

Federated Learning for Tabular Data using TabNet: A Vehicular Use-Case

William Lindskog

Corporate Research and Development
DENSO Automotive Deutschland GmbH
Munich, Germany
w.lindskog@eu.denso.com

Christian Prehofer

Corporate Research and Development
DENSO Automotive Deutschland GmbH
Munich, Germany
c.prehofer@eu.denso.com

arXiv:2405.02060v1 [cs.LG] 3 May 2024

Abstract—In this paper, we show how Federated Learning (FL) can be applied to vehicular use-cases in which we seek to classify obstacles, irregularities and pavement types on roads. Our proposed framework utilizes FL and TabNet, a state-of-the-art neural network for tabular data. We are the first to demonstrate how TabNet can be integrated with FL. Moreover, we achieve a maximum test accuracy of 93.6%. Finally, we reason why FL is a suitable concept for this data set.

Index Terms—Federated Learning, Feature Extraction, TabNet, Time Series Classification

I. INTRODUCTION

Federated Learning (FL) is a collaborative machine learning concept which advocates local computing and model transmission. Instead of sending raw data to central (or partially central) servers for computing, FL ensures that data are kept on the edge nodes and can reduce overhead communication. Google developed FL as a countermeasure to regulators’ incentives to increase protection of consumer data [1]. Since then, research has mainly focused on the theoretical foundations of the concept, its optimization algorithms, and comparing it to conventional machine learning on standard data sets.

FL as a technique is starting to gain traction in various industries e.g. healthcare [2], [3], internet-of-things (IoT) [4]–[7], and lately Intelligent Connected Vehicles (ICVs) [8], [9], see Figure 1 for conventional vehicular FL scheme. FL for ICVs has seen great improvement in its system design where researchers propose peer-to-peer networks [10], vehicle-to-everything communication [11], and network optimization with asynchronous update schemes [12]. Research has also identified FL as a suitable concept for vehicular networks and fleet management [9], [13]. Despite the extensive work being conducted in FL research, it lacks real-world implementations. Applied FL is a step in the direction of protecting end-user data. By collaboratively training a shared model, many cooperating ICVs can encounter more events than a single vehicle, and potentially train a superior model.

Research that seeks to apply FL to vehicles generally centers around image data. A vehicle produces a plethora of data, much of which can be seen as time series data e.g. acceleration, velocity, engine temperature. This data can be

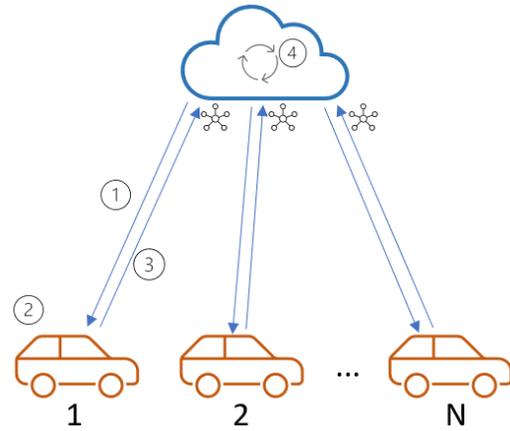


Fig. 1. A FL vehicular scheme: (1) Server sends out initial model, (2) vehicles train model on local data, (3) send back model updates, and (4) server aggregates model updates. This concludes a FL round.

used to improve consumer comfort, and to increase safety by applying accurate fault diagnosis and predictive maintenance. In IoT, we see that research is heavily invested in time series analysis, where FL is now beginning to be integrated [7], [14]. Applying feature extraction on time series data has proven to be advantageous for time series classification, resulting in work with tabular data [15], see Figure 2 for example.

id	x_feature_1	...	x_feature_N	y_feature_1	...	y_feature_N
A
B

Fig. 2. Example list of features in tabular form [16]

In search for suitable algorithms for tabular data, we find that Google’s recent TabNet algorithm [17] can be superior to other state-of-the-art methods e.g. Random Forests and XGBoost (which already have FL implementations). [17] show that TabNet outperforms the other algorithms in regression and classification tasks. However, TabNet is yet to be combined with FL, and we demonstrate how this can be done. In our work, we therefore construct a framework, in which both TabNet and FL can be used together. We apply a feature

extraction pre-processing step on time series to work with tabular data. We limit our work to TabNet and FL, and compare it with other studies. Our direct contributions to research are: (1) Integrating TabNet with FL, (2) showing that TabNet and FL can achieve similar performance to state-of-the-art methods, and (3) a conceptual framework combining FL and TabNet that can be used for tabular data and converted time series to tabular data.

