participated to the federation in our study due to some limits or slow connection between users in the platform. Therefore, studying such scenarios with a large number of users would allow more widespread adoption in the future. Also, we are not able to quantify the contribution of each user who takes part in the federation, as the evaluation of the interoperability of the data plays quite an important role in making what happens in the federation clearer for the user.

8 Acknowledgments

Without the guidance and support from all supervisors, this thesis would not have been possible. I want to thank everyone, particularly Nacho and Damien Gengler for their time and commitment. A special recognition for Mary-Anne Hartley has to be expressed; her dedication and enthusiasm has made this work a pleasure.

We would like to thank the DISCO development crew, Ignacio Aleman, Nacho and others for making working on this project a great team effort as it was really motivating to see the platform lively and being regularly updated.

We also thank Moustapha Cisse and Teta Bahunde as well as MLO, iGH and the AMMI program for the opportunity to work on this project as well as the support along the period.

References

[1] M. Brack and T. Castillo, "Data sharing for public health: Key lessons from other sectors," 2015.

[2] I. D. D. O. (IDDO), "Ebola data platform," 2021. [Online]. Available: https://www.iddo.org/research-themes/ebol

[3] MLO-iGH, "Deai : Decentralized collaborative machine learning," 2020. [Online]. Available: https://github.com/epfml/disco

[4] D. Roschewitz, M. Hartley, L. Corinzia, and M. Jaggi, "Ifedavg: Interpretable data-interoperability for federated learning," *CoRR*, vol. abs/2107.06580, 2021. [Online]. Available: https://arxiv.org/abs/2107.06580

[5] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," 2016. [Online]. Available: https://arxiv.org/abs/1602.05629

[6] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *CoRR*, vol. abs/1902.04885, 2019. [Online]. Available: http://arxiv.org/abs/1902.04885

[7] S. W. F. G. Nguyen Truong, Kai Sun and Y. K. Guo, "Privacy preservation in federated learning: An insightful survey from the gdpr perspective," *Computers Security*, nov 2021. ISSN 0167-4048. doi:10.1016/J.COSE.2021.102402.

[8] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *CoRR*, vol. abs/1908.07873, 2019. [Online]. Available: http://arxiv.org/abs/1908.07873

[9] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, "Federated learning of deep networks using model averaging," *CoRR*, vol. abs/1602.05629, 2016. [Online]. Available: http://arxiv.org/abs/1602.05629

[10] X. Gansel, M. A. Mary, and A. van Belkum, "Semantic data interoperability, digital medicine, and e-health in infectious disease management: a review," *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 38, pp. 1023–1034, 2019.

[11] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, and e. a. D'Oliveira, Rafael G. L., "Advances and open problems in federated learning," 2019. [Online]. Available: https://arxiv.org/abs/1912.04977

[12] B. L. Weyer J, Grobbelaar A, "Ebola virus disease: History, epidemiology and outbreaks." *Current Infectious Disease Reports,*, p. 17(5):1–8, 2015.

[13] K. M. B Goldacre, S Harrison and C. Heneghan, "Who consultation on data and results sharing during public health emergencies," sep 2015 centre for evidence-based medicine background briefing.

[14] T. F. K. R. B. F. P. S. Andres Colubri, Tom Silver, "Transforming clinical data into actionable prognosis models: Machine-learning framework and field-deployable app to predict outcome of ebola patients," *PLoS Neglected Tropical Diseases*, 2016. doi: 10.1371/journal.pntd.0004549.

[15] M. S. V. W. A. F. T. S. J. G. S. R. F. G. D. S. G. A. C. L. A. Colubri, M. A. Hartley and P. C. Sabeti, "Machine-learning prognostic models from the 2014–16 ebola outbreak: Data-harmonization challenges, validation strategies, and mhealth applications," *EClinicalMedicine*, p. 11:54–64, may 2019. ISSN 25895370.

[16] A.-M. T. H. H. O.-W. M. S. B. M. A. A.-D. M. F. Mary-Anne Hartley, Alyssa Young, "Predicting ebola severity: A clinical prioritization score for ebola virus disease," *PLoS Neglected Tropical Diseases*, feb 2017. ISSN 19352735. doi: 10.1371/ journal.pntd.0005265.