Exploring Federated Learning for Speech-based Parkinson's Disease Detection

Athanasios Sarlas
Alexandros S. Kalafatelis
thanossar@pms.uoa.gr
alexkalafat@pms.uoa.gr
Department of Ports Management and Shipping
National and Kapodistrian University of Athens
Psachna, Evia, Greece

Georgios Alexandridis
Department of Digital Industry Technologies
National and Kapodistrian University of Athens
Psachna, Evia, Greece
gealexandri@uoa.gr

Michail-Alexandros Kourtis
Institute of Informatics and Telecommunications
National Centre for Scientific Research "DEMOKRITOS" (NCSRD)
Aghia Paraskevi, Attica, Greece
akis.kourtis@iit.demokritos.gr

Panagiotis Trakadas
Department of Ports Management and Shipping
National and Kapodistrian University of Athens
Psachna, Evia, Greece
ptrakadas@pms.uoa.gr

ABSTRACT

Parkinson's Disease is the second most prevalent neurodegenerative disorder, currently affecting as high as 3% of the global population. Research suggests that up to 80% of patients manifest phonatory symptoms as early signs of the disease. In this respect, various systems have been developed that identify high risk patients by analyzing their speech using recordings obtained from natural dialogues and reading tasks conducted in clinical settings. However, most of them are centralized models, where training and inference take place on a single machine, raising concerns about data privacy and scalability. To address these issues, the current study migrates an existing, state-of-the-art centralized approach to the concept of federated learning, where the model is trained in multiple independent sessions on different machines, each with its own dataset. Therefore, the main objective is to establish a proof of concept for federated learning in this domain, demonstrating its effectiveness and viability. Moreover, the study aims to overcome challenges associated with centralized machine learning models while promoting collaborative and privacy-preserving model training.

CCS CONCEPTS

• **Computing methodologies** → **Speech recognition**; **Multi-agent systems**; *Supervised learning by classification.*

KEYWORDS

Federating Learning, Parkinson's Disease, Speech Articulation

ARES 2023, August 29–September 01, 2023, Benevento, Italy

ACM ISBN 979-8-4007-0772-8/23/08.
https://doi.org/10.1145/3600160.3605088

ACM Reference Format:
Athanasios Sarlas, Alexandros S. Kalafatelis, Georgios Alexandridis, Michail-Alexandros Kourtis, and Panagiotis Trakadas. 2023. Exploring Federated Learning for Speech-based Parkinson's Disease Detection. In *The 18th International Conference on Availability, Reliability and Security (ARES 2023), August 29–September 01, 2023, Benevento, Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3600160.3605088

1 INTRODUCTION

Parkinson's disease (PD) is the second leading neurodegenerative disorder, currently affecting as high as 3% of the world population above the age of 65 [16]. Typical neuropathological features of PD include bradykinesia and resting tremors, alongside non-motor symptoms that can detrimentally affect the patients' Quality of Life (QoL) [22]. In addition, phonatory symptoms are widely considered as early indications of neurological diseases. In the case of PD, research suggests that as many as 80% of PD patients can clinically manifest oral communication disorders during the early stages of the disease, including dysphonia, imprecise articulation and dysprosody. These voice alterations are associated with impaired muscles caused by PD, which are responsible for speech coordination [3, 24].

The early detection of patients with PD, or with a high risk of developing it in their lifetime, can significantly enhance patient outcomes and QoL by enabling timely interventions to delay disease progression [16, 20]. Recent advances in machine learning (ML) and speech recognition have exhibited promising results in detecting PD by analyzing speech patterns, facilitating early differential diagnosis [10, 21, 23]. However, conventional centralized ML approaches present substantial risks regarding patient data privacy, raising valid concerns about their security and scalability. Specifically, centralized methods introduce inherent vulnerabilities regarding the likelihood of potential unauthorized access or misuse of patient data used for training the ML models. Moreover, the incorporation of multiple sources that generate vast volumes of data can lead to challenges about the scalability of model training, requiring considerable computational resources [6]. These challenges emphasize the critical need to explore alternative methodologies, that prioritize patient data privacy, while at the same time ensuring efficient and scalable model training and inference.

