Latency-aware blockage prediction in vision-aided federated wireless networks

Ahsan Raza Khan[1]*, Iftikhar Ahmad[1], Lina Mohjazi[1],
Sajjad Hussain[1], Rao Naveed Bin Rais[2], Muhammad Ali Imran[1,2]
and Ahmed Zoha[1]

[1]James Watt School of Engineering, University of Glasgow, Glasgow, United Kingdom, [2]Artificial
Intelligence Research Center (AIRC), Ajman University, Ajman, United Arab Emirates

Introduction: The future wireless landscape is evolving rapidly to meet ever-increasing data requirements, which can be enabled using higher-frequency spectrums like millimetre waves (mmWaves) and terahertz (THz). However, mmWave and THztechnologies rely on line-of-sight (LOS) communication, making them sensitive to sudden environmental changes and higher mobility of users, especially in urban areas.

Methods: Therefore, beam blockage prediction is a critical challenge for sixth-generation (6G) wireless networks. One possible solution is to anticipate the potential change in the wireless network surroundings using multi-sensor data (wireless, vision, lidar, and GPS) with advanced deep learning (DL) and computer vision (CV) techniques. Despite numerous advantages, the fusion of deep learning,computer vision, and multi-modal data in centralised training introduces many challenges, including higher communication costs for raw data transfer, inefficient bandwidth usage and unacceptable latency. This work proposes latency-aware vision-aided federated wireless networks (VFWN) for beam blockage prediction using bimodal vision and wireless sensing data. The proposed framework usesdistributed learning on the edge nodes (EN) for data processing and model training.

Results and Discussion: This involves federated learning for global model aggregation that minimizes latency and data communication cost as compared to centralised learning while achieving comparable predictive accuracy. For instance, the VFWN achieves a predictive accuracy of 98.5%, which is comparable to centralised learning with overall predictive accuracy 99%, considering that no data sharing is done. Furthermore, the proposed framework significantly reduces the communication cost by 81.31% and latency by 6.77% using real-time on device processing and inference.

1 Introduction

Millimetre wave (mmWave) and terahertz (THz) communications are considered prospective contenders for future sixth-generation (6G) wireless communication networks. With larger bandwidth, the mmWave and THz technologies have the ability to meet the ever-increasing data requirements, supporting emerging technologies, including smart healthcare, industry 4.0, holographic telepresence, virtual or augmented reality (VR/VR) and autonomous vehicles Khan et al. (2022); Nishio et al., 2021. Furthermore, the higher frequencies also provide ultra-reliable low-latency communication (URLLC), massive connectivity and higher throughput. However, higher frequencies and growing antenna arrays carry significant control overhead, preventing them from achieving their full potential. Furthermore, the propagation properties of future wireless systems will drastically change, as mmWaves and THz signals are notorious for their poor penetration and significant power loss when reflecting from surfaces Alrabeiah et al. (2020a). This emphasises the need for antenna directivity, which requires huge antenna arrays and Line-of-Sight (LOS) links between the base station (BS) and the end user. Therefore, beam-forming is likely to be used in 5G and 6G communication networks owing to its remarkable features, including higher spatial reuse, increased throughput, higher capacity, and minimised interference Al-Quraan et al. (2021).

LOS links are necessary for mmWave and THz communications networks to provide sufficient receive signal power. Moving items in the surrounding area that block these LOS linkages may distort the communications channel or cause a sharp decline in link quality. This is because of the considerable penetration loss of mmWave and THz transmissions, resulting in low reliability and higher latency Charan and Alkhateeb (2022). The challenge of link blockage can be overcome by developing the sense of wireless networking sounding to anticipate the potential blockage and perform proactive handover (PHO). The traditional approach adopted for PHO is based on the combination of machine learning and wireless sensing (channel response, received signal strength) Choi et al. (2017); Alkhateeb et al. (2018). Furthermore, precise beam alignment can be achieved using advanced technologies like sensing, AI-driven position estimation, beam failure detection, real-time tracking, and proactive handover. However, acquiring reliable channel state information (CSI) is difficult due to the delay in measuring and reporting pilot-based sensing Lu et al. (2020). Therefore, AI-assisted sensing and positioning can transform the pilot-based channel acquisition into location-aware channel acquisition to simplify the optimal beam search. The challenge with traditional artificial intelligence (AI) driven beam and blockage prediction is associated with substantial control overhead and reactive management. Hence, the use of multi-modal data (wireless, vision, lidar, and GPS), the fusion of deep learning (DL) and computer vision (CV) is an emerging trend to solve the challenge of link blockage for future wireless networks Nishio et al., 2021. The fusion of multi-modality allows the wireless system to have a sense of the surrounding environment that is envisioned to play a vital role in future wireless communication, especially in blockage prediction, PHO,

