FedBID and FedDocs: A Dataset and System for Federated Document Analysis

Daniel Perazzo[1,†], Thiago de Souza[1,†], Pietro Masur[1], Eduardo de Amorim[1], Pedro de Oliveira[1], Kelvin Cunha[1], Lucas Maggi[1], Francisco Simões[1,2], Veronica Teichrieb[1] and Lucas Kirsten[3]
[1]*Voxar Labs, Centro de Informática, Universidade Federal de Pernambuco, Recife/PE, Brazil*
[2]*Visual Computing Lab, Departamento de Computação, Universidade Federal Rural de Pernambuco, Recife/PE, Brazil*
[3]*HP Inc., Porto Alegre/RS, Brazil*

Keywords: Federated Learning, Document Analysis, Privacy, Dataset.

Abstract: Data privacy has recently become one of the main concerns for society and machine learning researchers. The question of privacy led to research in privacy-aware machine learning and, amongst many other techniques, one solution gaining ground is federated learning. In this machine learning paradigm, data does not leave the user's device, with training happening on it and aggregated in a remote server. In this work, we present, to our knowledge, the first federated dataset for document classification: FedBID. To demonstrate how this dataset can be used for evaluating different techniques, we also developed a system, FedDocs, for federated learning for document classification. We demonstrate the characteristics of our federated dataset, along with different types of distributions possible to be created with our dataset. Finally, we analyze our system, FedDocs, in our dataset, FedBID, in multiple different scenarios. We analyze a federated setting with balanced categories, a federated setting with unbalanced classes, and, finally, simulating a siloed federated training. We demonstrate that FedBID can be used to analyze a federated learning algorithm. Finally, we hope the FedBID dataset allows more research in federated document classification. The dataset is available in https://github.com/voxarlabs/FedBID.

1 INTRODUCTION

The fast development of learning-based techniques, especially Deep Learning (DL), inspired new applications in different areas of computing (Xu et al., 2019). These applications became more accurate and efficient through increasingly robust learning neural networks and access to more data. For example, numerous applications on computer vision, using deep learning techniques, emerged in recent years in fields such as robotics, autonomous driving, and assistive technologies (Sünderhauf et al., 2018).

Nevertheless, for distributed training with edge devices, deep learning has traditionally followed a centralized training approach, in which third-party servers were responsible for the entire training process and where all data is sent and stored in the server (Huang et al., 2017). In this type of training, the user's data needs to be sent, stored, and processed on servers to train the neural network, which needs to be retrained continuously over time (Mayr et al., 2019). However, this type of centralized training is highly susceptible to security breaches due to all data being stored in a single server.

In this scenario, Federated Learning (FL) is a new paradigm to distribute neural network training with data privacy as a main requirement (McMahan et al., 2017). In this setting, part or all of the training process is addressed in remote devices, minimizing the exchange of sensitive data from users since customers' data is never sent to the network or accessed by an external agent directly. The models are updated over time, adding new information that reflects the data distribution for each user.

Over the last few years, different strategies and architectures have been proposed in FL for different areas (Liu et al., 2020; Li et al., 2019; Zhu et al., 2021). However, consolidated benchmarks and datasets are lacking in the FL literature to validate applications (Caldas et al., 2018). In most cases, datasets are either not publicly available, overly simple, or do not have adequate distribution to assess FL challenges (Caldas et al., 2018).