In an effort to address the aforementioned challenges, the current work outlines the migration of an existing state-of-the-art PD detection system [10], based on patient speech data, under a Federated Learning (FL) context. FL is a decentralized learning paradigm, enabling collaborative and local model training on client devices, without the need of data exchange, sharing only model parameters with a central server for aggregation. In this way, FL addresses issues of both data governance and privacy [6, 19], while recent research indicates that models trained under a FL approach have the ability to attain and possible overcome the performance levels of centrally trained models [4, 13].

The use of FL in PD detection also presents a significant contribution to the broader application of FL in healthcare. Privacy concerns are paramount when dealing with sensitive medical data, and FL provides a promising solution by ensuring that patient data remain decentralized and secure. This is achieved by training ML models locally on client devices, with only model parameters being shared with a central server for aggregation. By avoiding the need for raw data exchange, FL offers enhanced privacy protection compared to traditional centralized approaches.

The novelty of this study lies in the exploration of FL as a viable framework for PD detection using only speech data, collected from patients and healthy individuals. While previous studies on this domain have solely focused on conventional ML paradigms, this work aims to validate the feasibility of applying FL to address the limitations of data privacy and scalability.

The remainder of this paper is organized as follows; Section 2 provides a comprehensive review of the current state-of-the-art in ML-driven PD and FL. Section 3 introduces the methodology, including information regarding the utilized patient data, the model architecture, the examined FL strategies and the experimental setup. Section 4 presents the results of the study and discusses their impact, while Section 5 concludes the paper and summarizes the key findings of this work, highlighting potential directions for future research.

2 RELATED WORK

The application of ML algorithms for PD detection has gained significant attention in recent years. Several studies have explored the use of different data modalities (e.g. handwritten text, speech, motor and imaging data) to uncover relevant features that may further aid in the clinical diagnosis of PD and its atypical representations. Recently, there has been a growing interest on the use of speech analysis, showcasing promising results, while also providing valuable insight into the progression of the disease [7].

In this respect, Celik and Omurca [2] employed several ML algorithms such as logistic regression, support vector machines, gradient boosting and random forests alongside with principal component analysis to classify PD using a set of 26 speech features. The outcome of this study demonstrated that logistic regression attained the highest accuracy score when using the full feature set, while SVM performed best with a linear kernel. In a separate study, Quan et al. [17] introduced a bidirectional long short-term memory (Bi-LSTM) model for PD detection, focusing on capturing dynamic articulation features rather than relying solely on static ones. Their results showcased that the Bi-LSTM model outperformed traditional ML approaches, achieving significantly higher accuracy scores. Janbakhshi & Kodrasi [10] recently introduced a deep learning (DL) approach that achieves state-of-the-art results in PD detection from speech data. More specifically, their methodology aims at mitigating the effects of speaker variabilities (which are not related to PD) by obtaining (speaker) identity-invariant representations. The learning objective is achieved through the adversarial training of an auto-encoder and a speaker identification task. In conclusion, all of the aforementioned studies highlight the potential of ML and DL in assisting healthcare professionals in clinical environments, by enabling early PD diagnosis.

The increasing demand for privacy-preserving ML solutions in clinical settings has spurred the exploration of alternative approaches like FL, which offers a collaborative training method for ML models, while ensuring decentralized and secure storage of sensitive data. Recent studies have demonstrated the feasibility of FL in various healthcare applications. For example, Jorge et al. [11] employed FL to detect motor symptoms related to Freezing of Gait (FoG). The authors evaluated the FL model compared to a centralized approach and observed similar accuracy rates, demonstrating that FL can effectively detect FoG while preserving patient privacy. Likewise, Dipro et al. [5] assessed the performance of the VGG16, VGG19 and InceptionV3 DL models in FL, trained on single-photon emission computed tomography data, with VGG19 achieving the highest accuracy rate. Recently, Arasteh et al. [1] proposed the use of a pre-trained Wav2Vec model in FL context. The conducted experiments utilized speech data from three different languages, involving the rapid repetition of three specific syllables. Their findings demonstrated that the FL model outperformed all the individual conventional models.

