algorithms

Article

Insights into Multi-Model Federated Learning: An Advanced Approach for Air Quality Index Forecasting

Duy-Dong Le [1,*], Anh-Khoa Tran [2], Minh-Son Dao [2,*], Kieu-Chinh Nguyen-Ly [1], Hoang-Son Le [1], Xuan-Dao Nguyen-Thi [1], Thanh-Qui Pham [1], Van-Luong Nguyen [1] and Bach-Yen Nguyen-Thi [1]

[1] University of Economics, Ho Chi Minh City 700000, Vietnam
[2] National Institute of Information and Communications Technology, Tokyo 184-8795, Japan
[*] Correspondence: dongld@ueh.edu.vn (D.-D.L.); dao@nict.go.jp (M.-S.D.)

Abstract: The air quality index (AQI) forecast in big cities is an exciting study area in smart cities and healthcare on the Internet of Things. In recent years, a large number of empirical, academic, and review papers using machine learning (ML) for air quality analysis have been published. However, most of those studies focused on traditional centralized processing on a single machine, and there had been few surveys of federated learning (FL) in this field. This overview aims to fill this gap and provide newcomers with a broader perspective to inform future research on this topic, especially for the multi-model approach. In this survey, we went over the works that previous scholars have conducted in AQI forecast both in traditional ML approaches and FL mechanisms. Our objective is to comprehend previous research on AQI prediction including methods, models, data sources, achievements, challenges, and solutions applied in the past. We also convey a new path of using multi-model FL, which has piqued the computer science community's interest recently.

Keywords: multi-model; federated learning; AQI prediction; big data; IoT

check for updates

Citation: Le, D.-D.; Tran, A.-K.; Dao, M.-S.; Nguyen-Ly, K.-C.; Le, H.-S.; Nguyen-Thi, X.-D.; Pham, T.-Q.; Nguyen, V.-L.; Nguyen-Thi, B.-Y. Insights into Multi-Model Federated Learning: An Advanced Approach for Air Quality Index Forecasting. *Algorithms* **2022**, *15*, 434. https://doi.org/10.3390/a15110434

Academic Editors: Danilo Pelusi and Asit Kumar Das

Received: 10 October 2022
Accepted: 8 November 2022
Published: 17 November 2022

1. Traditional Approaches in AQI Prediction

Air is a vital need for all life on Earth's existence and development . It impacts one's health as well as the economy. Air quality heavily depends on natural and anthropomorphic resources such as volcanic eruptions, forest fires, climate change, ozone holes, industrialization, urbanization, and transportation emissions. Many pollutants, such as $SO_2, NO_2, CO_2, NO, CO, NO_x, PM_{2.5}$, and PM_{10} can be found in the atmosphere. A vast number of researches on air pollution forecast and AQI prediction on a global scale focus on pollutants forecasting. Nevertheless, most of them utilized single ML methods. Table 1 lists some of these works.

In Table 1, the first column labeled "Authors" is the references and authors. The second column titled "Methods" contains the models and their citations. In the column names "Model, Accuracy and Data's features" we list some models that had been implemented in the references, their metrics and some data that had been used. In the last column, we collect some conclusions from the review papers.

In general, the methods that have been applied in AQI prediction can be classified as following:

- Kind of models: statistical, artificial neural network (ANN), deep neural network (DNN), hybrid, and ensemble.
- Kind of data: temporal, spatial and spatiotemporal

Table 1. Traditional approaches in AQI prediction.

Authors	Methods	Model, Accuracy Data's Features	Main Conclusions
[1] Bai, L., et al. (2018).	[2]: SARIMA, ANN, FTS [3]: FFMLP, MLP-GA, MLP [4]: MLP [5]: BPNN [6]: PCA–ANN [7]: GFM-NN	[2] ANN MAE: 2.70 MSE: 12.79 RMSE: 3.58 $PM_{10}, O_3, CO_2, SO_2,$ NO_2	- The most popular statistical method uses (AI) models. - The accuracy of ANNs is higher than other statistical models, but they are usually be local optima. - The accuracy of ANNs can be improved by adding data: meteorology, geographic, time-scale, emission pattern of sources. - Some hybrid models can improve the accuracy.
[8] Masood, A. and Ahmad, K. (2021).	ANN FUZZY SVM DNN	[9–14] DNN PM RMSE: 7.27 PM R^2: 0.96 [10,14–16]: DNN O_3 RMSE: 3.51 O_3 R^2: 0.92 [9,10]: DNN CO RMSE: 0.95 CO R^2: 0.69×10^{-5} The above RMSE and R^2 are mean values of the refs	- The most frequently applied input parameter is API - The best performing AI-based model is the DNN. - Fuzzy logic, DNN and SVM are the three commonly used AI-based techniques - DNN, SVM, and Fuzzy techniques showed better accuracy in forecasting PM concentrations - DNN outperformed the other AI techniques for this pollutant category. DNN improved spatial and temporal stability for a multi-step ahead forecasting of pollutants - Ability to exploit high-level features from raw air quality and meteorological data
[17] Baklanov, A., and Zhang, Y. (2020).	[18] Multi-models Ensemble forecasting: A new Ozone Ensemble Forecast System (OEFS), [19] Dynamic Linear Regression for ensemble (DLR), [20] MarcoPolo–Panda	[18] OEFS O_3 RMSE: 16.34 [19] DLR O_3 RMSE: 10.57 [20] MarcoPolo – Panda O_3 RMSE: 32.8 NO_2 RMSE: 21.8 PM_{25} RMSE: 30.2	- Ensemble forecasting has shown significant statistical improvements for both O3 and PM2.5 forecasts over any individual forecast [21]. - Multi-model ensemble air quality forecasting has been emerging for AQF on global and regional scales - Multi-model ensemble results provide a range and an indication of the robustness of the forecasts and help to improve the accuracy of chemical weather and air quality forecasting.
[22] Masih, A., et al. (2019).	Ensemble learning (EL) Classical Regression (CR) NN SVM Lazy	R^2: EL:0.79 CR: 0.74 SVM :0.67 NN: 0.6	- The high accuracies achieved with ML algorithms explains it all why these algorithms are appropriate and should be preferred over traditional approaches - Should use critical pollutants (NO_x and SO_2) ensemble learning techniques to improve model accuracy
[23] Liao, K., et al. (2021).	Multi-layer neural network (NN) Land use regression (LUR) Multiple linear regression (MLR) Hybrid methods	[24] NN $PM_{2.5}$ error: 30%–60% [25] Variety of NN algorithms PM_{10} RMSE: 15.700 PM_{10} MAE: 9.047 PM_{10} R^2: 0.840 [26] DNN $PM_{2.5}$ MAPE: 11.93%	- ANN methods were preferred in studies of PM and O_3 - LUR were more widely used in studies of NO_x - Multi-method hybrid techniques gradually became the most widely used approach between 2010 and 2018 - The most difficult element of research on air pollution prediction is likely to be the interaction between pollutants, which will be based on a mixed technique to forecast numerous contaminants concurrently in the future.
[27] Liao, Q., et al. (2020).	Artificial Neural Network (ANN) Linear and Logistic Regression	SVR shows better performance in prediction of AQI while RFR gives the better performance in predicting the NOx concentration	- Linear and Logistic Regression are the choices of many researchers for the prediction of AQI and air pollutants concentration - The future scope may include consideration of all parameters that is meteorological parameters, air pollutants while predicting AQI or forecasting the future concentration level of different pollutants.

Statistical models have played a massive role in predicting air quality, especially statistical models with the application of AI techniques [1,8,17,22,23,27]. ANNs have been shown to outperform traditional statistical models that do not use AI. Unfortunately, the biggest drawback of this type is that it often falls into local optimal stages. To overcome this drawback, many studies have been used, such as DNN [8,23], hybrid and ensemble [17,22,28,29]. The AQI forecast should assess the implications of air pollution on several sectors, such as health, agriculture, land transportation, aviation, and energy, among others, and issue warnings and recommendations based on various thresholds, risks, and cost functions [30]. Some works concentrate on RNN [31–33] and the Spatial-Temporal Network can deal with the complicated non-linear spatial and temporal correlations [34,35]

Some conclusions of previous studies are very significant, such as: Forecasting tasks often employ neural networks and support vector machine-based algorithms, pollution estimation is typically carried out using ensemble learning and linear regression-based methodologies [36]. Between 2010 and 2018, multi-method hybrid strategies significantly surpassed single-method approaches in terms of usage [37]. Land-use Regression (LUR) models are an important tool for integrating traffic and geographic information to characterize variability in exposures [38,39]. Some reviews are limited in scope within a country or province, such as in India [40], and Hebei of China [41]

Among these studies, DNN is the preferred application direction of many current researchers in the field of AQI prediction. DNNs are considered superior to traditional AI techniques because they are stable, give lower model errors, and respond well to data diversity. A collection of algorithms known as the "ensemble learning technique" trains several models to tackle a single issue by integrating the outcomes that each model produces. Strong and weak predictors are combined in ensemble learning approaches, which are more generalizable and less susceptible to overfitting. Overall, these methods outperform single-base learning methods such as ANNs and SVMs.