II. DATA

The data we use, have been collected by [18]. By installing a smartphone inside the vehicle cabin, as in Figure 3, they registered the trilateral acceleration for further analysis of the asphalt/road conditions. The trilateral accelerations are given by an accelerometer, integrated in the smartphone, which continuously collects time series data. The sampling rate that they use is 100Hz.

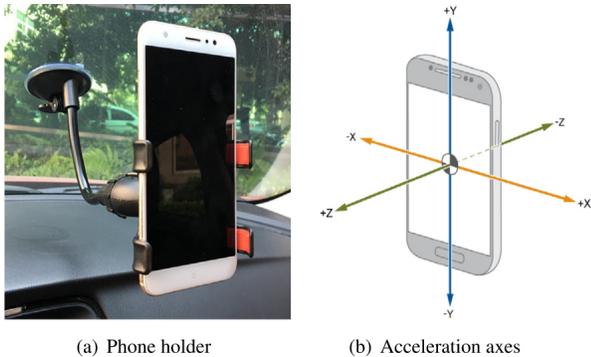


Fig. 3. Example of car phone holder used to collect data and directions considered by the accelerometer sensor [18]

In their paper, they divide and specify three distinct classification tasks for which they allocate independent labelled data sets, which we also consider for our work.

TABLE I
CLASS DISTRIBUTION OF TIME SERIES FOR THE THREE DATA SETS

Data set	Class	Series Length			Examples	Distribution
		Min	Max	Mean		
Regularity	Regular	66	2371	238	762	50.73%
	Deteriorated	190	4201	534	740	49.27%
Pavement Type	Flexible	66	2371	246	816	38.65%
	Cobblestone	284	1543	518	527	24.97%
	Dirt Road	274	1045	484	768	36.38%
Obstacles	Speed Bump	178	730	330	212	27.14%
	Vertical Patch	114	279	191	222	28.43%
	Raised Markers	111	462	256	187	23.94%
	Raised Crosswalk	258	736	457	160	20.49%

The first data set is the *Asphalt Regularity* data set. It considers two classes: (1) Regular pavement where driver comfort is high and (2) deteriorated where they observe irregularities and roughness in the road. Second, the *Pavement Type* data set. [18] identifies three pavement type classes: (1) Flexible pavement, (2) cobblestones, and (3) dirt roads. Flexible pavements are defined as a mixture of asphalt and

bituminous material. The final data set, *Asphalt Obstacles*, includes common obstacles on roads with specific characteristics, see Figure 4. In their study, they identify a great variability for examples of the same class (width, height, size, and material).

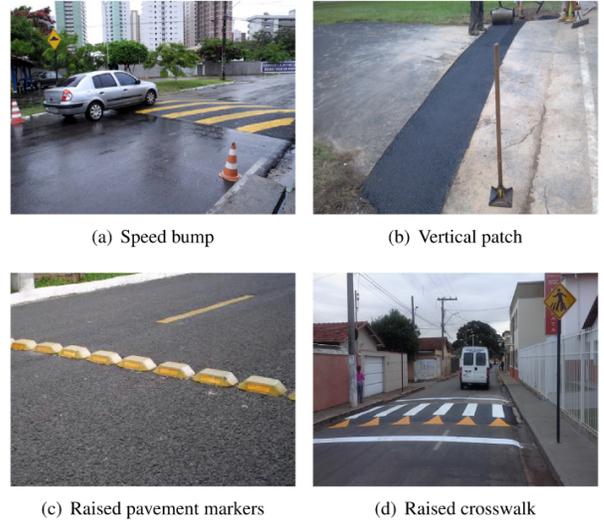


Fig. 4. Photos that illustrate the four class labels of the Asphalt-Obstacles data set [18]

III. METHOD

Initially we manually extract features from all time series using the work from [16]. Thereafter, the high level architecture of our processing scheme can be described shortly as:

- 1) Distribute copies of initial server model to clients (nodes).
- 2) Train local model on local training set.
- 3) Add new instances to local training set from local pool of instances to reflect real-world.
- 4) Validate local model on local validation set.
- 5) Transmit model updates to server.
- 6) Server aggregates model updates and evaluates on test set.