Judging from the above analysis and to the best of our knowledge, there seems to be limited prior work on the application of FL for PD detection using speech data. While FL has been introduced and explored in healthcare, its specific application for PD detection remains relatively unexplored. In this respect, the current work aims to bridge this gap by investigating the feasibility and effectiveness of FL in the context of early PD diagnosis, utilizing speech data that consist of recordings from both random dialogues and reading text aloud. By leveraging the FL scenario, the objective is to to address the challenges and open questions associated with centralized ML models, enabling a collaborative and privacy-preserving model training.

3 METHODOLOGY

This section illustrates the experimental methodology for PD detection based on speech data provided by the Mobile Device Voice Recordings at King's College London (MDVR-KCL) dataset [9], which contains voice recordings from 37 individuals with early and advanced PD, as well as healthy controls. The recordings were conducted in a realistic setting resembling a phone call, where participants were asked to read a lengthy article and have a spontaneous dialogue. The recordings are stored in uncompressed waveform audio file format, with a sampling rating of $44.1kHz$. According to

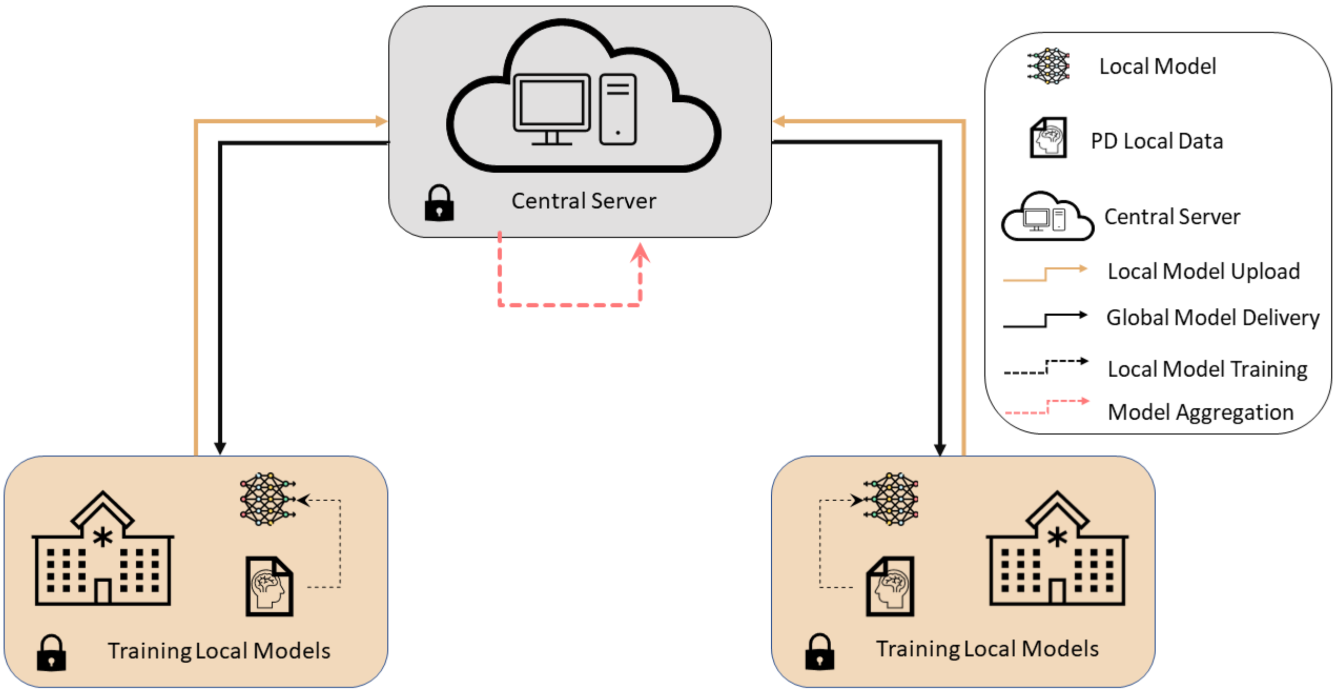

Figure 1: A high-level view of the federated learning environment

the dataset creators, labeling has been performed in compliance with Hoehn & Yahr (H&Y) scores, as well as on the UPDRS II part 5 and the UPDRS III part 18 scales [9].