and network resource allocation to ensure seamless connectivity Nishio et al. (2019).

1.1 Related work

Despite of numerous benefits, the true potential of mmWaves and THz can only be exploited by ensuring LOS link to maintain seamless connectivity. Hence, many solutions have been proposed to address the connectivity issues in future wireless networks. For instance, the studies in Choi et al. (2017); Alrabeiah and Alkhateeb (2020) demonstrated the use of wireless sensing data and advanced DL models to efficiently differentiate the LOS and blocked links. However, wireless sensing deals with blockage prediction problems from a reactive perspective, which is unsuitable for future wireless networks. Hence, the work Alkhateeb et al. (2018) took a step forward to take the beam blockage prediction as a proactive problem for successful handover in mmWave communication. The proposed scheme utilised a beamforming vector using Gated Recurrent Unit (GRU) for link blockage prediction. However, the uni-modal wireless data fails to meet future application requirements.

The use of multi-modality and fusion of DL and CV is an emerging area of research in vision-aided wireless communication. The use of multi-modality, especially vision sensing, provides information about the surrounding environment and brings awareness to wireless systems, which could potentially help in beam management and blockage prediction. Therefore, the authors in Alrabeiah et al. (2020b) developed vision wireless (ViWi), a deep learning framework for vision-aided wireless communication. The ViWi framework utilises wireless and vision-sensing data to provide a holistic view of the wireless network surroundings to predict the potential blockages proactively. The work in Alrabeiah et al. (2020a), utilises the received signal strength and RGB images to train a DL model using the ViWi framework. The proposed technique used a residual neural network (ResNet-18) for successful beam blockage prediction. The study in Charan et al. (2021b), proposed a novel DL architecture for future beam blockage prediction, leveraging the wireless beam sequences and their corresponding RGB images. The proposed solution combines the CV and DL to develop an understanding of wireless surroundings to highlight the potential of visual data for the future wireless network. Similarly, the authors in Charan et al. (2021a) developed a bimodal DL approach for proactive beam blockage prediction. The proposed scheme utilised object detection to determine the user's location by observing the consecutive RGB images in conjunction with wireless channels for blockage prediction to perform handover to maintain seamless connectivity.

1.2 Motivation and contributions

As discussed earlier, using multi-modal data and fusion of DL and CV is an emerging trend to solve connectivity issues in future wireless networks. It is envisioned that vision sensing allows the wireless system to develop a sense of its surroundings to predict the linkage blockage proactively. Despite its numerous advantages, the fusion of DL, CV and the use of multi-modality introduces many

challenges. For instance, the DL model for beam blockage prediction requires a massive amount of data for model training. In a future wireless system, BS equipped with vision sensors will be highly densified to provide maximum coverage in urban areas. Therefore, centralised model training requires raw data transfer to a centralised server, causing inefficient bandwidth usage, higher communication and storage costs with additional network footprints. Furthermore, the centralised inference mechanism also incurs unacceptable latency caused by the propagation delay.