2. Federated Learning in AQI Prediction

FL is training a ML model with cooperation between many participants (terminals or organizations, hereinafter referred to as clients) participating simultaneously without sharing data. At each training round, clients received global model parameters and perform local training with their data. Then clients return model parameters to aggregate a global model according to the Formula (1) [42,43].

$$min\ F(w), where\ F(w) := \sum_{k=1}^{m} (p_k F_k(w)) \tag{1}$$

$$and\ F_k(w) = \frac{1}{n_k} \sum_{j_k=1}^{n_k} (w; x_{j_k}, y_{j_k}) \tag{2}$$

where:

m: number of participants
$p_k \geq 0$ and $\sum_k pk = 1$
F_k: Local optimization function on participant k in Formula (2)
n_k: number of data samples

When the clients are terminal devices (mobiles, computers, Unmanned Aerial Vehicles or UAVs for short . . .) we call this relationship cross-devices, and cross-silos when the clients are organizations (hospitals, banks, schools . . .) [44]. The communication between clients is established in the following architecture: centralized, decentralized and hierarchy [45]. In addition, this architecture can also be classified according to the distributed data design of the system, including: Vertical FL, Horizontal FL and Transfer FL [46].

Since 2020, a few authors have investigated the use of FL in AQI prediction [47–51] or data missing prepossessing [52] for this task. Most authors agreed that FL outperformed the traditional ML methods in their research. Table 2 gives an overview of these works.

Table 2. Federated learning methods in AQI forecast.

Authors	Methods	Data Processing	Evaluation Metrics	Main Conclusions
[47] P. Chhikara, et al.	Centralized FL (FedAvg) with LSTM in each UAVs. Compare with other traditional training ML: SVM, KNN, Decision tree, ANN	- AQI from Dehli Inida (2015 to 2020) with 2009 timesteps - Train: 0.75 - Test: 0.25	- RMSE: 56.222 - MAE: 41.219 - MAPE: 24.184	- Traditional ML models miss the temporal dependencies between the data - The proposed model predicts the future AQI value with a minor error compared to other ML models. - The outcomes illustrate the efficacy of the proposed scheme to predict the AQI of a given area
[48] P. Chhikara, et al.	- FedAvg with CNN-LSTM in each UAVs. Compare with: RNN, GRU, Vanilla LSTM, Stacked LSTM, Bidi LSTM. - POS to find hazardous zone	Same as [47] Link to data [a]	- RMSE:221.682 - MAE: 200.668 - MAPE: 1217.897	- The proposed model is better than the others in term of the given erea - Model compression should be applied to reduce it's size - Need to find communication efficient FL frameworks for long term UAV monitoring.
[49] Yi Liu, et al.	- Air: Light-weight FL-based UAV (FedAvg) - Ground: graph convolution neural network combines spatio-temporal model (LSTM) - Compare with 2D CNN, 3D CNN, AQNet, SVM	Air: 5298 haze images: - Train: 0.8 - Test: 0.2 Ground: 6 months 2019 AQI from China - Train: 5/6 - Test: 1/6 Link to data [b]	RMSE: - Real-time: 3.212 - After 2 h: 4.589 - After 4 h: 6.357 - After 6 h: 9.145	- The proposed method not only realizes high-precision AQI monitoring, but also reduces UAV energy consumption. - Need to design some novel model compression techniques to deploy large-scale and complex DNN to UAVs -Need to solve the expensive communication cost for UAV to achieve long term monitoring
[50] Karisma Trinanda Putra	Federated Compressed Learning (FCL): - Compressed Sensing - FedAvg with LSTM Compare with Centralized Learning (CL)	During 9/2020 by 1000 sensors across Taiwan, and 4 sensors of the prototype with 5 features: PM1.0, PM2.5, PM10, temperature, and humidity	RMSE: - FCL: 5.044 - CL: 4.480	- The data consumption is reduced by more than 95%, error rate below 5% - The FCL will generates slightly lower accuracy compared with centralized training - The data could be heavily compacted and securely transmitted in WSNs
[51] Do-Van Nguyen, et al.	- FL CRNN Model, compare with Auxiliary CNN - Proposed spatial averaging aggregation function of federated learning paradigm	Kanto region, Japan, from 2018 to 2021, 15 dimentions Link to data[c]	See Figures 8–13 in [51]	- FL CRNN models can capture spatial-temporal local information and be able to share knowledge among participating cities - Can transfer knowledge to newly added participants - Epochs should be large enough to fully capture knowledge from each local side
[53] Huang, G., et al.	A new cross-domain prediction model FL: SSA-LSTM, FL-DPLA-SSA-LSTM.	Hourly air pollutants and meteorological data from 12 cities in the Fenhe River and Weihe River Plains in China in 2020		The prediction performance of the proposed model is significantly better than all comparison models
[54] Chen, Y., et al.	Asynchronous Online FL (ASO-Fed).	Air pollutants collected from multiple weather sensor devices distributed in 9 locations of Beijing with features such as thermometer and barometer	MAE: 36.71 SMAPE: 0.42	ASO-Fed has lower SMAPE errors than all other models as the dropout rate increases and the performance of ASO-Fed is relatively stable. However, as expected, if one of the nodes never sends updates to the central server, the model does not generalize. This explains the poor performance as the dropout rate increases.

[a] https://www.kaggle.com/rohanrao/air-quality-data-in-india (accessed on 6 November 2022); [b] http://www.weather.com.cn/air (accessed on 6 November 2022); [c] http://soramame.taiki.go.jp (accessed on 6 November 2022);

Extending LSTM model in [47] to CNN-LSTM, P. Chhikara, et al. [48] developed a PSO-based method that uses UAVs swarm intelligence to identify the city's most dangerous zone. A plan for using UAV swarms to detect the quality of the air in the sky is put forth.

The suggested solution uses FL to ensure privacy while monitoring and forecasting the AQI in great detail. Every day, a city calculates and stores the AQI value. They can forecast the AQI index for that city for the next 'n' days based on the collected data. The FL idea is used to install the whole suggested system in a decentralized network. The suggested CNN-LSTM model predicts future AQI index with relatively low error, which cannot be conducted by the current single model (aggregated with all UAVs' combined weights in the swarm). The suggested CNN-LSTM model is compared against the other five models using different error measures. The outcome demonstrates that the suggested method, which combines FL with deep learning, is more accurate and one of the pioneering attempts in forecasting the air quality of a specific location.

Moreover, using UAVs liked [47,48], but Yi Liu, et al. [49] suggests a novel framework for fine-grained 3D air quality monitoring and forecasting based on FL for aerial-ground air quality sensing. This system specifically uses a lightweight Dense-MobileNet model for airborne data to perform energy-efficient end-to-end learning from haze characteristics of hazy photos captured by UAVs for forecasting AQI scale distribution. Additionally, the FL Framework not only enlarges the extent of UAV swarms' monitoring but also enables many organizations or institutions to cooperatively build a well-trained global model to monitor AQI without sacrificing privacy. They suggest a Graph Convolutional Neural Network-based Long Short-Term Memory (GC-LSTM) model for ground sensing systems to accomplish precise, immediate, and future AQI inference. The aerial-ground sensing system can infer the AQI accurately thanks to the GC-LSTM model, which makes use of the topological structure of the ground monitoring station to capture the spatiotemporal correlation of historical observation data. Numerical findings demonstrate that the suggested framework may enable accurate and energy-efficient AQI sensing without compromising the privacy of raw data through comprehensive case studies on a real-world dataset.

In the use of smart city sensing, Karisma Trinanda Putra, et al. [50] introduces a unique edge computing framework called Federated Compressed Learning (FCL), which enables fast data creation while protecting data privacy for PM2.5 forecasts. The suggested method incorporates safe data transmission while inheriting the fundamental concepts of the compression methodology and regional cooperative learning. As a result, it could decrease the volume of data while maintaining data privacy. By utilizing the FCL edge computing architecture, this study seeks to construct a wireless sensor network system that is powered by green energy. It is also one of the fundamental technologies for the deployment of reconfigurable and customizable sensing devices. As a result, prototypes are created to verify the performance of the suggested framework. The outcomes demonstrate a reduction in data consumption of more than 95% with an error rate of less than 5%. Finally, compared to centralized training, the prediction results based on the FCL will produce somewhat lower accuracy. However, the information might be safely sent and tightly compressed in WSNs.

Meanwhile, Do-Van Nguyen, et al. [51] provide a framework for distributed learning that supports cross-spatial training among participants from cities and prefectures, for example. Convolutional Recurrent Neural Networks (CRNN) are trained locally in each area with the goal of predicting the local Oxidant alert level, while an aggregated global model improves the distilled information from every part of a region. The study shows that although adaptive structure at the predictive component of the deep neural network model may capture various environmental monitoring stations' configurations in local regions, planned common elements of the CRNN can be fused worldwide. In order to enhance the accuracy of the whole FL system, some experiment findings also point to strategies for maintaining a balance between local deep neural network training epochs and synchronous training rounds for FL. The findings also demonstrate that by transferring a shared global model, additional participating locations may train and rapidly achieve optimal local models.