Since the objective of the current work is to assess PD from voice recordings in a FL context, we didn't develop a new model from scratch, but instead chose to adapt an existing, centralized ML model in a FL setting. Out of the many models available, we chose the one of Janbakhshi et al. [10] (presented in Section 2), both because of the state-of-the-art results achieved and the availability of the source code on a public repository, which permits direct experimentation. In the FL scenario considered in our experiments, 2 clients participate in the multi-step training process, with each of them accessing a distinct part of the overall dataset.

Figure 1 illustrates a high-level view of the FL environment considered in this work. The lower part of the diagram depicts two distinct clients, which could as well be two different institutions (clinics, hospitals etc). Each client has exclusive access to its own patient data, which are not shared in-between them or with any other repository. For PD detection, the client requests from a specific ML model repository (upper part of Figure 1) a particular model, which is subsequently downloaded on its premises. The model is then trained on the local data and once training is complete, the updated model parameters are uploaded back into the server, which aggregates them into the existing model, following a specific strategy (FL strategies are going to be discussed later in this section). All data communications are performed using secure channels (e.g. SSL). In this way, data privacy is assessed on two levels; (i) on the application level, as only model parameters are communicated back to the server and not patient data and (ii) on the communication level, as data exchange is encrypted. Furthermore, this architecture is also scalable, in the sense that more clients may be easily added. As it is going to be discussed next, FL strategies can operate for an arbitrary number of clients and can also perform parameter aggregation asynchronously, meaning that it is not necessary for all clients to train their local models at the same time.

In FL-based training of the selected model [10], each client independently trains an auto-encoder, consisting of an encoder and a decoder whose parameters are denoted as θ_e and θ_d, respectively. The objective of the auto-encoder is to reconstruct the original voice input, minimizing the reconstruction loss L_{ae}. To achieve this, a convolutional neural network architecture is used, aiming at computing low-dimensional representations of voice data. More specifically, the auto-encoder consists of four convolutional layers with progressively increasing numbers of feature maps. Each layer is followed by max-pooling, batch normalization, and leaky ReLU activation functions. The final layer of the encoder is a fully connected layer that generates a final representation vector of size 128. The decoder part consists of transposed convolutional and interpolation layers, which are stacked in reverse order of the encoder. These layers aim to reconstruct the input from the final bottleneck representation.

Following, each client i trains a speaker identity (ID) module with parameters θ^i_{id}, utilizing the bottleneck representation obtained from their respective local auto-encoder. Adversarial training is employed to simultaneously minimize the auto-encoder reconstruction loss L_{ae} and maximize the speaker ID loss L_{id}, ensuring that the learned representations capture both the speaker-specific information and the reconstructed input. The optimization objective for each client is expressed according to Equation 1 that follows,

$$\arg\min_{\theta_e^i,\theta_d^i} \arg\max_{\theta_{id}^i} \left[(1-\lambda)L_{ae}^i(\theta_e^i,\theta_d^i) - \lambda L_{id}^i(\theta_e^i,\theta_{id}^i)\right] \quad (1)$$

where λ controls the trade-off between the two losses.

Simultaneously, each client employs the bottleneck representation from their auto-encoder to train the PD classifier, whose parameters are denoted as θ_{pc}^i. The PD classifier is optimized based on a task-specific loss L_{pc}, which enables it to identify pathological speech patterns accurately. Equation 2 summarizes the optimization objective for each client

$$\arg\min_{\theta_e^i,\theta_d^i,\theta_{pc}^i} \left[(1-\alpha)L_{ae}^i(\theta_e^i,\theta_d^i) + \alpha L_{pc}^i(\theta_e^i,\theta_{pc}^i)\right] \quad (2)$$

where α determines the relative importance of the reconstruction loss and the PD classifier loss.