Therefore, this paper proposes a latency-aware vision-aided federated wireless network (VFWN) for proactive beam blockage prediction using multi-modal data. The proposed framework uses federated learning (FL), a distributed learning paradigm for DL model training without any data sharing. This distributed learning mechanism offers privacy by design, on-device inference, and collaborative intelligence McMahan et al. (2017). Our proposed scheme consists of a federated server (FS) with multiple base stations (BS) equipped with vision sensors capable of data processing locally. In this approach, model parameters are shared with a centralised entity instead of sharing raw data to obtain an optimal global model. This distributed model training mechanism can potentially solve the cost, latency and scalability issues, especially in ultra-dense networks. Furthermore, an extensive comparative analysis in terms of predictive accuracy, energy efficiency and latency is done, and results are compared with centralised learning. To the best of the authors' knowledge, no prior work studied the impact of FL and performed energy efficiency and latency analysis, especially for vision-aided wireless communication. The main contributions of this work are as follows.

- Using multi-modal data, we propose a novel VFWN framework for beam blockage prediction for future 6G mmWave communication. The proposed framework has the ability to learn generalised DL models without data sharing to achieve scalability with massive user participation, solving communication costs and latency issues.
- For extensive analysis, the energy efficiency of transferring raw data in centralised training and model parameter sharing in FL training is done. Energy efficiency is the average electrical energy consumption for raw data or model parameters transfer *via* wireless link, measured in a kilo-watt hour per gigabyte (kWh/GB). Furthermore, latency analysis is done based on centralised and on-device inference.
- The performance of the proposed architecture is evaluated using the publically available dataset ViWi, which consists of bi-modal vision and wireless sensing data Alrabeiah et al. (2020b). The evaluation results confirm the importance of VFWN for beam blockage prediction with a significant improvement in energy efficiency using edge processing and low latency using on-device inference.

1.3 Paper organisation

The rest of the paper is organised as follows: Section 2 discusses the proposed VFWN, and Section 3 explains the simulation setup, and performance metric. Section 4 presents performance evaluation,

results and discussions, whereas Section 5 gives the conclusions and future work direction.

2 Proposed vision–aided federated wireless networK (VFWN) framework

This work considers a wireless communication system with edge nodes (EN) from the set of $n \in \{1, 2, 3, \ldots N\}$, where N is the maximum number of EN participating in the data collection and training process. Each EN is a wireless BS equipped with a vision sensor, capturing the surroundings of the wireless system at regular intervals. Furthermore, each BS or EN has an antenna array that enables the beam-forming technology to serve users by selecting the optimal beam. The system consists of a mobile user (car) and a potential blocking object (bus), which blocks the LOS link.

The EN has a local copy data $|D_n| \equiv D$, where D_n is the subset of the dataset at $n - th$ device and the entire dataset is given by $D = \sum_{n=1}^{N} |D_n|$. In the centralised mechanism for VAWC, the dataset D_n is shared with the centralised server for model training. This results in additional communication costs and latency. Hence, we propose VFWN, which uses distributed learning for model training. The system-level architecture of VFWN is shown in Figure 1. In our proposed framework, the EN are capable of processing and model training under the supervision of a federated server (FS), a centralised entity that orchestrates the training process. Each EN trains a DL model to obtain the model parameters w (weights and bias) using a local dataset. At the start of the training process, FS broadcasts the initial model to EN, starting the training process. The model parameters are updated on each EN using local training and shared with the FS for global aggregation. In order to obtain the optimal global model, the FS optimise the global cost function Wang et al. (2018), given as:

$$F(w) = \frac{\sum_{n=1}^{N} |D_n| \, F_n(w)}{\sum_{n=1}^{N} |D_n|}, \tag{1}$$

where $F_n(w)$ is the cost function on the $n-th$ device. The optimal global model parameters are obtained by achieving the minimum global cost function given by:

$$w^\star = \arg \min_{w} F(w). \tag{2}$$

To obtain the minimum in Eq. 2, an iterative model training process is adopted, which involves several communication rounds between the EN and FS.

2.1 Model training in proposed VFWN

The entire training and aggregation process consists of four main steps, as shown in Figure 1. In step 1, the FS initialises the global model in the first communication round and broadcasts it to EN, starting the training process. In the subsequent communication rounds, the FS broadcast the updated model parameters to EN. In step 2, the EN receives a copy of the global model and performs the local training using data present on the EN. In step 3, the local model parameters are shared with the FS for aggregation. It is worth mentioning that the model parameters obtained on each EN are

FIGURE 1
Proposed VFWN for beam blockage prediction for future wireless networks. In this system, we assume each BS acts as an edge nodes (EN) equipped with vision and wireless sensing to capture real-time data, capable of local processing and model training.

distant since the data on each EN is different from the others. In the final step, model aggregation is done to update the global model.