For the first time, a conditional GAN imputation technique is proposed in this research inside a FL framework by Zhou, X., et al. [52] to resolve data sets from various data owners without sharing them. Additionally, they use the Wasserstein distance and "Hint mask" method to enhance the performance of the standard conditional GAN. The outcomes of the experiments demonstrate that their GAN-based imputation techniques may deliver the best results. Additionally, their federated GAN imputation technique performs better than the GAN imputation method that was locally trained for each participant, indicating the viability of their imputation model. By facilitating private multi-institutional partnerships, their suggested federated GAN technique can enhance access to air quality data, hence improving model quality. They further explore how the geographic distribution of data across cooperating participants affects model quality, and, surprisingly, they discover that the GAN training process using a FL framework performs more consistently.

A review from Neo, E.X., et al. [55] imposes a framework on the basis of data from environmental, health, and AQI prediction for integrated environmental and health impact assessment systems, as well as big data learning and prediction abilities of AI and FL. In the framework of early health care service preparedness prediction and hospital administration, it is also crucial for efficient medical services.

Finally, Chen, Y., et al. [54] provide an Asynchronous Online FL (ASO-Fed) system, in which the edge devices carry out online learning with continuous local data streaming and a centralized server collects model parameters from clients. To address the issues posed by heterogeneous edge devices with variable computing demands as well as lagging or failing edge devices, their approach updates the central model in an asynchronous fashion. On three actual non-IID streaming datasets from the real world and a benchmark picture dataset created in simulation, they conduct comprehensive tests. The outcomes demonstrate the value of the model on rapidly convergent and good prediction performance.

Most of the FL architectures usually used in Table 2 are centralized FL with deep learning models implemented in the clients, which will be discuss in the next section. These studies also demonstrate the trend of applying UAVs with onboard sensors [47,48] and cameras [49] or smart sensing [51].

3. Insights into Federated Learning

Many surveys of FL have been conducted worldwide. They often focus on: architectures, algorithms, optimization, system, benchmarks, challenges..., and future research [43–45,56–61]. Some surveys specialize in each area such as: , Healthcare [45], Internet of Thing [56], Mobile Edge Network [58], Blockchain [59], Smart-Cities [60], and Securities [61]. The following are some highlights of them:

3.1. FL Architectures and Algorithms

The architectures of FL are usually divided into three types: centralized, decentralized, and hierarchical. Figure 1 presents the usual topology FL applied in these studies.

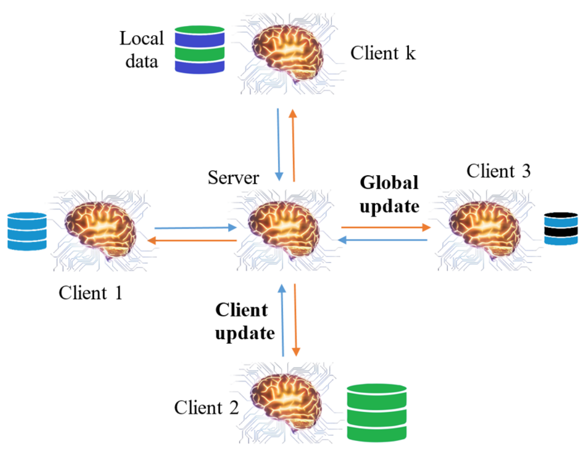

Figure 1. Centralized FL architecture [43,56,57].

As shown in Figure 1, a centralized FL system consists of a central server connecting with several clients in a star network. In the training round, some selected clients take part in the simultaneous training of a model using their local data, batches, epochs, and learning rate. Then, updated clients are then sent to the central server by the participated clients. The server aggregates a global model, and the global update is forwarded to the clients for the next training cycle. The training process is ended after several rounds.

The server is regarded as the essential element of the network for coordinating the distribution of client updates to global aggregation while maintaining the confidentiality of the clients' training data. It aggregates a global model by averaging all the participated model updates. It is called Federated Averaging or FedAvg for short, which was first introduced by McMahan, et al. in 2016 [42]. The models used in the works listed in Table 2 are usually deep learning ones: Long Short-Term Memory (LSTM) [47,50], Convolutional Recurrent Neural Networks (CRNN) [48,49,51], and Generative Adversarial Nets (GAN) [52]. The pseudo-code of the whole training process is given in Algorithm 1 [42].

Algorithm 1 FedAvg centralized

ServerSite: ▷ Run on Server
initialize model parameters W_0
for each training round t **do**
  random select clients
  for each client k **in parallel do**
    $W^k_{t+1} \leftarrow$ **ClientSite**(k, W_t) ▷ Client update
  end for
  $W_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} W^k_{t+1}$ ▷ Global update
end for

ClientSite(k, W_t): ▷ Run on Clients
  Fit local model with it dataset, batch size, epochs, and learning rate to get new model update W^k_{t+1}
Return (W^k_{t+1}) **to server**

Centralized FL architecture faces the main drawback when the number of clients is enormous [44]. This drawback is called a bottleneck because the vast communications link to the server leads to overload. There is another architecture without a server called a decentralized system, shown in Figure 2.

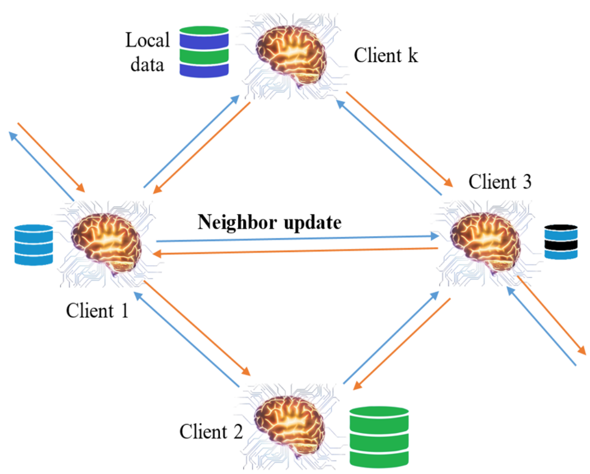

Figure 2. Decentralized FL [43,56,57].

Decentralized FL is a network architecture with no server to manage the training process. As seen in Figure 2, all clients participated in a peer-to-peer (P2P) topology to carry out a training model. The clients also carry out local training in each communication round using their data set. Then, to reach an agreement on the global update, each client executes a model aggregation utilizing the model updates received from nearby clients through P2P communication. Decentralized FL may be made to entirely or partially rely on how it is used (see FL graph in [62]). The whole training process can be written in pseudocode as in (Algorithm 2).

Algorithm 2 FedAvg Decentralized

```
each node initialize model parameters W_0
for each training round t do
    for each node i in paralleled do
        W^i_{t+1} ← ClientSite(i, W_t)
    end for
    for each neighbour k of node i do
        W^i_{t+1} ← Σ_{k=1}^{K} (n_k/n) W^{ik}_{t+1}        ▷ Exchange update
    end for
end for

ClientSite(i, W_t):
   Fit local model with it dataset, batch size, epochs, and
   learning rate to get new model update W^i_{t+1}
Return (W^i_{t+1})
```

Another topology is the hybrid of centralized and decentralized design known as hierarchy [45], which has been developed as shown in Figure 3. This topology's goal is to work with many clients, which sometimes is a bottleneck of centralized FL, and to lower communication expenses in decentralized FL.

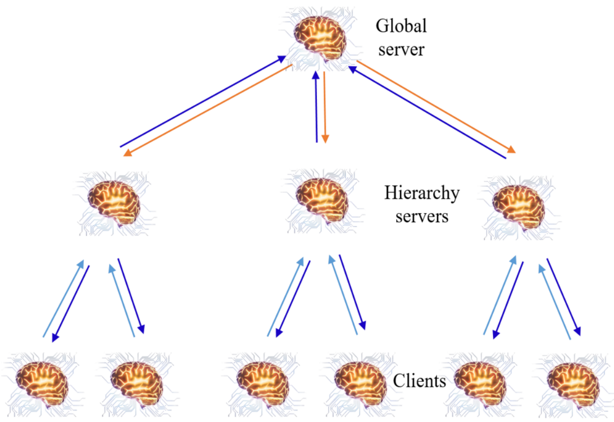

Figure 3. Hierarchical FL [63,64].

The server will sometimes become overloaded in a centralized design with just one server. While in a decentralized network, every participant will communicate with each other during each training round, which will result in very high communication cost across the network. These issues are known to be resolved by hierarchical architecture [63,64]. At each communication round, not all clients send model updates to the global server, but just a group of clients transfers their model updates to hierarchy servers (group or hierarchy server). The group servers then aggregate hierarchical updates and send them to the global server. Thus, this architecture significantly balances the communication load to the global server. Additionally, it resembles the hierarchical organizations (silos) in real life, making applying FL simpler. The pseudocode of this training type can be found in Algorithm 3.