Once the clients have completed training on their individual parts of the dataset, a FL strategy is utilized, in order to combine their learned parameters. Many relevant strategies exist in the literature, so in this work, which serves as a proof-of-concept, we examine three of the most prevalent of them, namely; (i) federated averaging (FedAvg) [14], (ii) federated averaging with server momentum (FedAvgM) [8] and (iii) adaptive federated optimization using Adam (FedAdam) [18].

In the FedAvg case, and when stochastic gradient descend is used for optimization with a fixed learning rate η, each client i computes the average gradient on its local data (Equation 3) ,

$$g_i = \nabla \frac{1}{n_i} \sum_{k\in\mathcal{P}_i} L(k;\theta_t) \quad (3)$$

where θ_t are the model parameters at iteration t, n_i are the total samples in the data partition $\mathcal{P}_i$ available at client i and L is the loss function. Then those gradients are communicated to the server via a secure channel, which in turn, aggregates them according to Equation 4, thereby ensuring that the fused parameters capture the collective information from all clients

$$\theta_{t+1} \leftarrow \theta_t - \eta \sum_{i=1}^{N} \frac{n_i}{n} g_i \quad (4)$$

where N is the total number of participating clients ($N = 2$ in the case we examine) and n is the total number of samples in the dataset. The benefits of the clients' sharing of gradient values g_i with the server, instead of the whole dataset, are threefold; (i) utilization of network resources, as it is only necessary to share value instead of the whole dataset, (ii) privacy preservation, because potentially sensitive data are not shared and (iii) reduced computational burden on the clients' side, since they need to perform computations on their share of data only.

FedAvgM [8] extends the previous strategy by considering a momentum term β when calculating model parameters, adding the relevant term to Equation 4 and producing Equation 5 below

$$\theta_{t+1} \leftarrow \theta_t - \eta \sum_{i=1}^{N} \frac{n_i}{n} g_i + \beta\Delta\theta \quad (5)$$

As in the case of centralized SGD, its federated counterpart can also get stuck in areas of the parameter space that have no gradient. Therefore, the momentum term allows FedAvgM to build inertia in the parameter search, overcoming the oscillations of the noisy gradients. Naturally, β is a hyper-parameter of this strategy. Finally, FedAdam [18] is the FL-based implementation of the widely used Adam optimization algorithm [12].

Once the aggregated parameters θ have been determined, a PD speech classifier θ_{pcl} is trained to minimize the task-specific loss L_{pcl}, which is then used for inference (classify pathological speech in unseen data). The optimization objectives for each client are adapted to consider their local data and the model parameters are updated based on the employed federated learning strategy (Equation 6)

$$\arg\min_{\theta} L_{pcl}(\theta,\theta_{pcl}) \quad (6)$$

Model training and evaluation are based on the same protocol used in [10]. The difference is that while the upstream task (speaker identity-invariant representation with adversarial training) is trained centrally (as in the original paper), the downstream task (PD classification, using the pre-trained representations of the upstream task) is trained in a FL context. Data are split into distinct training-validation-test sets according to a $60\% - 20\% - 20\%$ ratio, following a 10 fold cross-validation protocol. Other training hyper-parameters include the batch size which has ben set to 128, the learning rating ($\eta = 0, 02$) and the number of epochs (20). Lastly, performance is assessed at the chunk level (patient recordings are split into non-overlapping parts) using the accuracy metric.

4 RESULTS

Prior to examining the various FL strategies discussed in the previous Section, we performed a centralized training of the model on our dataset and computed the chunk level accuracy at 65.19%. This result is considerably lower when compared to the respective value (75.4%) reported in [10] and the difference is attributed to the peculiarities of the dataset used in our experiments (Section 3). The Spanish recordings extracted from the PC-GITA database [15] considered in [10] contain speech segments of PD patients and healthy controls reading text excerpts. On the other hand, the MDVR-KCL dataset contains longer talks between patients/healthy controls and doctors, resembling a more realistic examination scenario. However, from the model's perspective, the data in our case contain "artifacts" not present in [10] and hence the difference in performance.