On the server-side, we therefore keep one test set which is used to evaluate the aggregated server model after each communication round, where a communication round is illustrated in Figure 1. On each client, we keep three local data sets:

- 1) (Local) Training set, grows in size over elapsed processing time. The reason why it grows is because a real-world operating vehicle would experience more situations as it goes.
- 2) (Local) Pool set from which we draw random samples from and add to local training set, if pool instances still exist for respective client.
- 3) Validation set, for tuning hyperparameters in TabNet.

We randomly select a small arbitrary number of instances and add to add them to the local training set. We continue this process until there are no more instances in the local

pool set. The FL environment is constructed using Flower [19]. It is an end-to-end FL framework, designed to enable a seamless transition from experimental research to real edge devices. We set up the FL environment on one machine using Flower’s built-in Virtual Client Engine (VCE) that enables the virtualization of Flower Clients to maximise utilization of the available hardware. Moreover, we choose to use FedAvg [1] as FL optimization algorithm.

A. TabNet

TabNet is based on the idea of representing a deep neural network as decision trees (DTs). DTs (and other similar algorithms e.g. XGBoost [20] and Random Forests) are known to perform well on tabular data. It uses tabular input as data for classification- or self-supervised tasks and applies sparse features selection which can enable interpretability and better learning [21]. The use of sparse feature selection includes processing a subset of features in step-wise order, and later aggregating information from subset feature processing. The selected features are linearly transformed and together with additional bias they form TabNet’s decision boundaries. TabNet is trained using gradient decent-based optimization and it incorporates an encoder-decoder architecture [17].

IV. EVALUATION AND RESULTS

For our evaluation, we use an Alienware, Ubuntu 21.04, 16core CPU, Intel Core i7-10870H processor at 2.20GHz clock speed, and a 16GB NVIDIA GeForce RTX 3080 GPU.

When configuring TabNet’s hyperparameters, we follow the guidelines in their article [17]. Nevertheless, we recognize that they work with larger data sets than ours. Therefore, our base settings differ from what they advocate, mainly in terms of width of decision prediction layer and attentive embedding. We use a value of five for both of these values, while [17] recommend a value between 8-64. Our choice of lower values than recommended is motivated by initial attempts to tune TabNet’s hyperparameters. For our data sets, it seems that a slight increase in the width of decision prediction layer and attentive embedding leads to overfitting the model. Everything else we leave as default setting, presented in their ablation study. We choose to evaluate the federated set-up using 2-3 clients. When adding more than three clients to the FL set-up, our machine sometimes runs out of GPU memory.

Section IV is divided into three parts, one for each data set. For each data set, we present test accuracy and test cross-entropy loss over 100 communication rounds. The accuracy and loss is computed on the server-side after one communication round and aggregation of model updates. Furthermore, we also include confusion matrices for the data sets with more than two classes. In a two-client set-up, we wait for each client to send its updates to the server. Using three clients, we sample two out of three to send their updates. Lastly, we split each data set into four subsets:

- 1) (Training set) 10% of total data, equally distributed to clients. Each client’s training set increase with 10 instances per communication round.
- 2) (Pool set) 60% of total data, equally distributed to clients. Each client’s pool set decreases with 10 instances per communication round.
- 3) (Validation set) 10% of total data, equally distributed to clients.
- 4) (Test data set) 20% of total data, kept on the server for evaluation.

A. Asphalt Regularity

The first data set is a balanced data set with two classes (see Table I). The server sends out an initial models to each client. The clients sample random instances from their respective pool set and adds them to the training set for the next local training round. Eventually the clients run out of samples from the pool set but the training continues. This is also true for the other two data sets.

We illustrate the test accuracy results and cross entropy loss in Figure 5. The maximum test accuracy reached is 93.6% and we achieve this with two clients. Adding an additional client reduces maximum test accuracy slightly to 91.2%. The test accuracy is greatly varying until approximately communication round 30. Thereafter, it seems to increase linearly until communication round 70-80 and later converges towards maximum test accuracy.