Algorithm 3 FedAvg Hierarchy

ServerSite: ▷ Run on Server
initialize model parameters W_0
for each training round t **do**
  for each hierarchy server h **do**
    random select clients
    for each client k **in parallel do**
      $W_{t+1}^k \leftarrow$ **ClientSite**(k, W_t)
    end for
    $W_{t+1}^h \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} W_{t+1}^k$ ▷ Hierarchy update
  end for
  $W_{t+1} \leftarrow \sum_{h=1}^{H} \frac{n_h}{n} W_{t+1}^h$ ▷ Global update
end for

ClientSite(k, W_t): ▷ Run on Clients
  Fit local model with it dataset, batch size, epochs, and learning rate to get new model update W_{t+1}^k
Return (W_{t+1}^k) **to server**

3.2. FL Categories

Depending on the data distributions of the clients, we can classify FL into three categories: horizontal FL (HFL), vertical FL (VFL) and federated transfer learning (FTL) [43,44,46,65]. HFL is sample-based FL where data sets share the same feature but are different in samples. Otherwise, VFL is a feature-based FL, which applies to the cases where two data sets share the same sample ID but different features; it is also called heterogeneous FL. FTL is a combination of HFL and VFL; it applies to the two data sets that are not only in samples but also in features. Figure 4a–c display diagram of VFL, HFL and FTL, respectively.

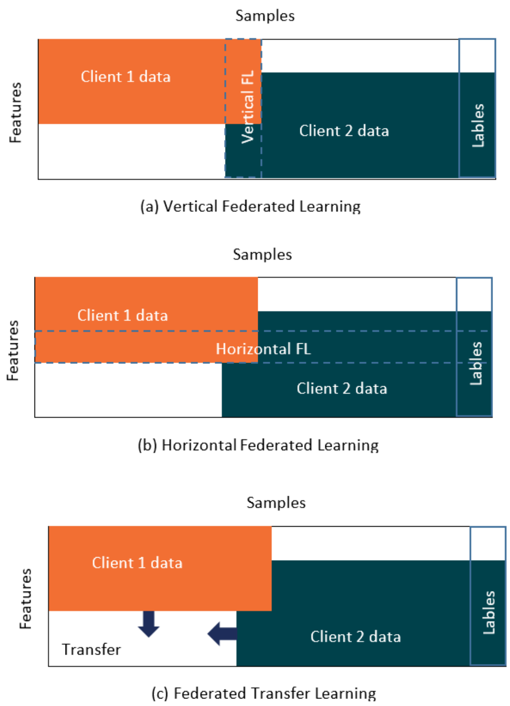

Figure 4. FL Categories [65].

3.3. FL Domains

FL has been applied worldwide. In this subsection, some representative works are presented.

- FL in IoT: Nguyen, D.C., et al. [56] was inspired by the dearth of a thorough study on FL's application in IoT. The authors first discuss current developments in FL and IoT to close this gap and offer thoughts on how they may be combined. FL is used in important IoT services, such as IoT data sharing, data offloading and caching, attack detection, localization, mobile crowdsensing, and IoT privacy and security. They also share the most recent advancements in integrated FL-IoT applications in several important use case areas, such as smart healthcare, smart transportation, UAVs, smart city, and smart industries, which have since caught the public interest.
  Jiang, J., et al. [60] made a survey to provide an outline of smart city sensing and its existing difficulties. They also discussed how FL could help to solve those difficulties. Both the state-of-the-art methods for FL and their use in smart city sensing are covered in detail; clear insights on unresolved problems, difficulties, and possibilities in this area are given as advice for the researchers looking into this topic.
- FL in Mobile: Some discussions on FL in mobile can be found in [43,57–59]. Lim, W.Y., et al. [58] provided a tutorial on FL and an in-depth analysis of the problems with FL implementation. The impetus for mobile edge computing is explained at the outset of the study, along with how FL may be used as an enabling technology for group model training at mobile edge networks. The fundamentals of DNN model training, FL, and system architecture for FL at scale are then discussed. The authors then offer thorough assessments, analyses, and comparisons of several implementation strategies for newly developing difficulties in FL. The cost of communication, resource distribution, data privacy, and data security are among the problems. Additionally, they discuss obstacles, future research prospects, and the use of FL for privacy-preserving mobile edge network optimization.
- FL in Healthcare: Rieke, N. et al. [45] foresee a federated for future digital health. The authors share their opinion with the community to provide context and detail about the advantages and impacts of FL for medical applications and highlight important considerations and challenges of implementing FL for future digital healthcare. There have been some lasted works in this scope, such as: a cloud-edge Network [66], a survey [67], and a survey in privacy preservation [68].
- Securing FL: Enthoven, D., et al. [61] examine FL's current weaknesses and then conduct a literature assessment of potential attack strategies aimed at FL's privacy protection features. A fundamental taxonomy is then used to characterize these assault strategies. They also offer a literature review of the most recent FL defense techniques and algorithms designed to counter these assaults. By the relevant fundamental defense concept, these defensive tactics are grouped. The use of a single defensive approach, according to the authors, is insufficient to offer sufficient defense against all known assault techniques.

4. Challenges of Federated Learning

- Communication cost: Because federated networks may have a vast number of devices, the computation time of FL may take longer than local training. Even though there have been several studies addressing this issue, it is imperative to build communication-efficient approaches as a result of increasing clients. Some crucial factors to consider in this situation are reducing the total number of communication rounds, finding fast convergence algorithms, and making as small model updates as possible at each round. To address this issue, adding split learning to FL is one of the effective solutions. The following are the two examples of this direction.
  Pathak, R.,et al. [69] first examine various earlier methods for federated optimization and demonstrate that, even in straightforward convex settings with deterministic updates, their fixed points do not always have to coincide with the stationary points

of the original optimization problem. They provide FedSplit, a set of methods for solving distributed convex minimization with an additive structure that is based on operator splitting processes, to address these problems. They demonstrate that these methods have the right fixed points, which correspond to the solutions of the initial optimization problem, and they describe the speeds at which these methods converge under various conditions. These techniques are resistant to the inaccurate calculation of intermediate local quantities, according to their theory. They support their argument with a few straightforward experiments that show how effective their plans are in action.

In addition, Thapa, C., et al. proposed SplitFed [70], a novel distributed ML approach called splitfed learning (SFL) that combines (FL) and split learning (SL), eliminating the inherent drawbacks of each approach, as well as a refined architectural configuration incorporating differential privacy and PixelDP to improve data privacy and model robustness. According to their study and actual findings, SFL offers comparable test accuracy and communication effectiveness to SL while having a much shorter calculation time per global epoch for numerous clients. The effectiveness of its communication through FL also increases with the number of clients, just like in SL. Additionally, the effectiveness of SFL with privacy and resilience safeguards is assessed in more extensive experimental situations.

- Heterogeneity: FL faces a considerable challenge when operating in various devices and data of the whole system [71–73]. Indeed, increasingly intelligent devices can connect to train the FL system. These devices have different hardware and software architectures (e.g., storage capacity, CPU topologies, power consumption level, operating system software, network bandwidth). In addition, each type of device is also designed to collect or generate different amounts and types of data, thus data nowadays is usually not independent and identically distributed (non IID data). To accommodate the heterogeneity of the system, it is necessary to develop clients' selection solutions during the training process to ensure convergence of the whole system. Dealing with the diversity in statistical heterogeneity, multi-model FL should be considered [44].

  Yu, F., et al. [74] created a solid structure-information alignment across collaborative models. The authors offer a unique FL approach to address the heterogeneity of the FL system. To provide explicit feature information allocation in various neural network topologies, they specifically construct a feature-oriented regulation mechanism (Net). Matchable structures with similar feature information may be initialized at the very early training stage by using this regulating strategy on collaborative models. Dedicated cooperation methods further provide ordered information delivery with clear structure matching and full model alignment during the FL process in both IID and non-IID scenarios. In the end, this framework efficiently improves the applicability of FL to numerous heterogeneous contexts while offering good convergence speed, accuracy, and computation/communication efficiency.

  Moreover, to address the heterogeneity of computational capabilities of edge devices, Wang, C., et al. [75] suggested a novel heterogeneous FL framework built on multi-branch deep neural network models. This framework enables the client devices to choose the best sub-branch model for their computing capabilities. They also provide MFedAvg, a model training aggregation approach that uses branch-wise averaging-based aggregation. Extensive experiments on MNIST, FashionMNIST, MedMNIST, and CIFAR-10 demonstrate that their suggested approaches can achieve satisfactory performance with guaranteed convergence and efficiently use all the resources available for training across different devices with lower communication cost than its homogeneous counterpart.

  Abdelmoniem, A.M., and Canini, M. [73] also concentrate on reducing the degree of device heterogeneity by suggesting AQFL, a straightforward and useful method that uses adaptive model quantization to homogenize the customers' computational resources. They assess AQFL using five standard FL metrics. The findings demonstrate that AQFL achieves roughly the same quality and fairness in diverse environments as the model trained in homogeneous conditions.

- Deep learning model architecture: DNN is proven to outperform other models in previous research, but its architecture in both traditional ML methods and FL was often pre-fixed [44,76]. Such a pre-defined setting carries many subjective factors that lead to the consequence that the models may fall into local optimal states. To address this challenge, [77] proposed a unified deep learning framework that suits data in multiple modalities and dimensions; the authors proved that using the proposed framework yields better performance than traditional ML approaches. In addition, [44] recommended using neural architecture search (NAS) to find the appropriate deep learning architecture for each task on each client.
- Securing FL system: Clients can learn a common global model cooperatively using FL without disclosing their training data to a cloud server. Malicious clients, however, can tamper with the global model and cause it to forecast inaccurate labels for test cases. Through ensemble, FL, Cao, X., et al. [78] close this gap. They utilize the technique to train numerous global models, each of which is learned using a randomly chosen subset of clients, given any basic FL algorithm. They use the global models' majority vote when predicting a testing example's label. They demonstrate that their ensemble FL is demonstrably safe against malicious clients using any FL base algorithm. In particular, a testing example's label predicted by their ensemble global model is provably unaffected by a finite number of malevolent clients. They also demonstrate the tightness of their derived limit. On the MNIST and Human Activity Recognition datasets, they assess their methodology. For instance, when 20 out of 1000 clients are malicious, their technique may obtain a verified accuracy of 88% on MNIST.