Table 1 summarizes the results of model evaluation on the downstream task, for the FL strategies outlined in Section 3. Results are reported on a per-client basis (on the part of the test dataset available at each client) and aggregated. A first observation is that the centralized model clearly outperforms the FL strategies; only FedAdam on the part of the dataset available at the first client seems to achieve a similar accuracy. Nevertheless, it is our belief that this situation does not necessarily designate a weakness of the FL strategies over centralized training, as the latter approaches are dependant on an number of hyper-parameters (some of which have been discussed in Section 2) that have not been extensively searched for their optimal values (we have resorted to some widely used common values). Additionally, the small number of clients

Table 1: Accuracy results for the various FL strategies

	FedAvg	FedAvgM ($\beta = 0.3$)	FedAvgM ($\beta = 0.7$)	FedAdam ($\eta = 0.1, \eta_l = 0.1$)	FedAdam ($\eta = 0.1, \eta_l = 0.7$)
1st Client	58.67%	60.58%	59.69%	63.08%	60.02%
2nd Client	58.84%	57.42%	53.78%	59.48%	63.03%
Aggregate	58.74%	59.19%	57.07%	61.49%	61.35%

involved (the minimum 2 for FL) along with size of the dataset had their impact on the outcome; FL strategies would have performed better had the dataset size and number of clients been bigger.

Apart from the general remarks of the previous paragraph, Table 1 allows us to inspect the performance of the different strategies in more detail. For example, the simple FedAvg strategy achieves comparable performance on the dataset parts of both clients and the aggregate model performance is in close vicinity to their average. The addition of a small momentum term ($\beta = 0.3$) to the federated SGD boosts the model's throughput, resulting in a 2% accuracy gain. Further increasing the momentum term ($\beta = 0.7$) does not seem to yield better results, as performance slightly deteriorates. The FL adaptation of Adam, the second optimization algorithm to be considered in the current work, achieves even better results, which are very close to the ones reported by the centralized model (65.19%).

Overall, the findings above suggest that FL strategies, if properly adjusted, exhibit a potential for effective model training and aggregation, by leveraging distributed data. The aggregate model accuracy indicates the collaborative learning capability of FL, as each client's local model contributes to the overall performance enhancement. It should be noted, that even though the centralized model outperformed all FL strategies in terms of accuracy, it is important to consider the trade-offs associated with this approach. FL addresses concerns related to data privacy and scalability by keeping the raw data decentralized and aggregating parameters updates instead. Consequently, FL strategies provide viable alternatives in cases where data privacy is crucial or computational constraints are of paramount importance.

5 CONCLUSIONS

In this work, the training and inference of a state-of-the-art PD detection model under different FL strategies on patient speech data has been outlined. Even though the objective was to address challenges associated with the feasibility of the approach, certain FL strategies achieved comparable results to a centrally trained model, thereby demonstrating the potential of collaborative and privacy-preserving model training for PD detection.

Despite slightly affecting performance, it is evident that the FL strategies provide a decentralized and secure framework for training ML models, ensuring the privacy of patient data. These results are a clear indication for the potential of FL for further application in the healthcare domain. The ability of some FL strategies to attain comparable performance levels to centralized models highlights their effectiveness in leveraging distributed speech data while addressing data privacy concerns.

Moving forward, there are several areas for improvement and exploration in the application of FL for PD detection using speech data. One direction for future work is to increase the number of clients participating in the FL process. By involving a larger number of clients, it would be possible to incorporate more diverse and representative speech data into the training process. This can help improve the generalizability of the trained model and enhance its performance on a wider range of patients.

Additionally, further investigation into the impact of different FL strategies on the performance of the models could also be sought after. Exploring a broader range of strategies and evaluating their effects on model accuracy, convergence speed, and communication overhead can provide valuable insights into optimizing the aggregation process. This optimization is crucial for achieving efficient and effective model training in FL scenarios.