2.2 Local training and aggregation in VFWN

Local model training is similar to traditional model training, where the data from each EN is used for local training. Considering the vision and wireless sensing data, a convolutional neural network (CNN) is used, which consists of a 2D convolution layer with 3×3 feature detector, and 16 output filters. The next is the 2D max-pooling layer, followed by the flattened layer. The results of the flattened layer passed through a dense layer with 100 neurons and rectified linear unit (Relu) as an activation function, followed by a dropout layer and output layer with sigmoid activation. In local training, ADAM is used as an optimiser with a batch size of 32 and the hyper-parameter is kept the same on each device to make the system simple and computationally efficient.

Once the EN shares the local model parameters with FS, aggregation is done to obtain the optimal global model. In this study, we used federated averaging (FedAVG), one of the most efficient and commonly used aggregation algorithms in FL McMahan et al. (2016). In each communication round, multiple local epochs are executed on the subset of local data (small batches), reducing the frequent communication between the FS and EN. Once the global model is shared with EN, local training is done using the data on each client. The updated model parameters are shared with FS, where the model aggregation is done in the weighted averaging manner as given in Algorithm 1.

Required: T = Communication Round, N = Client Count, mini-batch size (B), local epochs (E)
1: **Server Execution:**
2: Initialise the server weights w_0
3: **for** Each round $t = 1...T$ **do**
4:     **for** Client count $n = 1, ...N$ **do**
5:         $w_n^t \leftarrow n.ClientUpdate(k, w_{t-1})$
6:     **end for**
7:     $w_t \leftarrow \sum_{n=1}^{N} \frac{n_n}{n} w_n^t$
8: **end for**

9: **ClinetUpdate(k,w):**
10: **for** Each epoch= 1,2,...,E **do**
11:     batches $\leftarrow$ dataset in B batches
12:     **for** b= 1,2,...,B **do**
13:         $w \leftarrow w - \eta \Delta l(w; b)$
14:     **end for**
15: **end for**
16: Send local w to server

Algorithm 1. Model Aggregation Algorithm in VFWN.

Note: In VFWN, once the FS obtain the optimal global model using FedAVG algorithm, the model is deployed on each EN for real-time on-device inference. The EN model takes vision and wireless sensing data as input and predicts the link, either LOS/NLOS using. In contrast to VFWN, the EN shares the data with a centralised server for inference; hence the proposed architecture ensures low latency using edge computing.

3 Simulation setup for VFWN

To evaluate the performance of the proposed VFWN, we utilise one of the scenarios in a publically available dataset ViWi Alrabeiah et al. (2020b). The ViWi is a parametric, systematic and scalable data framework for vision-aided studies that produce the customised dataset for DL model training. This framework uses wireless Insite ray tracing software and the 3D game engine Blender to generate high-fidelity synthetic data. The synthetic dataset consists of wireless and

TABLE 1 ViWi dataset samples distribution for VFWN model training. To evaluate the performance of the global model, a global testset of 1,000 data samples is used, which is never used in the training process. The dataset is almost balanced for both LOS and N-LOS scenarios.

	Training Samples	Testing Samples	Total Samples
EN-1	1173	297	1470
EN-2	1653	387	2040
EN-3	1174	316	1490
	4000	1,000	5000

vision sensing samples collected simultaneously, with different scenarios based on the camera locations (distributed or co-located) and the camera view (direct or blocked). The distributed data generation setup consists of three cameras mounted on three different BS covering the entire street with minimum overlapping. The outdoor scenario is simulated using Blender and wireless Insite to collect the visual and wireless datasets on predefined trajectories.