In this context, we present a new dataset that

[†]Denotes equal contribution

Perazzo, D., de Souza, T., Masur, P., de Amorim, E., de Oliveira, P., Cunha, K., Maggi, L., Simões, F., Teichrieb, V. and Kirsten, L.
FedBID and FedDocs: A Dataset and System for Federated Document Analysis.
DOI: 10.5220/0011658700003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 5: VISAPP*, pages 551-558
ISBN: 978-989-758-634-7; ISSN: 2184-4321

contains a well-defined non-IID distribution. The BID (de Sá Soares et al., 2020) is a dataset for identity document classification, containing annotations for different types of Brazilian documents. Through analyzing BID, we could identify data patterns and classes that we can map to existing challenges in Federated Learning. In this way, we propose FedBID, a new dataset in federated document classification containing $21,600$ examples annotated from an accurate non-IID distribution, to evaluate new challenges not mapped by previous datasets. Among these challenges, FedBID can consider test scenarios with a severe imbalance of examples in different users, with variations related to classes, number of samples, image quality, and annotation reliability. Furthermore, FedBID was designed to handle various tasks such as labeling the document type, orientation, source organization, and document owner. We also constructed a system, FedDocs, for federated document classification and evaluated the accuracy of FedDocs in our dataset.

As the main contributions, we highlight the following:

- FedBID, the first federated document classification dataset (Section 4);
- FedDocs, a system for federated learning for document classification (Section 3);
- Finally, we evaluate our proposed system on FedBID, with various configurations for federated learning. (Section 5).

2 RELATED WORKS

Federated learning is a new paradigm for decentralized learning (McMahan et al., 2017) and a relatively new research field, with many challenges and open problems still present in the research community, as shown in recent surveys (Kairouz et al., 2021; Li et al., 2020). Below we list research related to our work.

Federated Learning Datasets. Due to its decentralized training procedure, federated learning datasets differ from traditional machine learning datasets, requiring data partitions for various devices. Many researchers have proposed and released datasets and evaluation benchmarks for federated learning. An essential factor for these datasets is for them to mimic non-IID distributions since they frequently occur in the real world and are still a challenge for federated learning algorithms (Shoham et al., 2019). A longstanding issue with federated learning datasets is that creating a federated non-IID dataset by artificially partitioning datasets commonly used for machine learning (e.g., CIFAR, MNIST) can create datasets that do not follow a realistic non-IID distribution. Due to this issue, (Caldas et al., 2018) created a benchmark comprised of multiple federated datasets and metrics. For example, their federated version of MNIST (FEMNIST) is partitioned according to the writer of the digit. In contrast, their dataset Sentiment140, for sentiment analysis, is partitioned based on the Twitter user. Furthermore, (He et al., 2020) introduced FedML, a framework for federated learning that includes default datasets for users to get started. Finally, (Koh et al., 2021; Luo et al., 2019) created datasets focused on real-world and in-the-wild data.

Federated Learning Applications. Numerous applications were proposed for algorithms using federated learning. For example, in NLP, Google has used federated learning to improve query suggestions in the “Google Keyboard” while preserving user privacy (Yang et al., 2018). Other applications of federated learning were also in emoji prediction (Ramaswamy et al., 2019), also by Google, and speech recognition (Paulik et al., 2021), by Apple. In computer vision, FedVision (Liu et al., 2020) introduced a platform for object detection using federated learning. Furthermore, one field that has been receiving attention in federated learning is healthcare, where patient data is very sensitive (Antunes et al., 2022). Finally, with the expansion of edge devices and the internet of things, federated learning is increasingly receiving attention due to possible applications for training on edge devices (Kontar et al., 2021)

Document Analysis. Document analysis is a longstanding problem in computer vision with more than two decades of research (Liu et al., 2021). Although in the past decades, researchers focused their work on traditional computer vision techniques, such as image processing and pattern matching, to perform their analysis (Love et al., 2013), more recently, machine learning approaches have shown promising results (Li et al., 2021). For example, (Kang et al., 2014) interpreted document classification as an image classification problem and used Convolutional Neural Networks (CNNs) to perform the task. Researchers aiming to stimulate progress in the area created numerous new datasets, competitions, and benchmarks for document analysis using data from many different countries (Burie et al., 2015; Bulatovich et al., 2022a; Chernyshova et al., 2021; Bulatovich et al., 2022b; Polevoy et al., 2022). In this context, (de Sá Soares et al., 2020) created a document classification based on Brazilian personal documents. Document identification is a fascinating case study for federated learning since privacy is a significant concern among users because malicious users can use the leaked information associated with identification details to perpetrate financial fraud, false identity, and many other crimes.