Furthermore, the integration of additional modalities, such as motor and imaging data, along with speech data, holds great potential for enhancing the overall accuracy and robustness of PD detection systems. By incorporating multiple modalities, it becomes possible to capture a more comprehensive view of the disease and its manifestations. However, integrating multimodal data within the FL framework requires the development of novel aggregation techniques to effectively combine the information from different modalities. Exploring and designing such aggregation methods that can leverage the strengths of each modality can further enhance the performance of FL-based PD detection systems.

In summary, future work in this area can focus on increasing the number of clients, exploring more FL strategies, and incorporating additional modalities to improve the accuracy, generalizability, and robustness of PD detection systems. These advancements have the potential to pave the way for more personalized and effective early diagnosis of PD, ultimately improving patient care and outcomes.

In conclusion, this work demonstrated the feasibility and effectiveness of FL in PD detection using speech data, showcasing its potential as a viable framework for privacy-preserving and scalable ML training in the healthcare domain. While the FL strategies' accuracy was slightly lower than of the centralized model, the benefits of data privacy and scalability offered by FL make it a promising approach for future research and application in healthcare.

ACKNOWLEDGMENTS

This research initiative is supported by the European Union's Horizon Europe Framework Programme for Research and Innovation, under the OASEES project (Grant Agreement No. 101092702). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union, which cannot be held responsible for them.

REFERENCES

[1] Soroosh Tayebi Arasteh, Cristian David Rios-Urrego, Elmar Noeth, Andreas Maier, Seung Hee Yang, Jan Rusz, and Juan Rafael Orozco-Arroyave. 2023. Federated learning for secure development of AI models for Parkinson's disease detection using speech from different languages. *arXiv preprint arXiv:2305.11284* (2023).

[2] Enes Celik and Sevinc Ilhan Omurca. 2019. Improving Parkinson's disease diagnosis with machine learning methods. In *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*. Ieee, 1–4.

[3] Rita Chiaramonte and Marco Bonfiglio. 2020. Acoustic analysis of voice in Parkinson's disease: a systematic review of voice disability and meta-analysis of studies. *Revista de neurologia* 70, 11 (2020), 393–405. https://doi.org/10.33588/rn.7011.2019414

[4] Judith Sáinz-Pardo Díaz and Álvaro López García. 2023. Study of the performance and scalability of federated learning for medical imaging with intermittent clients. *Neurocomputing* 518 (2023), 142–154. https://doi.org/10.1016/j.neucom.2022.11.011

[5] Sumit Howlader Dipro, Mynul Islam, Abdullah Al Nahian, Moonami Sharmita Azad, Amitabha Chakrabarty, and Tanzim Reza. 2022. A Federated Learning Based Privacy Preserving Approach for Detecting Parkinson's Disease Using Deep Learning. In *2022 25th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 139–144.

[6] Anastasios Giannopoulos, Panagiotis Gkonis, Petros Bithas, Nikolaos Nomikos, George Ntroulias, and Panagiotis Trakadas. 2023. Federated Learning for Maritime Environments: Use Cases, Experimental Results, and Open Issues. (2023). https://doi.org/10.36227/techrxiv.22133549.v1

[7] Pedro Gómez-Vilda, Jiri Mekyska, José M Ferrández, Daniel Palacios-Alonso, Andrés Gómez-Rodellar, Victoria Rodellar-Biarge, Zoltan Galaz, Zdenek Smekal, Ilona Eliasova, Milena Kostalova, et al. 2017. Parkinson disease detection from speech articulation neuromechanics. *Frontiers in neuroinformatics* 11 (2017), 56.

[8] Tzu-Ming Harry Hsu, Hang Qi, and Matthew Brown. 2019. Measuring the Effects of Non-Identical Data Distribution for Federated Visual Classification. arXiv:1909.06335 [cs.LG]

[9] Hagen Jaeger, Dhaval Trivedi, and Michael Stadtschnitzer. 2019. Mobile device voice recordings at King's College London (MDVR-KCL) from both early and advanced Parkinson's disease patients and healthy controls. *Zenodo* (2019).