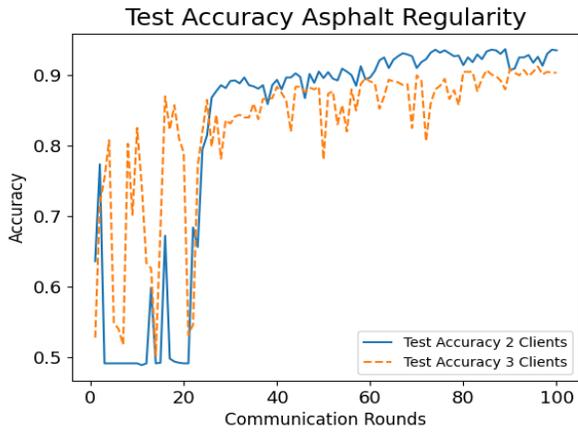
B. Pavement Type

The second data set is less balanced than the first one. There number of time series stemming from cobblestones are significantly less than the other two classes: (1) flexible and (2) dirt road. For this data set, we also include a confusion matrix to better illustrate correct and incorrect predictions. The confusion matrix shows the predictions the server in a two-client set-up since we achieve maximum test accuracy with two clients. Maximum test accuracy of 86.7% is achieved, again with two clients. For three clients, maximum test accuracy is 83.6%, and this is shown in Figure 6.

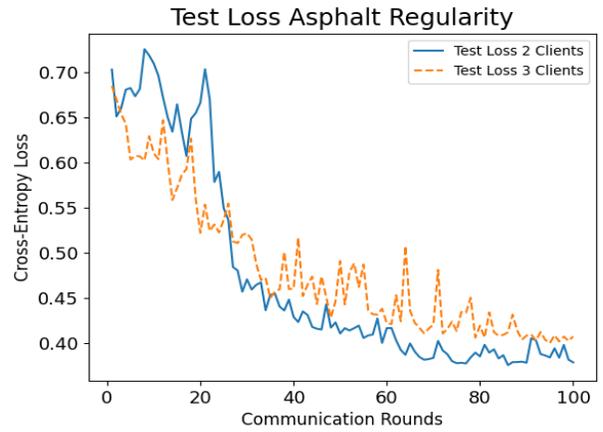
The confusion matrix in Figure 8 shows that data representing travel over *cobblestones* are hardest for TabNet to predict, incorrectly classifying 20.42% of all examples. This can be compared to 15.64% for *dirt roads* and 6.08% for *flexible roads*. Lastly, TabNet finds distinguishing cobblestone from flexible pavement, and vice versa, the easiest. It only misclassifies these examples as the counterpart in 1.05% and 1.69% of the time.

C. Asphalt Obstacles

The asphalt obstacles and final data set includes four classes. It is less balanced than the asphalt regularity data set but more balanced than the pavement type data set (see Table I). Again, we plot a confusion matrix, using the predictions from one client in a two-client set-up. We also illustrate the test accuracy and cross-entropy loss of the given number of clients in Figure 7. Maximum test accuracy of 68.0% is reached with two clients. For three clients, maximum test accuracy is 64.5%. The confusion matrix (Figure 9) shows that TabNet finds raised markers hardest to predict, followed by

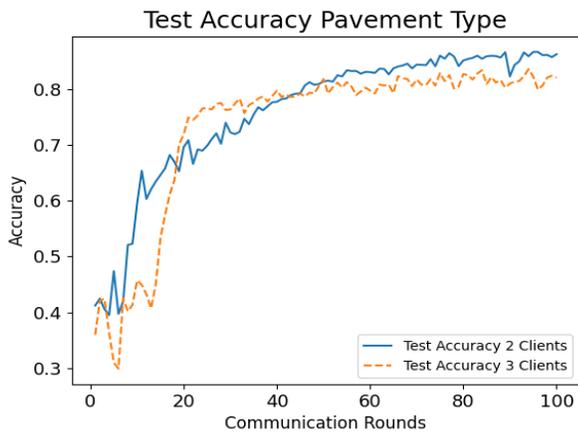


(a) Test accuracy for asphalt regularity data set.

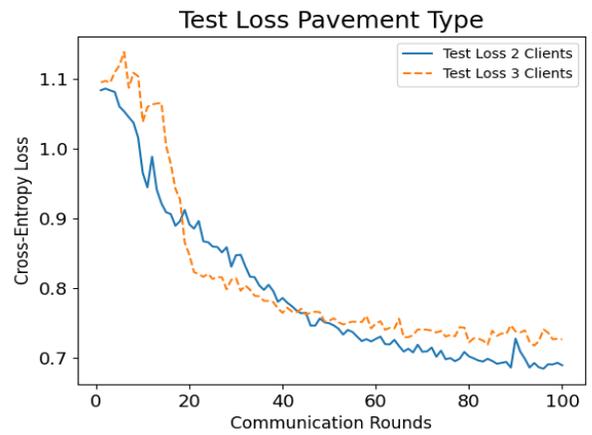


(b) Test loss for asphalt regularity data set.