5. Benchmarks for Federated Learning

Numerous platforms and datasets are created to facilitate the operation of FL, which is advantageous for newcomers. Those benchmarks can be found in [43,44,58], and here we describe some popular ones.

- TensorFlow Federated (TFF) [79]: This is an open-source platform for FL and other decentralized data processing. It is led by Google and has grown in prominence in recent years. TFF allows developers to experiment with innovative algorithms by simulating the included FL algorithms on their models and data. Researchers will discover the beginning points and comprehensive examples for a wide range of studies.
- Leaf [80]: Leaf is an open-source framework for federated settings. It comprises a collection of federated datasets, a wide range of statistical and system metrics, and a collection of model implementations. Researchers and practitioners in fields such as FL, meta-learning, and multi-task learning will be able to evaluate novel solutions under more reasonable presumptions thanks to this platform.
- Flower [81]: A more smooth transition from an experimental study in simulation to system research on a large cohort of actual edge devices is made possible by Flower. It is a revolutionary end-to-end FL platform. Regarding simulation and real-world technology, Flower offers individual strength in both areas and allows experimental implementations to move between the two extremes as needed.
- PySyf: Ryffel, T., et al. [82] describe and explore a novel paradigm for privacy-preserving deep learning. The system prioritizes data ownership and secure processing, and it presents a useful representation based on command chains and tensors. This abstraction enables the implementation of complicated privacy-preserving structures such as FL, Secure Multiparty Computation, and Differential Privacy while providing a familiar deep learning API to the end user. They provide preliminary findings from the Boston Housing and Pima Indian Diabetes datasets. While privacy measures other than Differential Privacy do not affect prediction accuracy, the

current implementation of the framework creates a large performance burden that will be addressed at a later stage of development. The authors feel that their work is a significant step toward developing the first dependable, universal framework for privacy-preserving deep learning.

- FedML: He, C., et al. [83] present FedML, an open research library and benchmark for FL algorithm development and fair performance comparison, in this paper. FedML offers three computing paradigms: edge device on-device training, distributed computing, and single-machine simulation. FedML also encourages various algorithmic research by the design of flexible and generic APIs and detailed reference baseline implementations (optimizer, models, and datasets). FedML, they think, will provide an efficient and reproducible method for building and assessing FL algorithms that will assist the FL research community.
- Sherpa.ai: Barroso, N.R., et al. [84] introduced the Sherpa.ai FL framework, which is based on a comprehensive understanding of FL and differential privacy. It is the outcome of research into how to adapt the ML paradigm to FL as well as the development of methodological recommendations for producing artificial intelligence services based on FL and differentiated privacy. Using classification and regression use cases, they also demonstrate how to adhere to the methodological principles with the Sherpa.ai FL framework.
- EMNIST: A benchmark database is presented by Cohen, G., et al. [85]. It is a derivative of the entire NIST dataset known as Extended MNIST (EMNIST), which uses the same conversion paradigm as the MNIST dataset. The end result is a collection of datasets that represent more difficult classification problems containing letters and digits, but which share the same picture structure and characteristics as the original MNIST job, allowing direct interoperability with all current classifiers and systems. Benchmark results are reported, along with validation of the conversion procedure by comparison of classification results on converted NIST and MNIST digits.
- FedEval: Di Chai, et al, [86] suggested a thorough approach to FL system evaluation in their works. Particularly, they first offer the ACTPR model, which identifies five metrics—accuracy, communication, time efficiency, privacy, and robustness—that cannot be disregarded in the FL assessment. Then, using FedEval, a benchmarking system they developed and put into use, it is possible to compare and evaluate previous works in a systematic manner while maintaining a constant experimental environment. The authors next present a thorough benchmarking analysis between FedSGD and FedAvg, the two most used FL methods. According to the benchmarking results, FedSGD and FedAvg both have benefits and drawbacks when using the ACTPR model. For instance, the non-IID data issue has little to no effect on FedSGD, while in their studies, FedAvg has an accuracy loss of up to 9%. However, when it comes to communication and time consumption, FedAvg outperforms FedSGD. Finally, they uncover a series of key findings that will be highly beneficial for researchers working in the FL region.
- OARF: An Open Application Repository for FL (OARF) is presented by Sixu Hu, et al, [87]. It is a benchmark set for federated ML systems. OARF simulates more realistic application situations. The benchmark suite is heterogeneous in terms of data quantity, dispersion, feature distribution, and learning task difficulty, as observed by their categorization. The thorough analyses with reference implementations highlight potential areas for future FL system development. The key components of FL, such as model correctness, communication cost, throughput, and convergence time, have been constructed as reference implementations and reviewed. They made several intriguing discoveries as a result of these tests, such as the fact that FL may significantly boost end-to-end throughput.

From the papers we gathered, there have been some separated AQI datasets, as shown at the end of Table 1. To the best of our knowledge, there is presently no standard dataset designed expressly for AQI analysis, even though there are several benchmarks for FL in general. Future studies in this area must consider this problem.

6. Multi-Models Federated Learning

Forecasting results from a single model are worse and have less integration with other methods [1]. The outcomes in [17,22] demonstrate the better result when applying ensemble methods. [44] suggested using multi-models FL to deal with the heterogeneous statistic. Currently, some studies on the multi-model in AQI prediction are found in [88–90], but these applications are implemented on a single machine. Several authors have studied multi-model FL for other fields, as shown in [91–93] and [86,94,95]. In this section, some of them are presented with the hope that they will help convey new directions for AQI forecasting in the future.

Bhuyan, N. and Moharir, S., [91] show that several unrelated models can be trained concurrently in a federated environment. The authors expand the FedAvg method to accommodate multi-model training. They also suggest two novel approaches for client selection in multi-model training in federated settings: RanklistMulti-UCB and Pareto-Multi-UCB. Additionally, they use FedAvg to examine the effectiveness of these three strategies when trained on a single model. Both synthetic and real-world datasets are used for this. They conclude that Ranklist-Multi-UCB and Pareto-Multi-UCB perform better than the single model FedAvg when the number of clients each round is low.

Zhao, Y., et al. [92] utilized the multimodal in cooperated with semi-supervised FL to IoT devices in their research. Specifically in the client site, they offer a multimodal and semi-supervised FL framework for training auto-encoders to extract shared or correlated representations from several local data modalities. Furthermore, they provide a multimodal FedAvg technique for aggregating local auto-encoders trained on different input modalities. With the support of auxiliary labeled data on the server, they apply the learned global auto-encoder for a downstream classification assignment. They experimentally test their system using several modalities such as sensory data, depth camera films, and RGB camera videos. The experimental results demonstrate that incorporating input from many modalities into FL can increase classification performance. Furthermore, they can use labeled data from only one modality for supervised learning on the server and apply the learned model to training datasets from other modalities to achieve decent F1 scores (e.g., with the best performance being more excellent than 60%), especially when combining achievements from both unimodal and multimodal clients.

Smith, V., et al. provide the MOCHA method to address multi-task FL in [94]. They first demonstrate that multi-task learning is a logical option for dealing with statistical difficulties in the federated scenario. Second, they create a brand-new technique called MOCHA to address a general issue with multi-task learning. They offer MOCHA convergence assurances that properly take into account these particular system problems and give information on actual performance. Finally, they use a novel benchmarking suite of federated datasets to show MOCHA's improved empirical performance. The proposed method is said to have challenges in FL as: high communication cost, stragglers, and fault tolerance.

While [94] develops multi-model approach for multi-task, Xie, M., et al. [86] provide a novel multi-center aggregation technique to achieve a multi-model for a single task. It generates the best match between users and centers while learning several global models from the data. Then, the authors define it as a bi-level optimization issue that can be successfully resolved by the stochastic expectation maximization technique. Experiments on several benchmark datasets of FL demonstrate that their technique beats several well-known FL rivals.

Muhammad, A., et al. [95] suggest a FedMD modification called the Robust Multi-Model FL (RMMFL) framework, which is used with the same set of assumptions and greatly enhances the output of each individual model. RMMFL modified the FedMD training procedure in two ways: To soften the output predictions, a high entropy aggregation approach is first introduced. Second, a weighted ensemble approach is utilized to balance each client model's predictions according to its performance. The CIFAR/MNIST benchmark datasets are used extensively in research on heterogeneous models, and the simulation findings from their work reveal that RMMFL has an accuracy advantage of 5% over the standard approach.