3 FedDocs

In this section, we explain our FedDocs system. To do so, we will detail more about its architecture, the process in which the clients are trained, and how the server aggregates the weights.

3.1 Architecture of the System

Our application simulates a remote Federated Learning (FL) architecture in different devices and networks. The module comprises two applications: a desktop client and a server app. Both are written in Python and ran over the docker virtualization engine. The docker image contains the Flower library (Beutel et al., 2020), a general FL framework. Flower provides tools, data structures, and protocols to perform federated learning training across different devices, using the gRPC (Google Remote Procedure Call) framework (Marculescu, 2015).

Both applications (Client and Server modules) are the primary tools of our architecture and can be used in a real scenario to deploy the applications on different devices. Moreover, the architecture provides tools to adapt various applications or datasets for supervised image classification quickly. In our case, more specifically, dealing with image document classification.

3.2 Client Training

The desktop client is responsible for locally training the received model from the server, using the available data, and sending back the updated parameters to the server. We perform this training similarly to the standard centralized setting, where each local data sample contributes to the learning progress in every epoch. The process finishes when it achieves a stop criteria, such as the maximum number of local epochs or a minimum loss value. In this paper, we use the number of local epochs processed on each device with their local dataset as the stop criteria.

After the training is concluded in all clients, the trained model weights are sent back from the clients to the server for the resulting aggregated model. Then the server joins the *knowledge* and *information* of each client neural network model into one *global* neural network model.

3.3 Server Aggregation

The server coordinates the training process, aggregates the client parameters (i.e., weights) in each round, and retrieves the training metrics (i.e., accuracy, number of samples). We compute the total accuracy by taking the mean of the accuracy for each device on their local test set. To implement this task, we use the strategy abstraction in Flower, where each strategy provides instructions to train and evaluate models on clients, and perform the aggregationon the server.

As aforementioned, the connected clients send the updated weights to the server for the aggregation phase. Hence, no data leaves the local device, and only the model parameters go through the network.

In each round (i.e., training iteration), the server aggregates the local client model's weights through an aggregation method, in our case FedAvg (McMahan et al., 2017). FedAvg makes a weighted average of the local models and generates a new aggregated global model. Then, this new aggregated global model is sent back to the clients for evaluation. Finally, new clients are selected for the next training cycle. The process repeats until the model converges or achieves a stop criteria.

4 FedBID

In this section, we describe our new dataset FedBID, based on the BID Dataset (de Sá Soares et al., 2020).

4.1 Summary of BID Dataset

BID Dataset (Brazilian Identity Document Dataset) is a dataset composed of images of the Brazilian Driver's License (a.k.a. CNH), Natural Person registry document (a.k.a. CPF), and the Brazilian identification document (a.k.a. RG) that aims to help researchers with numerous challenges of computer vision for document automated processing, such as classification and segmentation, etc. We used this dataset since it was generated using a process that anonymizes the original publicly available data, creating a dataset that complies with Brazilian data privacy laws (LGPD, 2018). Based on the original BID dataset, we manually created a new set of labels and a distribution of documents per various clients, as we proceed to detail shortly.

The setup of the original BID dataset allows a classification problem between the previously stated document classes (CPF, CNH, and RG) and their respective back and front counterparts. In this manner, we have the following six classes: CPF_Front, CPF_Back, CNH_Front, CNH_Back, RG_Front, and RG_Back.

4.2 Distribution and Methodology

BID was initially conceived for the utilization of segmentation ML systems. Each BID image has a corresponding label with information from the image segmentation. Furthermore, images of different documents are categorized into different folders. This setting, although sufficient for traditional classification systems, lacks some information that may be relevant, such as orientation, which can be important for some applications, such as OCR systems.