3.1 Dataset description and data distribution

The proposed VFWN framework utilises the distributed camera with a blocked view to generate the dataset for model training. The scenario considers an urban street having a wireless user (car), potential blockages (buses), and three BS equipped with vision sensors. Each vision sensor has a specific field-of-view (FoV) that captures additional information about the wireless system surroundings. The received signal strength and vision sensor data are used to categorise the user as either blocked (N-LOS) or LOS and provide ground truth for model training.

In this study, each BS is treated as EN (client) participating in the model training process. Therefore, the data (visual and wireless) captured at each BS is used for local training and never shared with the centralised server. Moreover, the entire dataset consists of 5000 RGB images with wireless channel state information, and each EN's data distribution is quite different, as given in Table 1. The global testset of 1,000 data samples is used to evaluate the performance of the proposed scheme. Furthermore, the results are compared with the traditional centralised approach, where data from each EN is shared with the central server for model training.

3.2 Performance metrics

To evaluate the performance of VFWN for beam blockage prediction, the average classification accuracy is used as the baseline metric. However, to avoid the accuracy paradox in the classification problem, other metrics like precision, recall and F1-score are also considered, which are given by:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{4}$$
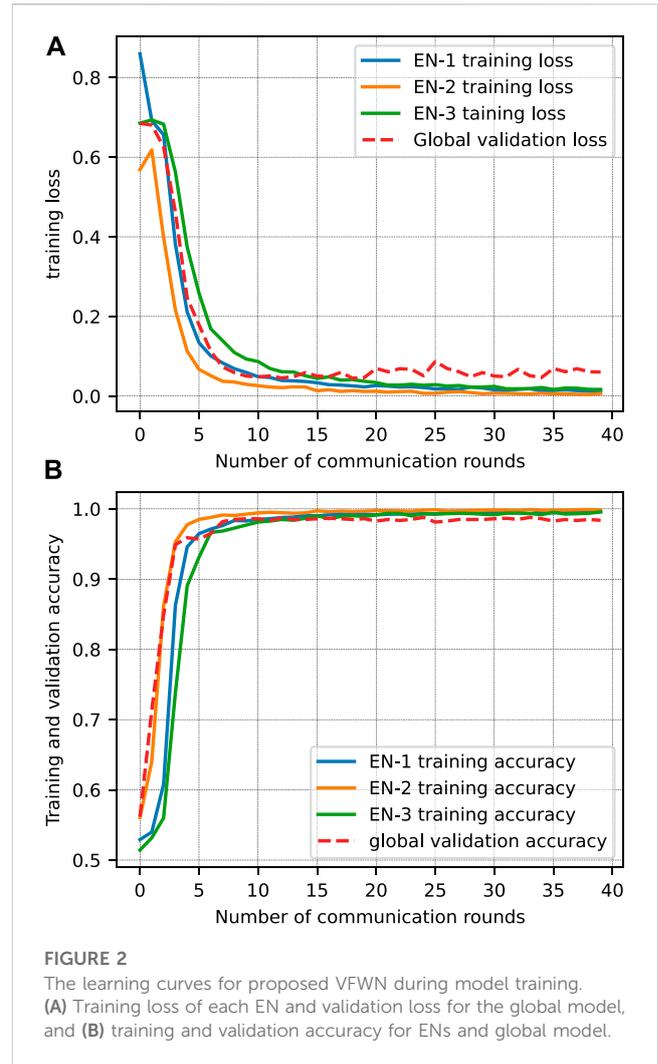
FIGURE 2
The learning curves for proposed VFWN during model training. **(A)** Training loss of each EN and validation loss for the global model, and **(B)** training and validation accuracy for ENs and global model.

and

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{5}$$

where TP denotes true positive, TN true negative, FP false positive, and FN false negative. Furthermore, we also used the confusion matrix (CM) to further elaborate the classification results, where rows of CM denoted the targeted labels, whereas the column represents the predicted labels. The TP of a class is in the diagonal of the confusion matrix, whereas the TN is the sum of all rows and columns, excluding the row and column of a particular positive class. Furthermore, to evaluate the efficiency of VFWN for beam blockage prediction problems, extensive analysis of communication cost and latency is done, and results are compared with a centralised model.