Fig. 5. Performance of asphalt regularity data set.

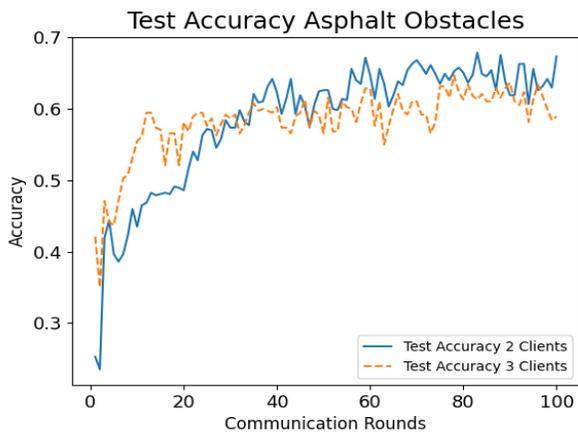


(a) Test accuracy for pavement type data set.

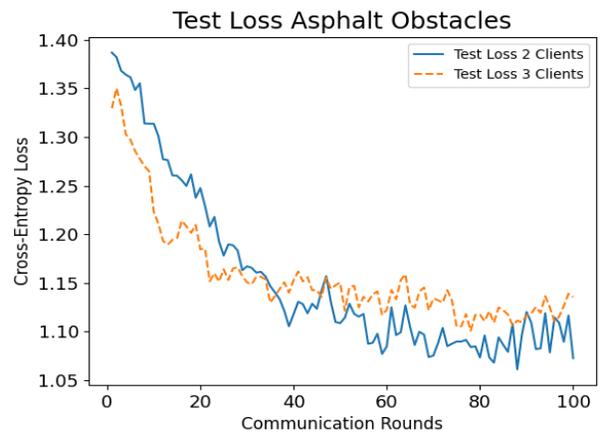


(b) Test loss for pavement type data set.

Fig. 6. Performance of pavement type data set.



(a) Test accuracy for asphalt obstacles data set.



(b) Test loss for asphalt obstacles data set.

Fig. 7. Performance of asphalt obstacles data set.

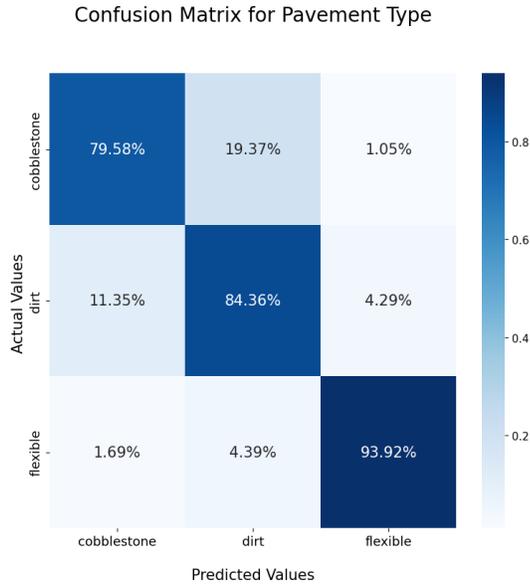


Fig. 8. Confusion matrix for pavement type data set.

raised crosswalk. Moreover, TabNet mistakes raised markers for vertical patches in almost a third of the examples, raised crosswalks for speed bumps in a fourth of the examples, and speed bumps for raised crosswalks in a fifth of the examples. At the same time, TabNet is 85.19% accurate when predicting vertical patches, only mistaking them for raised crosswalks and speed bumps in 3.70% of the time, and for raised markers in 7.41% of the time.

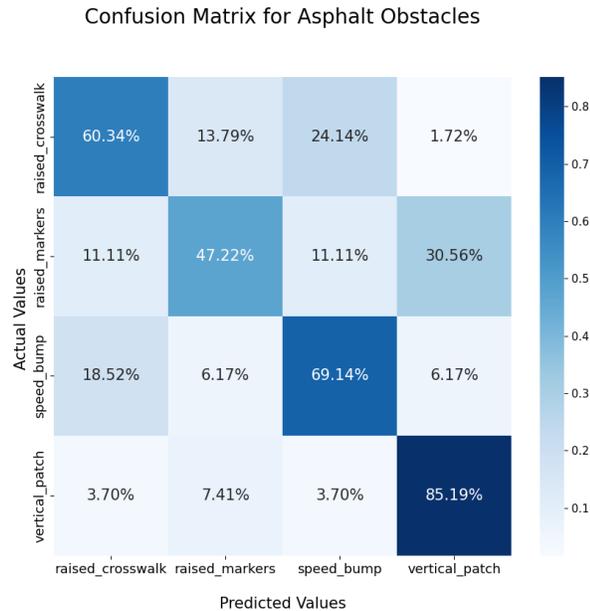


Fig. 9. Confusion matrix for asphalt obstacles data set.