[10] Parvaneh Janbakhshi and Ina Kodrasi. 2021. Supervised speech representation learning for Parkinson's disease classification. In *Speech Communication; 14th ITG Conference*. VDE, 1–5.

[11] J Jorge, PH Barros, R Yokoyama, D Guidoni, HS Ramos, N Fonseca, and L Villas. 2022. Applying Federated Learning in the detection of Freezing of Gait in Parkinson's disease. In *2022 IEEE/ACM 15th International Conference on Utility and Cloud Computing (UCC)*. IEEE, 195–200.

[12] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs.LG]

[13] Geun Hyeong Lee and Soo-Yong Shin. 2020. Federated learning on clinical benchmark data: performance assessment. *Journal of medical Internet research* 22, 10 (2020), e20891. https://doi.org/10.2196/20891

[14] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.

[15] Juan Rafael Orozco-Arroyave, Julián David Arias-Londoño, Jesús Francisco Vargas-Bonilla, María Claudia González-Rátiva, and Elmar Nöth. 2014. New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. European Language Resources Association (ELRA), Reykjavik, Iceland, 342–347. http://www.lrec-conf.org/proceedings/lrec2014/pdf/7_Paper.pdf

[16] Werner Poewe, Klaus Seppi, Caroline M Tanner, Glenda M Halliday, Patrik Brundin, Jens Volkmann, Schrag Anette-Eleonore, and Antony E Lang. 2017. Parkinson disease. *Nature Reviews Disease Primers* 3, 17013 (2017). https://doi.org/10.1038/nrdp.2017.13

[17] Changqin Quan, Kang Ren, and Zhiwei Luo. 2021. A deep learning based method for Parkinson's disease detection using dynamic features of speech. *IEEE Access* 9 (2021), 10239–10252.

[18] Sashank Reddi, Zachary Charles, Manzil Zaheer, Zachary Garrett, Keith Rush, Jakub Konečný, Sanjiv Kumar, and H. Brendan McMahan. 2021. Adaptive Federated Optimization. arXiv:2003.00295 [cs.LG]

[19] Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N Galtier, Bennett A Landman, Klaus Maier-Hein, et al. 2020. The future of digital health with federated learning. *NPJ digital medicine* 3, 1 (2020), 119. https://doi.org/10.1038/s41746-020-00323-1

[20] Jan Rusz, Tereza Tykalová, Michal Novotný, Evžen Růžička, and Petr Dušek. 2021. Distinct patterns of speech disorder in early-onset and late-onset de-novo Parkinson's disease. *npj Parkinson's Disease* 7, 1 (11 Nov 2021), 98. https://doi.org/10.1038/s41531-021-00243-1

[21] Antonio Suppa, Giovanni Costantini, Francesco Asci, Pietro Di Leo, Mohammad Sami Al-Wardat, Giulia Di Lazzaro, Simona Scalise, Antonio Pisani, and Giovanni Saggio. 2022. Voice in Parkinson's disease: a machine learning study. *Frontiers in Neurology* 13 (2022). https://doi.org/10.3389/fneur.2022.831428

[22] Malú Gámez Tansey, Rebecca L Wallings, Madelyn C Houser, Mary K Herrick, Cody E Keating, and Valerie Joers. 2022. Inflammation and immune dysfunction in Parkinson disease. *Nature Reviews Immunology* (2022), 1–17. https://doi.org/10.1038/s41577-022-00684-6

[23] John Michael Templeton, Christian Poellabauer, and Sandra Schneider. 2022. Classification of Parkinson's disease and its stages using machine learning. *Scientific Reports* 12, 1 (2022), 14036. https://doi.org/10.1038/s41598-022-18015-z

[24] Qiyue Wang, Yan Fu, Baiyu Shao, Le Chang, Kang Ren, Zhonglue Chen, and Yun Ling. 2022. Early detection of Parkinson's disease from multiple signal speech: Based on Mandarin language dataset. *Frontiers in Aging Neuroscience* 14 (2022). https://doi.org/10.3389/fnagi.2022.1036588