Using the BID as a baseline, we further build another system of classes. Since determining the orientation of a document is essential for many OCR systems to extract the text correctly, we added this type of class to the BID dataset. To do this, we manually annotate each document image with its orientation (0°, 90°, 180°, 270°) using an interface from (Goecks et al., 2021). Since we have four different orientations with six classes of documents, we created a set of 24 categories by a cartesian product.

Furthermore, some of the documents on the BID dataset contain information regarding the emission issuer. To condense this information into a single label file, we have built a script for automatically retrieving this information from the segmentation labels. It works by, for each image, matching its label file context with a list containing all the document issuers in Brazil. We can use the document issuer labels to simulate a cross-silo federated data distribution. For example, in Figure 2, we detail the distribution of the document issuers found on the dataset. Details about the FedBID dataset are shown in Table 1 and Figure 1. Separating these issuers is interesting since, with this information, we can perform a simulation of a siloed federated learning setting.

Table 1: FedBID dataset description.

Split	Samples	Average Per Class
All Data	21600	3600 ± 0
Train	15120	2520 ± 12.5
Test	6480	1080 ± 12.5

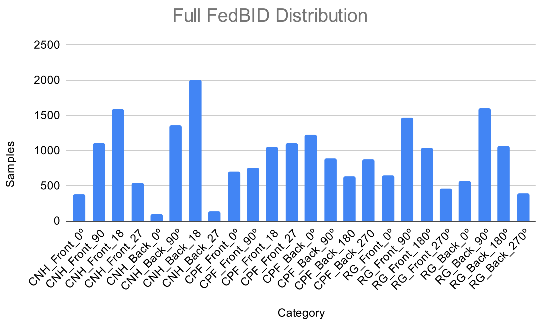

Figure 1: The distribution for the 24 classes. The FedBID distribution is unbalanced and non-IID. The classes in this histogram are labeled in the following manner: **"(document_class)_(Front/Back)_(angle)"**.

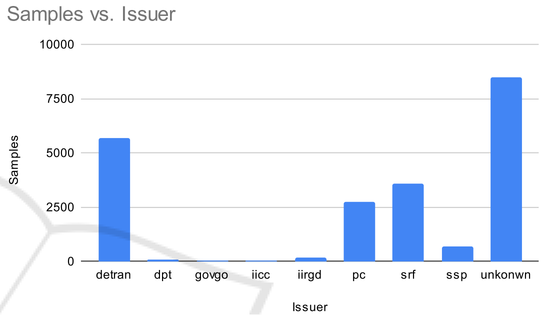

Figure 2: Details on the nine different issuers present on FedBID. Those documents that do not provide this information are displayed in the unknown column.

4.3 Available Distributions

To verify different settings for federated algorithms, we included different distributions of our data into different distributions to evaluate many aspects of federated training, such as initializing with a pre-trained model on some categories. In this setting, we use all 24 categories on each device, naming it a *balanced* dataset. To evaluate algorithms, we built three scenarios with 100, 500, and 1000 clients for each distribution below:

- Full Dataset: All samples from the entire dataset are partitioned almost equally between the devices, without repetition;
- 50% of train data for centralized training: We separate 50% from the entire dataset to pre-train a centralized model.

We also wanted to observe the effect of different categories on each device. To do so, we change the range of the number of categories in each device for these distributions, randomly removing some classes from each client and partitioning the training dataset into 100 clients. We name this distribution a *unbalanced* dataset.

5 EXPERIMENTS

In this section, we propose a series of experiments that aim to demonstrate how our dataset can be used to benchmark a federated document classification system, in our case, the FedDocs system.

5.1 Training Setup

We use our system, FedDocs, with the FedAvg aggregator for all experiments and the same training local configuration. The chosen model architecture is the EfficientNet B0 (Tan and Le, 2019). We resize the documents' image inputs to a resolution of 224x224 RGB pixels. The number of local epochs is set to $E = 1$, and the batch size is $B = 16$. We use the Adam (Kingma and Ba, 2014) optimizer with learning-rate $lr = 0.0005$.