The first observation we make is that TabNet performs well together with FL for all data sets. We clearly identify a convergence of performance as the local training sets grow, and as we communicate back and forth between the server and clients. Furthermore, we present a method using feature extraction and another, combining it with FL and TabNet. The importance of features could also serve as an interesting study. The feature extraction step in our framework is suitable for time series data that shall be converted to tabular data. This step can be skipped if the original data set consists of tabular data.

For pavement type and asphalt regularity data sets, we find that the results are positively surprising. We argue that the large variation for test accuracy in asphalt regularity data set stems from that the training sets are relatively small in the beginning. After approximately 40 rounds, when the pool set is almost empty, it starts to straighten out. In the pavement type data set, we see that TabNet combined with FL finds it most difficult to predict cobblestones and dirt roads. Intuitively, this is reasonable as both roads can be seen as quite "shaky". This is also in line with the results from [18].

We find that the maximum test accuracy for the asphalt obstacles data set is significantly lower than the results presented in [18]. This could be due to sub-optimal choice of TabNet hyperparameters. The ablation study conducted by [17] specifies values for hyperparameters for larger data sets > 10k instances. Our choices could therefore be sub-optimal and can be studied further. Moreover, the data set is quite small (only 781 time series). FL generally requires quite large data sets, and we see that for the first two larger data sets we perform relatively similar to other studies [18], [22]. Additionally, the two classes that were hardest to predict (raised markers and raised crosswalk) were the classes with the least instances in the data set. In this case, there can be a need to handcraft a few distinct features that can support TabNet in its classification and/or conduct a feature analysis study. Between raised markers and vertical patches there seems to be some strong correlation between feature significance. They are both each other's class which is misclassified the most. In more than 30% of the cases, TabNet mistakes raised markers for vertical patches. Studying Figure 4, one could argue that raised markers might share most characteristics with vertical patches e.g. height. Nevertheless, we acknowledge that TabNet can predict vertical patches well (85.19%).

The data sets are interesting to us, as we reason that ICVs should make optimal decisions about what road to take for an arbitrary trip that can maximize user comfort. Many co-operating ICVs can quickly learn characteristics of roads and additional anomalies (obstacles). This information is arguably useful for ICVs as they can choose to avoid uncomfortable roads, and potentially reduce trip duration. Additionally, by applying FL, we can reduce overhead communication and increase privacy by not transmitting trip data that can be used to derive the location of an operating vehicle.

VI. RELATED WORK

In the original study, [18] propose a combination of distance measures between time series to help classify them using a one-nearest-neighbor classifier. They successfully apply it in a one-vehicle setting but do not consider that vehicle can operate together as many ICVs. Further studies have come to show improved performance on the same data sets, with tweaked or novel algorithms, but similarly neglect the integration of FL [22], [23].

In [24], they emphasize the importance of principal component analysis (PCA) and feature extraction from time series. Nevertheless, they engineer the features themselves which is a time-consuming task. Our framework proposes extracting a large set of statistical features using the package presented in [16]. [15] showed how the work from [16] can be included in time series classification for predictive maintenance in IoT. They also included an active learning sampling strategy which increased the convergence speed. However, they did only consider centralized machine learning applications and not FL. In another work, related to vehicle operations, [25] showed how a two-scale regression model could be integrated with FL to predict energy demand for an arbitrary route of a vehicle. They even demonstrated how personalization added to the model by clustering groups of drivers. Nevertheless, they do handcraft the features, that the algorithm requires, themselves and their data set is quite small.

There is much research on how to integrate deep neural networks into DTs. Neural DTs [26], [27] use differentiable decision functions, instead of non-differentiable axis-aligned splits. Neural DTs do however not include automatic feature selection which can lead to worse performance. In their study, [28] demonstrated how Neural DTs can be combined with FL for the MNIST data set. In collaboration with Volvo, [12] extended the work on FL using Neural DTs and showed how it can be applied in a vehicular setting for BDD100k data set. Interestingly, their algorithm outperforms centralized learning. The data are nevertheless images and videos and therefore they disregard time series data.