Bhuyan, N., Moharir, S., and Joshi, G. [96] concentrate on the issue of simultaneously training many independent models with a common pool of customers using FL. They suggest two iterations of the well-researched FedAvg algorithm, known as MFA-Rand and MFA-RR, in the context of multiple models and demonstrate their convergence. They demonstrate that for MFARR, increasing the data sample size (for client-side SGD rounds) significantly speeds up convergence. They also suggest performance criteria for determining multi-model FL's advantage over single-model FL. They define the circumstances in which running MFA-Rand for M models concurrently is preferable to running FedAvg for one model for each model one at a time. To validate the findings of their analytical work, they conduct experiments in strongly convex and convex environments. They demonstrate the advantages of the multi-model FL in deep learning with experiments conducted in a non-convex environment. Additionally, they conduct tests that are unrelated to the suggested context. These were the tests with partial device participation and real-world situations. They see a benefit in simultaneously training many models here as well.

Through the reviews in this section, we can see an attractive direction of multi-model FL that could be applied for the AQI forecast in the future. Thus, later works in AQI analysis should have great concern for this method. We inferred from [86,92] the topology and algorithm for multi-model centralized FL as shown in Figure 5 and Algorithm 4, respectively.

Algorithm 4 Multi-model FedAvg Centralized

ServerSite: ▷ Run on Server
initialize M model parameters W_0^m
for each training round t **do**
  random select clients
  for each client k **in parallel do**
    $W_{t+1}^{Mk} \leftarrow$ **ClientSite**(k, W_t^M) ▷ Client update

    for each model m **in M do**

    $W_{t+1}^{m} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} W_{t+1}^{mk}$ ▷ Global update

  end for
end for

ClientSite(k, W_t^M): ▷ Run on Clients
for each model m **in M do**
  Fit local model m with it dataset, batch size, epochs, and learning rate to get new model update W_{t+1}^{mk}
Return (W_{t+1}^{Mk}) **to server**

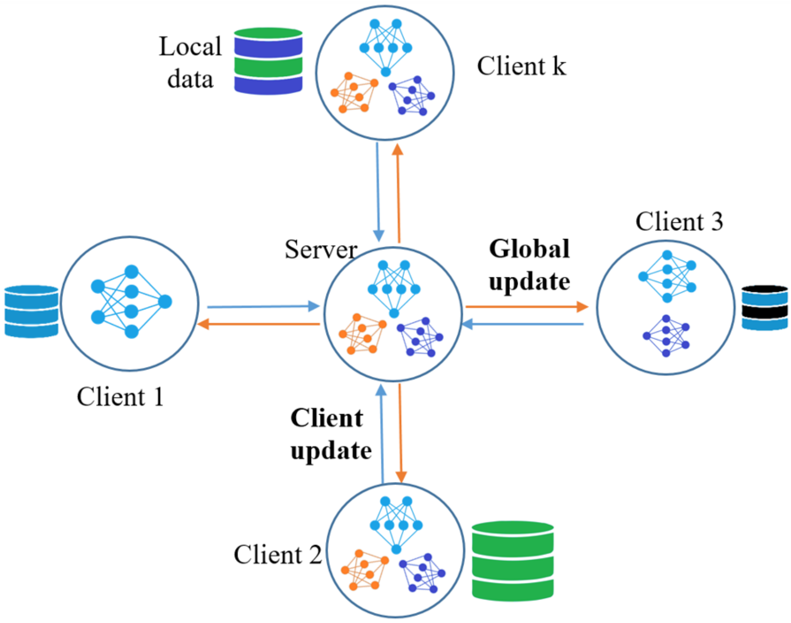

Figure 5. Multi-model centralized FL.

7. Conclusions

We have gathered several AQI prediction studies as well as FL research in this survey. We discovered through research and synthesis that, prior to 2020, the majority of AQI forecasting mechanisms relied on single ML techniques such as statistics, ANN, DNN, hybrid, and ensemble. Although commonly used, ANN and DNN were prone to failure in local optimum conditions. The anticipated outcomes could be enhanced by using some hybrid and ensemble approaches.

Since 2000, various works have integrated FL into AQI processing. We have identified the three different FL architectures: centralized FL, decentralized FL, and hierarchical FL as well as their corresponding algorithms. The system mostly used centralized FL, and the typical process model was DNNs. There has been a trend in the use of UAVs in this field. Additionally, we mentioned several benchmarks, challenges and solutions, and domains of FL.

Reviews demonstrate that multi-model FL improves the accuracy of other domains, thus it could be applied to enhance AQI predictions. Multi-model FL is a study area that needs to be focused on in future works. DNN should be incorporated into client models, since it is more responsive than conventional ML architectures. Additionally, DNN should be designed to adopt both partial and temporal data.

More importantly, to have a common baseline to assess new solutions, the scientific community should also be interested in developing a dedicated open dataset on AQI.

Author Contributions: Project administration and conceptualization, M.-S.D., A.-K.T.; writing—review and editing, D.-D.L., A.-K.T., M.-S.D., K.-C.N.-L., H.-S.L., X.-D.N.-T., T.-Q.P., V.-L.N., B.-Y.N.-T. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the University of Economic Ho Chi Minh City (UEH) Vietnam grant number 2022-09-19-1160.

Data Availability Statement: This is a survey, we did not generate any data.

Acknowledgments: We acknowledge the University of Economic Ho Chi Minh City (UEH) for funding this research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Bai, L.; Wang, J.; Ma, X.; Lu, H.H. Air Pollution Forecasts: An Overview. *Int. J. Environ. Res. Public Health* **2018**, *15*, 780. [CrossRef] [PubMed]
2. Rahman, N.H.A.; Lee, M.H.; Latif, M.T. Artificial neural networks and fuzzy time series forecasting: An application to air quality. *Qual. Quant.* **2015**, *49*, 2633–2647 [CrossRef]
3. Grivas, G.; Chaloulakou, A. Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. *Atmos. Environ.* **2006**, *40*, 1216–1229 [CrossRef]
4. Elangasinghe, M.A.; Singhal, N.; Dirks, K.N.; Salmond, J.A. Development of an ANN–based air pollution forecasting system with explicit knowledge through sensitivity analysis. *Atmos. Pollut. Res.* **2014**, *5*, 696–708 [CrossRef]
5. Bai, Y.; Li, Y.; Wang, X.; Xie, J.; Li, C. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmos. Pollut. Res.* **2016**, *7*, 557–566. [CrossRef]
6. Mishra, D.; Goyal, P. NO_2 forecasting models Agra. *Atmos. Pollut. Res.* **2015**, *6*, 99–106. [CrossRef]
7. Kurt, A.; Oktay, A.B. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Syst. Appl.* **2010**, *37*, 7986–7992. [CrossRef]
8. Masood, A.; Ahmad, K. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *J. Clean. Prod.* **2021**, *322*, 129072. [CrossRef]
9. Navares, R.; Aznarte, J.L. Predicting air quality with deep learning LSTM: Towards comprehensive models. *Ecol. Inf.* **2020**, *55*, 101019. [CrossRef]
10. Krishan, M.; Jha, S.; Das, J.; Singh, A.; Goyal, M.K.; Sekar, C. Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India. *Air Qual. Health* **2019**, *12* , 899–908. 00696-7. [CrossRef]
11. Li, S.; Xie, G.; Ren, J.; Guo, L.; Yang, Y.; Xu, X. Urban PM2. 5 concentration prediction via attention-based CNN–LSTM. *Appl. Sci.* **2020**, *10*, 1953. 10.3390/app10061953 [CrossRef]
12. Zhou, Y.; Chang, F.J.; Chang, L.C.; Kao, I.F.; Wang, Y.S. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *J. Clean. Prod.* **2019**, *209*, 134–145. [CrossRef]
13. Xayasouk, T.; Lee, H. Air pollution prediction system using deep learning. *WIT Trans. Ecol. Environ.* **2018**, *230*, 71–79. [CrossRef]
14. Li, X.; Peng, L.; Yao, X.; Cui, S.; Hu, Y.; You, C.; Chi, T. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environ. Pollut.* **2017**, *231*, 997–1004. envpol.2017.08.114. [CrossRef] [PubMed]
15. Freeman, B.S.; Taylor, G.; Gharabaghi, B.; The, J. Forecasting air quality time series using deep learning. *J. Air Waste Manag. Assoc.* **2018**, *68*, 866–886. [CrossRef]
16. Ma, J.; Li, Z.; Cheng, J.C.; Ding, Y.; Lin, C.; Xu, Z. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Sci. Total Environ.* **2020**, *705*, 135771. [CrossRef]
17. Baklanov, A.; Zhang, Y. Advances in air quality modeling and forecasting. *Glob. Transit.* **2020**, *2*, 261–270. [CrossRef]
18. Monache, L.D.; Deng, X.; Zhou, Y.; Stull, R. Ozone ensemble forecasts: 1. A new ensemble design. *J. Geophys. Res.* **2006**, *111*, D05307. [CrossRef]
19. Pagowski, M.; Grell, G.A.; Devenyi, D.; Peckham, S.E.; McKeen, S.A.; Gong, W.; Monache, L.D.; McHenry, J.N.; McQueen, J.; Lee, P. Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts. *Atmos. Environ.* **2006**, *40*, 3240–3250. [CrossRef]
20. Brasseur, G.P.; Xie, Y.; Petersen, A.K.; Bouarar, I.; Flemming, J.; Gauss, M.; Jiang, F.; Kouznetsov, R.; Kranenburg, R.; Mijling, B.; et al. Ensemble forecasts of air quality in eastern China—Part 1: Model description and implementation of the MarcoPolo–panda prediction system, version 1. *Geosci. Model Dev.* **2019**, *12*, 33–67. [CrossRef]
21. McKeen, S.; Chung, S.H.; Wilczak, J.; Grell, G.; Djalalova, I.; Peckham, S.; Gong, W.; Bouchet, V.; Moffet, R.; Tang, Y.; et al. Evaluation of several PM2.5 forecast models using data collected during the ICARTT/NEAQS 2004 field study. *J. Geophys. Res.* **2007**, *112*, D10S20. [CrossRef]
22. Masih, A. ML algorithms in air quality modeling. *Glob. J. Environ. Sci. Manag.* **2019**, *5*, 515–534.
23. Liao, K.; Huang, X.H.; Dang, H.; Ren, Y.; Zuo, S.; Duan, C. Statistical Approaches for Forecasting Primary Air Pollutants: A Review. *Atmosphere* **2021**, *12*, 686. [CrossRef]
24. Perez, P.; Trier, A.; Reyes, J. Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago, Chile. *Atmos. Environ.* **2000**, *34*, 1189–1196. [CrossRef]
25. Ceylan, Z.; Bulkan, S. Forecasting PM10 levels using ANN and MLR: A case study for Sakarya City. *Glob. Nest J.* **2018**, *20*, 281–290.
26. Belavadi, S.V.; Rajagopal, S.; Ranjani, R.; Mohan, R. Air Quality Forecasting using LSTM RNN and Wireless Sensor Networks. *Procedia Comput. Sci.* **2020**, *170*, 241–248. [CrossRef]
27. Liao, Q.; Zhu, M.; Wu, L.; Pan, X.; Tang, X.; Wang, Z. Deep Learning for Air Quality Forecasts: A Review. *Curr. Pollut. Rep.* **2020**, *6*, 399–409. [CrossRef]
28. Liu, H.; Yin, S.; Chen, C.; Duan, Z. Data multi-scale decomposition strategies for air pollution forecasting: A comprehensive review. *J. Clean. Prod.* **2020**, *277*, 124023. [CrossRef]