5.2 Results

In this section, we present the results of our experiments with our federated learning tests. All reported results are related to our test set.

5.2.1 Federated Training with a Balanced Dataset

In Figure 3, we present the training results of our solution in the BID dataset with no pre-trained model for different numbers of clients and use the *balanced* dataset distributed for each device. We evaluate the model after training it for one epoch.

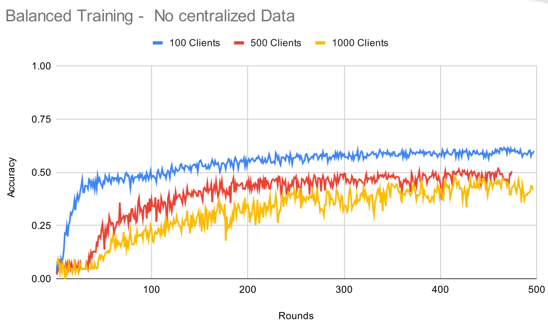

Figure 3: Federated results. Each curve represents a training with a different number of total clients: **(blue):** 100 clients; **(red):** 500 clients; **(yellow):** 1000 clients.

As presented in Figure 4, using a model pre-trained on 50%, the results are similar even if we increase the number of clients. Although, as we note, the federated training does not seem to help gain extra accuracy points, the training course appears to decrease the accuracy of the global model slightly.

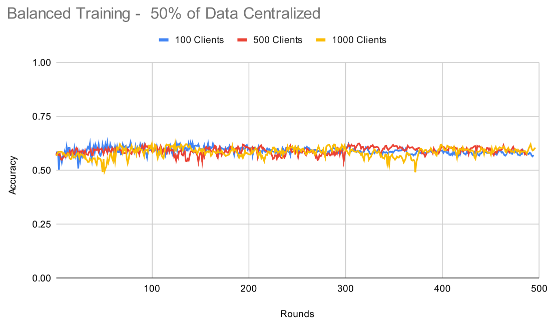

Figure 4: Federated training results starting from a centralized model pre-trained with 50% of training data. Each curve represents a training with a different number of total clients: **(blue):** 100 clients; **(red):** 500 clients; **(yellow):** 1000 clients.

In this case, we can see some challenges in our dataset that researchers could address with new federated techniques, for example, a catastrophic forgetting (Kirkpatrick et al., 2017) of a pre-trained model as seen in the work of (Shoham et al., 2019). Thus, FedBID could be considered a dataset contribution for future Federated Learning research.

5.2.2 Federated Training with an Unbalanced Dataset

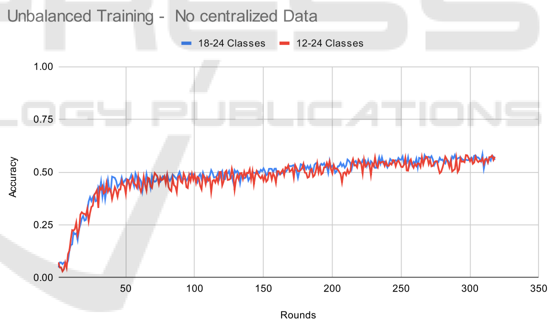

Figure 5: Unbalanced federated results without pre-trained. **(blue):** The classes vary between 18 and 24. **(red):** The classes vary between 12 and 24

In Figures 5 and 6, we present the results of training our model in the *unbalanced* dataset, with a varying range of possible categories on each device. We can see similar behavior to federated training in the *balanced* setting. In the tests without the pre-trained centralized models using all categories, it is noticeable that the convergence is slower since the models only reach more than 50% of accuracy near round 200, while for the *unbalanced* setting, we obtained it around round 100. Faster convergence is crucial since it saves clients' data transfer and battery consumption. The difference between varying the range of categories in each device is small, even with the reduction