Similarly to deep neural DTs, researchers have also developed federated algorithms of random forests [29], conventional DTs [30], SVM [31], and XGBoost [32]. The non-federated implementations for these algorithms have nevertheless been shown to perform worse than TabNet [17] on large data sets, which is why we seek to develop our own implementation of federated TabNet. To our understanding, there is no work comparing these FL algorithms on standardized data sets, which therefore would be an appropriate study in the future. [33] presents existing work on FL and mentions tree structure implementations for FL. Nevertheless, no superior algorithm is concluded, they only mention the work of [34].

VII. CONCLUSION

TabNet has shown superior performance compared to established models, but combination with FL has not been considered. To our understanding, we are the first to analyze TabNet with FL for tabular data. Moreover, we demonstrate how FL

can be applied to the data sets from [18]. Our framework, using the combination of TabNet and FL, is successfully applied to the data sets and shows promising results. We include a feature extraction step in our framework to convert the time series data to tabular data. Furthermore, we acknowledge a slight decrease in test accuracy compared to other relevant studies [18], [22]. Our test accuracies vary between 68.0 – 93.6% and the most difficult data set is the *asphalt obstacles* data set. Nevertheless, in the *asphalt obstacles* data set, we observe that for vertical patches, TabNet classifies the instances 85.19% correctly. FL does not rely on raw data transmission and supports the increasing demand for privacy preserving operations for edge devices. We also argue for why FL applied to the used data sets is important for ICVs.

In the future, we shall tune the hyperparameters for the smaller data sets at hand and evaluate different FL optimization algorithms performance on similar data sets. Furthermore, we shall investigate the combination of TabNet and FL on other benchmark data sets and extend our work to include regression tasks and not only use it for classification. In such a work, we aim to compare federated TabNet with current state-of-the-art federated algorithms that can be utilized for tabular data. Lastly, we shall compare TabNet and FL with a centralized approach and study the difference in terms of overall performance and communication efforts.

ACKNOWLEDGEMENT

This work was partly supported by the TRANSACT project. TRANSACT (<https://transact-ecsel.eu/>) has received funding from the Electronic Component Systems for European Leadership Joint Under-taking under grant agreement no. 101007260. This joint undertaking receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Belgium, Denmark, Finland, Germany, Poland, Netherlands, Norway, and Spain.

REFERENCES

- [1] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.
- [2] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein *et al.*, "The future of digital health with federated learning," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–7, 2020.
- [3] B. Pfitzner, N. Steckhan, and B. Arnrich, "Federated learning in a medical context: A systematic literature review," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 2, pp. 1–31, 2021.
- [4] L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong, "Federated learning for internet of things: Recent advances, taxonomy, and open challenges," *IEEE Communications Surveys & Tutorials*, 2021.
- [5] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for internet of things: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, 2021.
- [6] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A survey on federated learning for resource-constrained iot devices," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 1–24, 2021.
- [7] I. Kholod, E. Yanaki, D. Fomichev, E. Shalugin, E. Novikova, E. Filippov, and M. Nordlund, "Open-source federated learning frameworks for iot: A comparative review and analysis," *Sensors*, vol. 21, no. 1, p. 167, 2020.