29. Zhu, S.; Lian, X.; Liu, H.; Hu, J.; Wang, Y.; Che, J. Daily air quality index forecasting with hybrid models: A case in China. *Environ. Pollut.* **2017**, *231*, 1232–1244 . [CrossRef]
30. Patil, R.M.; Dinde, D.H.; Powar, S.K. A Literature Review on Prediction of Air Quality Index and Forecasting Ambient Air Pollutants using ML Algorithms. *Int. J. Innov. Sci. Res. Technol.* **2020**, *5*, 1148–1152 [CrossRef]
31. Bui, T.C.; Le, V.D.; Cha, S.K. A Deep Learning Approach for Forecasting Air Pollution in South Korea Using LSTM. *arXiv* **2018**, arXiv:1804.07891v3.
32. Kim, S.; Lee, J.M.; Lee, J.; Seo, J. Deep-dust: Predicting concentrations of fine dust in Seoul using LSTM. *arXiv* **2019**, arXiv:1901.10106
33. Huang, C.J.; Kuo, P.H. A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. *Sensors* **2018**, *18*, 2220. [CrossRef] [PubMed]
34. Soh, P.W.; Chang, J.W.; Huang, J.W. Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. *IEEE Access* **2018**, *6*, 38186–38199 [CrossRef]
35. Wang, H.; Zhuang, B.; Chen, Y.; Li, N.; Wei, D. Deep inferential spatial-temporal network for forecasting air pollution concentrations. *arXiv* **2018**, arXiv:1809.03964v1.
36. Govender, P.; Sivakumar, V. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmos. Pollut. Res.* **2020**, *11*, 40–56. [CrossRef]
37. Cabaneros, S.M.; Calautit, J.K.; Hughes, B.R. A review of artificial neural network models for ambient air pollution prediction. *Environ. Model. Softw.* **2019**, *119*, 285–304. [CrossRef]
38. Bekkar, A.; Hssina, B.; Douzi, S.; Douzi, K. Air-pollution prediction in smart city, deep learning approach. *J. Big Data* **2021**, *8*, 161. [CrossRef]
39. Karroum, K.; Lin, Y.; Chiang, Y.; Ben, Maissa, Y.; El haziti, M.; Sokolov, A.; Delbarre, H. A Review of Air Quality Modeling. *MAPAN* **2020**, *35*, 287–300. [CrossRef]
40. Garaga, R.; Sahu, S.K.; Kota, S.H. A Review of Air Quality Modeling Studies in India: Local and Regional Scale. *Curr. Pollut. Rep.* **2018**, *4*, 59–73. [CrossRef]
41. Wang, L.; Yang, J.; Zhang, P.; Zhao, X.; Wei, Z.; Zhang, F.; Su, J.; Meng, C. A Review of Air Pollution and Control in Hebei Province, China. *Open J. Air Pollut.* **2013**, *2*, 47. [CrossRef]
42. McMahan, H.B.; Moore, E.; Ramage, D.; HAMFson, S.; Arcas, B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), Lauderdale, FL, USA, 20–22 April 2017
43. Li, T.; Sahu, A.; Talwalkar, A.S.; Smith, V. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Process. Mag.* **2020**, *37*, 50–60. [CrossRef]
44. Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A.N.; Bonawitz, K.; Charles, Z.B.; Cormode, G.; Cummings, R.; et al. Advances and Open Problems in Federated Learning. *Found. Trends Mach. Learn.* **2021**, *14*, 1–210. [CrossRef]
45. Rieke, N.; Hancox, J.; Li, W.; Milletari, F.; Roth, H.R.; Albarqouni, S.; Bakas, S.; Galtier, M.; Landman, B.A.; Maier-Hein, K.H.; et al. The future of digital health with federated learning. *NPJ Digit. Med.* **2020**, *3*, 119. [CrossRef]
46. Zhu, H.; Xu, J.; Liu, S.; Jin, Y. Federated Learning on Non-IID Data: A Survey. *Neurocomputing* **2021**, *465*, 371–390. [CrossRef]
47. Chhikara, P.; Tekchandani, R.; Kumar, N.; Tanwar, S.; Rodrigues, J.J. Federated Learning for Air Quality Index Prediction using UAV Swarm Networks. In Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 7–11 December 2021; pp. 1–6.
48. Chhikara, P.; Tekch, ani, R.; Kumar, N.; Guizani, M.; Hassan, M.M. Federated Learning and Autonomous UAVs for Hazardous Zone Detection and AQI Prediction in IoT Environment. *IEEE Internet Things J.* **2021**, *8*, 15456–15467. [CrossRef]
49. Liu, Y.; Nie, J.; Li, X.; Ahmed, S.H.; Lim, W.Y.; Miao, C. Federated Learning in the Sky: Aerial-Ground Air Quality Sensing Framework With UAV Swarms. *IEEE Internet Things J.* **2021**, *8*, 9827–9837. [CrossRef]
50. Putra, K.T.; Chen, H.; Ogiela, M.R.; Chou, C.; Weng, C.; Shae, Z. Federated Compressed Learning Edge Computing Framework with Ensuring Data Privacy for PM2.5 Prediction in Smart City Sensing Applications. *Sensors* **2021**, *21*, 4586. [CrossRef]
51. Nguyen, D.V.; Zettsu, K. Spatially-distributed Federated Learning of Convolutional Recurrent Neural Networks for Air Pollution Prediction. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; pp. 3601–3608.
52. Zhou, X.; Liu, X.; Lan, G.; Wu, J. Federated conditional generative adversarial nets imputation method for air quality missing data. *Knowl. Based Syst.* **2021**, *228*, 107261. [CrossRef]
53. Huang, G.; Zhao, X.; Lu, Q. A new cross-domain prediction model of air pollutant concentration based on secure federated learning and optimized LSTM neural network. *Environ. Sci. Pollut. Res.* **2022**, 1–23. [CrossRef]
54. Chen, Y.; Ning, Y.; Slawski, M.; Rangwala, H. Asynchronous Online Federated Learning for Edge Devices with Non-IID Data. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 15–24.
55. Neo, E.X.; Hasikin, K.; Mokhtar, M.I.; Lai, K.W.; Azizan, M.M.; Razak, S.A.; Hizaddin, H.F. Towards Integrated Air Pollution Monitoring and Health Impact Assessment Using Federated Learning: A Systematic Review. *Front. Public Health* **2022**, *10*, 851553. [CrossRef] [PubMed]