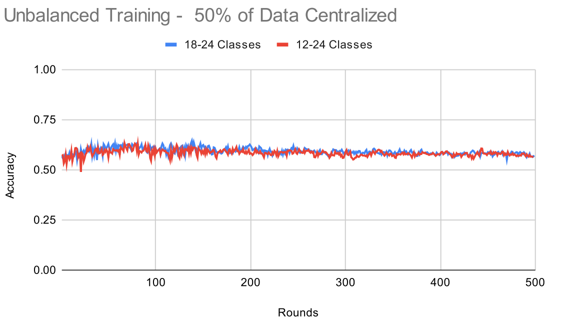

Figure 6: Unbalanced federated training results starting from a centralized model pre-trained with 50% of training data. **(blue):** The classes vary between 18 and 24. **(red):** The classes vary between 12 and 24

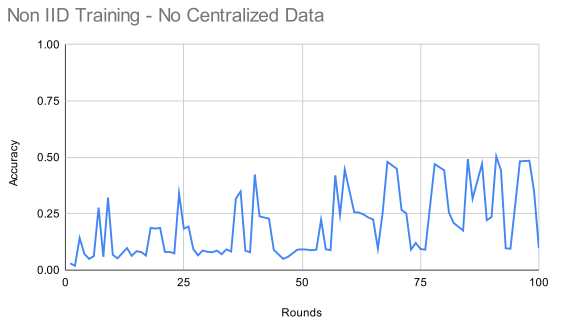

Figure 7: Non-idd federated training results using siloed partition where each contains data from one exclusive issuer.

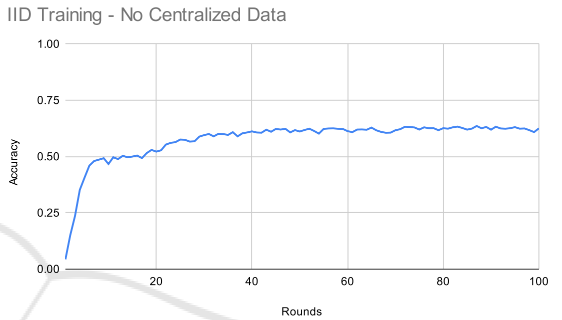

Figure 8: Federated training results using nine siloed partitions containing balanced IID data from all issuers.

of categories between them. However, increasing this unbalance in categories between the clients can make learning processing more challenging. These scenarios can be helpful to test new methods to deal with the convergence in *unbalanced* configurations, for example, using the FedProx technique (Li et al., 2020), since the local domain diverges from the global domain.

In the tests with the pre-trained model (Figure 4), we can observe a similar behavior as the experiments with all categories, with the results showing a tendency to the neural network models "forgetting" the centralized knowledge, and, as a consequence, don't learning between the rounds. Therefore, we can explore this scenario's previous challenges in the *unbalanced* scenario.

5.2.3 Federated Training with Silloed Data

In this test, our architecture uses the nine natural partitions from FedBID according to the document issuer, presented in Figure 2, where each one has only one type of document, but with the document style according to its owner, except the unknown partition which has classes from many unknown users and different document types. This setting is near to a real scenario of federated training where the data is non-IID. Our architecture results showed this distribution's impact on model convergence.

As shown in Figure 7, the model convergence does not increase gradually as in the IID scenarios. On the contrary, the curve appeared much fuzzier with jumps between 6% to 51% of accuracy without a clear indication of when the model will converge. This situation is caused due to the difference again between the local domains and the global, which, now, for this dataset, is the most diverse due to the data heterogeneity level.

On the other hand, in Figure 8, we tested the same number of partitions from the last experiment now with all data distributed in a balanced form, we can see a now expected smooth convergence with a gradually increasing accuracy. This result highlights the challenge of dealing with a non-IID distribution since this IID scenario is not possible in a real case due to privacy, opening a new set of federated challenges for document categorization.