- [8] Z. Du, C. Wu, T. Yoshinaga, K.-L. A. Yau, Y. Ji, and J. Li, "Federated learning for vehicular internet of things: Recent advances and open issues," *IEEE Open Journal of the Computer Society*, vol. 1, pp. 45–61, 2020.
- [9] J. Posner, L. Tseng, M. Aloqaily, and Y. Jararweh, "Federated learning in vehicular networks: opportunities and solutions," *IEEE Network*, vol. 35, no. 2, pp. 152–159, 2021.
- [10] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Federated learning for data privacy preservation in vehicular cyber-physical systems," *IEEE Network*, vol. 34, no. 3, pp. 50–56, 2020.
- [11] X. Zhou, W. Liang, J. She, Z. Yan, I. Kevin, and K. Wang, "Two-layer federated learning with heterogeneous model aggregation for 6g supported internet of vehicles," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5308–5317, 2021.
- [12] H. Zhang, J. Bosch, H. H. Olsson, and A. C. Koppisetty, "Af-dndf: Asynchronous federated learning of deep neural decision forests," in *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 2021, pp. 308–315.
- [13] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 2031–2063, 2020.
- [14] Y. Liu, S. Garg, J. Nie, Y. Zhang, Z. Xiong, J. Kang, and M. S. Hossain, "Deep anomaly detection for time-series data in industrial iot: A communication-efficient on-device federated learning approach," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6348–6358, 2020.
- [15] S. M. del Campo Barraza, W. Lindskog, D. Badalotti, O. Liew, and A. Toyser, "Active learning framework for time-series classification of vibration and industrial process data," in *Annual Conference of the PHM Society*, vol. 13, no. 1, 2021.
- [16] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh—a python package)," *Neurocomputing*, vol. 307, pp. 72–77, 2018.
- [17] S. O. Arik and T. Pfister, "Tabnet: Attentive interpretable tabular learning," in *AAAI*, vol. 35, 2021, pp. 6679–6687.
- [18] V. M. Souza, "Asphalt pavement classification using smartphone accelerometer and complexity invariant distance," *Engineering Applications of Artificial Intelligence*, vol. 74, pp. 198–211, 2018.
- [19] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, T. Parcollet, P. P. de Gusmão, and N. D. Lane, "Flower: A friendly federated learning research framework," *arXiv preprint arXiv:2007.14390*, 2020.
- [20] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.
- [21] X. Li, Y. Wang, and R. Ruiz, "A survey on sparse learning models for feature selection," *IEEE Transactions on Cybernetics*, 2020.
- [22] M. Middlehurst, J. Large, and A. Bagnall, "The canonical interval forest (cif) classifier for time series classification," in *2020 IEEE international conference on big data (big data)*. IEEE, 2020, pp. 188–195.
- [23] T. Kannan and H. Hoffmann, "Budget rnns: multi-capacity neural networks to improve in-sensor inference under energy budgets," in *2021 IEEE 27th Real-Time and Embedded Technology and Applications Symposium (RTAS)*. IEEE, 2021, pp. 143–156.
- [24] Y. Wang, I. L. Bennani, X. Liu, M. Sun, and Y. Zhou, "Electricity consumer characteristics identification: A federated learning approach," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3637–3647, 2021.
- [25] A. T. Thorgeirsson, S. Scheubner, S. Fünfgeld, and F. Gauterin, "Probabilistic prediction of energy demand and driving range for electric vehicles with federated learning," *IEEE Open Journal of Vehicular Technology*, vol. 2, pp. 151–161, 2021.
- [26] S. Wang, C. Aggarwal, and H. Liu, "Using a random forest to inspire a neural network and improving on it," in *Proceedings of the 2017 SIAM international conference on data mining*. SIAM, 2017, pp. 1–9.
- [27] P. Kotschieder, M. Fiterau, A. Criminisi, and S. R. Buló, "Deep neural decision forests," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1467–1475.
- [28] A. Sjöberg, E. Gustavsson, A. C. Koppisetty, and M. Jirstrand, "Federated learning of deep neural decision forests," in *International Conference on Machine Learning, Optimization, and Data Science*. Springer, 2019, pp. 700–710.
- [29] Y. Liu, Y. Liu, Z. Liu, Y. Liang, C. Meng, J. Zhang, and Y. Zheng, "Federated forest," *IEEE Transactions on Big Data*, 2020.
- [30] Q. Li, Z. Wen, and B. He, "Practical federated gradient boosting decision trees," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 04, 2020, pp. 4642–4649.
- [31] R.-H. Hsu, Y.-C. Wang, C.-I. Fan, B. Sun, T. Ban, T. Takahashi, T.-W. Wu, and S.-W. Kao, "A privacy-preserving federated learning system for android malware detection based on edge computing," in *2020 15th Asia Joint Conference on Information Security (AsiaJCIS)*. IEEE, 2020, pp. 128–136.
- [32] N. K. Le, Y. Liu, Q. M. Nguyen, Q. Liu, F. Liu, Q. Cai, and S. Hirche, "Fedxgboost: Privacy-preserving xgboost for federated learning," *arXiv preprint arXiv:2106.10662*, 2021.
- [33] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowledge-Based Systems*, vol. 216, p. 106775, 2021.
- [34] K. Cheng, T. Fan, Y. Jin, Y. Liu, T. Chen, D. Papadopoulos, and Q. Yang, "Secureboost: A lossless federated learning framework," *IEEE Intelligent Systems*, vol. 36, no. 6, pp. 87–98, 2021.