56. Nguyen, D.C.; Ding, M.; Pathirana, P.N.; Seneviratne, A.P.; Li, J.; Poor, H.V. Federated Learning for Internet of Things: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 1622–1658. [CrossRef]
57. Li, Q.; Wen, Z.; He, B. A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection. *arXiv* **2021**, arXiv:1907.09693.
58. Lim, W.Y.; Luong, N.C.; Hoang, D.T.; Jiao, Y.; Liang, Y.; Yang, Q.; Niyato, D.T.; Miao, C. Federated Learning in Mobile Edge Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 2031–2063. [CrossRef]
59. Ogundokun, R.O.; Misra, S.; Maskeliūnas, R.; Damaševičius, R. A Review on Federated Learning and ML Approaches: Categorization, Application Areas, and Blockchain Technology. *Information* **2022**, *13*, 263. [CrossRef]
60. Jiang, J.; Kantarci, B.; Oktug, S.F.; Soyata, T. Federated Learning in Smart City Sensing: Challenges and Opportunities. *Sensors* **2020**, *20*, 6230. [CrossRef]
61. Enthoven, D.; Al-Ars, Z. An Overview of Federated Deep Learning Privacy Attacks and Defensive Strategies. *arXiv* **2021**, arXiv:2004.04676.
62. Huang, Y.; Wu, C.; Wu, F.; Lyu, L.; Qi, T.; Xie, X. A Federated Graph Neural Network Framework for Privacy-Preserving Personalization. *Nat. Commun.* **2022**, *13*, 1–10
63. Liu, L.; Zhang, J.; Song, S.H.; Letaief, K.B. Edge-Assisted Hierarchical Federated Learning with Non-IID Data. *arXiv* **2019**, arXiv:1905.06641.
64. Lin, T.; Stich, S.U.; Jaggi, M. Don't Use Large Mini-Batches, Use Local SGD. *arXiv* **2020**, arXiv:1808.07217.
65. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated ML: Concept and Applications. *arXiv* **2019**, arXiv:1902.04885.
66. Khoa, T.A.; Nguyen, D.-V.; Dao, M.-S.; Zettsu, K. Fed xData: A Federated Learning Framework for Enabling Contextual Health Monitoring in a Cloud-Edge Network. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; pp. 4979–4988. [CrossRef]
67. Nguyen, D.C.; Pham, Q.; Pathirana, P.N.; Ding, M.; Seneviratne, A.P.; Lin, Z.; Dobre, O.A.; Hwang, W.J. Federated Learning for Smart Healthcare: A Survey. *arXiv* **2021**, arXiv:2111.08834.
68. Ali, M.; Naeem, F.; Tariq, M.A.; Kaddoum, G. Federated Learning for Privacy Preservation in Smart Healthcare Systems: A Comprehensive Survey. *arXiv* **2022**, arXiv:2203.09702.
69. Pathak, R.; Wainwright, M.J. FedSplit: An algorithmic framework for fast federated optimization. *arXiv* **2020**, arXiv:2005.05238.
70. Thapa, C.; Arachchige, P.; Çamtepe, S.A. SplitFed: When Federated Learning Meets Split Learning. *arXiv* **2022**, arXiv:2004.12088.
71. Abdelmoniem, A.M.; Ho, C.; Papageorgiou, P.; Canini, M. Empirical analysis of federated learning in heterogeneous environments. In Proceedings of the 2nd European Workshop on Machine Learning and Systems, Rennes, France, 5–8 April 2022.
72. Yang, C.; Wang, Q.; Xu, M.; Chen, Z.; Bian, K.; Liu, Y.; Liu, X. Characterizing Impacts of Heterogeneity in Federated Learning upon Large-Scale Smartphone Data. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021.
73. Abdelmoniem, A.M.; Canini, M. Towards Mitigating Device Heterogeneity in Federated Learning via Adaptive Model Quantization. In Proceedings of the 1st Workshop on Machine Learning and Systems, New York, NY, USA, 10 November 2019.
74. Yu, F.; Zhang, W.; Qin, Z.; Xu, Z.; Wang, D.; Liu, C.; Tian, Z.; Chen, X. Heterogeneous federated learning. *arXiv* **2020**, arXiv:2008.06767.
75. Wang, C.; Huang, K.; Chen, J.; Shuai, H.; Cheng, W. Heterogeneous Federated Learning Through Multi-Branch Network. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July; pp. 1–6.
76. Dong, L.D.; Nguyen, V.T.; Le D.T.; Tiep, M.V.; Hien, V.T.; Huy, P.P.; Trung, H.P. Modeling Transmission Rate of COVID-19 in Regional Countries to Forecast Newly Infected Cases in a Nation by the Deep Learning Method. In Proceedings of the 8th International Conference, FDSE 2021, Virtual Event, 24–26 November 2021.
77. Xi, P.; Shu, C.; Goubran, R.A. A Unified Deep Learning Framework for Multi-Modal Multi-Dimensional Data. In Proceedings of the 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Istanbul, Turkey, 26–28 June 2019; pp. 1–6.
78. Cao, X.; Jia, J.; Gong, N.Z. Provably Secure Federated Learning against Malicious Clients. *Proc. Aaai Conf. Artif. Intell.* **2021**, *35*, 6885–6893 [CrossRef]
79. Tensorflow Federated: ML on Decentralized Data. Available online: https://www.tensorflow.org/federated (accessed on 6 November 2022).
80. Caldas, S.; Wu, P.; Li, T.; Kone ˇcny, J.; McMahan, H.B.; Smith, V.; Talwalkar, A. Leaf: A benchmark for federated settings. *arXiv* **2018**, arXiv:1812.01097.
81. Beutel, D.J.; Topal, T.; Mathur, A.; Qiu, X.; Parcollet, T.; Gusmão, P.P.; Lane, N.D. Flower: A Friendly Federated Learning Framework. Available online: https://hal.archives-ouvertes.fr/hal-03601230/document (accessed on 6 November 2022).
82. Ryffel, T.; Trask, A.; Dahl, M.; Wagner, B.; Mancuso, J.V.; Rueckert, D.; Passerat-Palmbach, J. A generic framework for privacy preserving deep learning. *arXiv* **2018**, arXiv:1811.04017.
83. He, C.; Li, S.; So, J.; Zhang, M.; Wang, H.; Wang, X.; Vepakomma, P.; Singh, A.; Qiu, H.; Shen, L.; et al. FedML: A Research Library and Benchmark for Federated Machine Learning. *arXiv* **2020**, arXiv:2007.13518.
84. Barroso, N.R.; Stipcich, G.; Jim'enez-L'opez, D.; Ruiz-Millán, J.A.; Martínez-Cámara, E.; González-Seco, G.; Luzón, M.V.; Veganzones, M.A.; Herrera, F. Federated Learning and Differential Privacy: Software tools analysis, the Sherpa.ai FL framework and methodological guidelines for preserving data privacy. *Inf. Fusion* **2020**, *64*, 270–292. [CrossRef]

85. Cohen, G.; Afshar, S.; Tapson, J.C.; Schaik, A.V. EMNIST: An extension of MNIST to handwritten letters. *arXiv* **2017**, arXiv:1702.05373.
86. Xie, M.; Long, G.; Shen, T.; Zhou, T.; Wang, X.; Jiang, J. Multi-Center Federated Learning. *arXiv* **2020**, arXiv:2108.08647.
87. Hu, S.; Li, Y.; Liu, X.; Li, Q.; Wu, Z.; He, B. The oarf benchmark suite: Characterization and implications for federated learning systems. *arXiv* **2020**, arXiv:2006.07856.
88. Seaton, M.D.; O'Neill, J.; Bien, B.; Hood, C.; Jackson, M.; Jackson, R.; Johnson, K.; Oades, M.; Stidworthy, A.; Stocker, J.; et al. A Multi-model Air Quality System for Health Research: Road model development and evaluation. *Environ. Model. Softw.* **2022**, *155*, 105455. [CrossRef]
89. Qi, H.; Ma, S.; Chen, J.; Sun, J.; Wang, L.; Wang, N.; Wang, W.; Zhi, X.; Yang, H. Multi-model Evaluation and Bayesian Model Averaging in Quantitative Air Quality Forecasting in Central China. *Aerosol Air Qual. Res.* **2022**, *22*, 210247. [CrossRef]
90. Zhang, J.; Li, S. Air quality index forecast in Beijing based on CNN-LSTM multi-model. *Chemosphere* **2022**, *308*, 136180. [CrossRef]
91. Bhuyan, N.; Moharir, S. Multi-Model Federated Learning. In Proceedings of the 2022 14th International Conference on COMmunication Systems and NETworkS (COMSNETS), Bangalore, India, 4–8 January 2022; pp. 779–783.
92. Zhao, Y.; Barnaghi, P.M.; Haddadi, H. Multimodal Federated Learning on IoT Data. In Proceedings of the 2022 IEEE/ACM Seventh International Conference on Internet-of-Things Design and Implementation (IoTDI), Milano, Italy, 4–6 May 2022.
93. Xiong, B.; Yang, X.; Qi, F.; Xu, C. A unified framework for multi-modal federated learning. Neurocomputing **2022**, *480*, 110–118. [CrossRef]
94. Smith, V.; Chiang, C.; Sanjabi, M.; Talwalkar, A.S. Federated Multi-Task Learning. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
95. Muhammad, A.; Lin, K.; Gao, J.; Chen, B. Robust Multi-model Personalized Federated Learning via Model Distillation. In Proceedings of the 21st International Conference, ICA3PP 2021, Virtual Event, 3–5 December 2021.
96. Bhuyan, N.; Moharir, S.; Joshi, G. Multi-Model Federated Learning with Provable Guarantees. *arXiv* **2022**, arXiv:2207.04330.