6 CONCLUSION AND FUTURE WORKS

Our experiments show that our new FedBID dataset can be a benchmark for federated learning in Non-IID tasks. To perform the tests, we created an application named FedDocs to perform federated document classification. Our new federated dataset gives the first federated dataset for document classification. For the federated learning community, we also provide a dataset with an attractive property, a natural federated distribution based on issuers. As a result, researchers could explore cross-silo labeling in future works. In future work, we intend to construct federated learning algorithms for non-IID distributions exploring our dataset for future work.

ACKNOWLEDGMENT

This paper was achieved in cooperation with HP Inc. R&D Brazil, using incentives of the Brazilian Informatics Law (n°. 8.2.48 of 1991). The authors would like to thank Thomas da Silva Paula for being the project conceiver and Eduardo Chagas for the management support during this work.

REFERENCES

Antunes, R. S., André da Costa, C., Küderle, A., Yari, I. A., and Eskofier, B. (2022). Federated learning for healthcare: Systematic review and architecture proposal. *ACM Transactions on Intelligent Systems and Technology (TIST)*.

Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Parcollet, T., de Gusmão, P. P., and Lane, N. D. (2020). Flower: A friendly federated learning research framework. *arXiv preprint arXiv:2007.14390*.

Bulatovich, B. K., Vladimirovna, E. E., Vyacheslavovich, T. D., Sergeevna, S. N., Sergeevna, C. Y., Zuheng, M., Jean-Christophe, B., and Muzzamil, L. M. (2022a). Midv-2020: A comprehensive benchmark dataset for identity document analysis. *Computer Optics*, 46(2):252–270.

Bulatovich, B. K., Vladimirovna, E. E., Vyacheslavovich, T. D., Sergeevna, S. N., Sergeevna, C. Y., Zuheng, M., Jean-Christophe, B., and Muzzamil, L. M. (2022b). Midv-2020: A comprehensive benchmark dataset for identity document analysis. 46(2):252–270.

Burie, J.-C., Chazalon, J., Coustaty, M., Eskenazi, S., Luqman, M. M., Mehri, M., Nayef, N., Ogier, J.-M., Prum, S., and Rusiñol, M. (2015). Icdar2015 competition on smartphone document capture and ocr (smartdoc). In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1161–1165. IEEE.

Caldas, S., Duddu, S. M. K., Wu, P., Li, T., Konečnỳ, J., McMahan, H. B., Smith, V., and Talwalkar, A. (2018). Leaf: A benchmark for federated settings. *arXiv preprint arXiv:1812.01097*.

Chernyshova, Y., Emelianova, E., Sheshkus, A., and Arlazarov, V. V. (2021). Midv-lait: a challenging dataset for recognition of ids with perso-arabic, thai, and indian scripts. In *International Conference on Document Analysis and Recognition*, pages 258–272. Springer.

de Sá Soares, A., das Neves Junior, R. B., and Bezerra, B. L. D. (2020). Bid dataset: a challenge dataset for document processing tasks. In *Anais Estendidos do XXXIII Conference on Graphics, Patterns and Images*, pages 143–146. SBC.

Goecks, V. G., Waytowich, N., Watkins, D., and Prakash, B. (2021). Combining learning from human feedback and knowledge engineering to solve hierarchical tasks in minecraft. *arXiv preprint arXiv:2112.03482*.

He, C., Li, S., So, J., Zeng, X., Zhang, M., Wang, H., Wang, X., Vepakomma, P., Singh, A., Qiu, H., et al. (2020). Fedml: A research library and benchmark for federated machine learning. *arXiv preprint arXiv:2007.13518*.

Huang, Y., Ma, X., Fan, X., Liu, J., and Gong, W. (2017). When deep learning meets edge computing. In *2017 IEEE 25th international conference on network protocols (ICNP)*, pages 1–2. IEEE.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210.

Kang, L., Kumar, J., Ye, P., Li, Y., and Doermann, D. (2014). Convolutional neural networks for document image classification. In *2014 22nd International Conference on Pattern Recognition*, pages 3168–3172. IEEE.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526.

Koh, P. W., Sagawa, S., Marklund, H., Xie, S. M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R. L., Gao, I., et al. (2021). Wilds: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning*.

Kontar, R., Shi, N., Yue, X., Chung, S., Byon, E., Chowdhury, M., Jin, J., Kontar, W., Masoud, N., Nouiehed, M., et al. (2021). The internet of federated things (ioft). *IEEE Access*.

LGPD (2018). General law for personal data protection (lgpd). http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Accessed: 2022-09-19.

Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X., and He, B. (2021). A survey on federated learning systems: vision, hype and reality for data privacy and protection. *IEEE Transactions on Knowledge and Data Engineering*.

Li, T., Sahu, A. K., Talwalkar, A., and Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*.

Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M. J., et al. (2019). Privacy-preserving federated brain tumour segmentation. In *International workshop on machine learning in medical imaging*.

Liu, L., Wang, Z., Qiu, T., Chen, Q., Lu, Y., and Suen, C. Y. (2021). Document image classification: Progress over two decades. *Neurocomputing*, 453:223–240.

Liu, Y., Huang, A., Luo, Y., Huang, H., Liu, Y., Chen, Y., Feng, L., Chen, T., Yu, H., and Yang, Q. (2020). Fedvision: An online visual object detection platform powered by federated learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Love, P. et al. (2013). Document analysis. In *Research in the college context*, pages 99–112. Routledge.

Luo, J., Wu, X., Luo, Y., Huang, A., Huang, Y., Liu, Y., and Yang, Q. (2019). Real-world image datasets for federated learning. *arXiv preprint arXiv:1910.11089*.

Marculescu, M. (2015). Introducing grpc, a new open source http/2 rpc framework. https://developers.googleblog.com/2015/02/introducing-grpc-new-open-source-http2.html.

Mayr, A., Kißkalt, D., Meiners, M., Lutz, B., Schäfer, F., Seidel, R., Selmaier, A., Fuchs, J., Metzner, M., Blank, A., et al. (2019). Machine learning in production–potentials, challenges and exemplary applications. *Procedia CIRP*, 86:49–54.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR.

Paulik, M., Seigel, M., Mason, H., Telaar, D., Kluivers, J., van Dalen, R., Lau, C. W., Carlson, L., Granqvist, F., Vandevelde, C., et al. (2021). Federated evaluation and tuning for on-device personalization: System design & applications. *arXiv preprint arXiv:2102.08503*.

Polevoy, D. V., Sigareva, I. V., Ershova, D. M., Arlazarov, V. V., Nikolaev, D. P., Ming, Z., Luqman, M. M., and Burie, J.-C. (2022). Document liveness challenge dataset (dlc-2021). *Journal of Imaging*, 8(7):181.

Ramaswamy, S., Mathews, R., Rao, K., and Beaufays, F. (2019). Federated learning for emoji prediction in a mobile keyboard. *arXiv preprint arXiv:1906.04329*.

Shoham, N., Avidor, T., Keren, A., Israel, N., Benditkis, D., Mor-Yosef, L., and Zeitak, I. (2019). Overcoming forgetting in federated learning on non-iid data. *arXiv preprint arXiv:1910.07796*.

Sünderhauf, N., Brock, O., Scheirer, W., Hadsell, R., Fox, D., Leitner, J., Upcroft, B., Abbeel, P., Burgard, W., Milford, M., et al. (2018). The limits and potentials of deep learning for robotics. *The International journal of robotics research*, 37(4-5):405–420.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.

Xu, M., Liu, J., Liu, Y., Lin, F. X., Liu, Y., and Liu, X. (2019). A first look at deep learning apps on smartphones. In *The World Wide Web Conference*, pages 2125–2136.

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

Zhu, Z., Hong, J., and Zhou, J. (2021). Data-free knowledge distillation for heterogeneous federated learning. In *International Conference on Machine Learning*, pages 12878–12889. PMLR.