4 Performance evaluation, results, and discussions

As discussed earlier, the scenario under consideration consists of three BS equipped with a camera to cover the entire street. We

FIGURE 3
The confusion matrices **(A)** centralised learning, and **(B)** proposed VFWN.
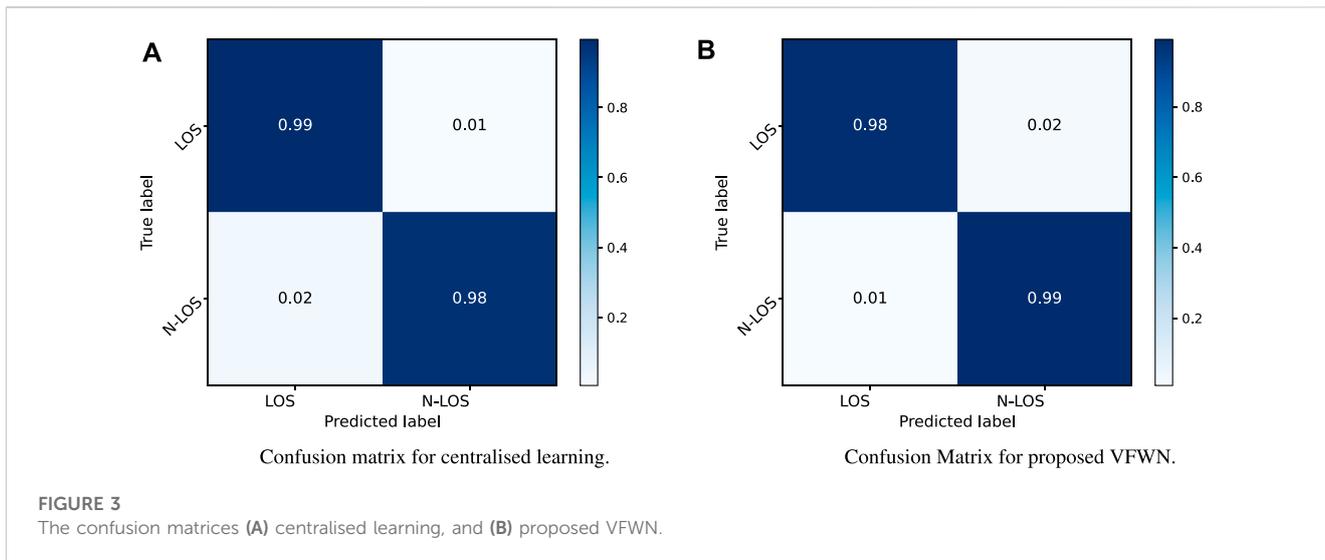
assumed that each BS has the local processing capability to train a DL model. Therefore, the FS initialise the global model and broadcasts to EN for local training. The EN trains the CNN using local data and shares the model parameters with FS for aggregation. This is an iterative process that continues until the model converges.

4.1 Blockage prediction results

In the FL setup, a CNN network is used for local training, consisting of a convolutional layer, a max-pooling layer, a dense layer and an output layer. The architecture of CNN is intentionally kept simple to make it computationally efficient with few model parameters. The FL training is done for 40 communication rounds with three local epochs, and a random search is performed to obtain the hyperparameters. The results in Figure 2, represent the learning curve for each client and a global model for each communication round. The FS evaluates the performance of the global model in each communication round using the global testset. The learning curve exhibits the classical behaviour with steep improvement in initial communication rounds until it reaches the slow monotonic behaviour after a few communication rounds. For instance, Figure 2A represents the training and validation loss of ENs and the global model, respectively. During the first ten rounds, there is a significant change in loss function with slow improvement in the remaining rounds. Furthermore, the similar trend of training and validation curves for both loss and accuracy represents the perfect model training.

We present the confusion matrices for both centralised learning and proposed VFWN to validate our results further. The testing of the proposed scheme is done on the global testset of 1,000 data samples, which are never used in the training process. The results in Figure 3A, present the confusion matrix of the centralised CNN model with an overall accuracy of 0.99% with average precision, recall and F-1 scores of 0.98, 0.99 and 0.99, respectively. Furthermore, the results in Figure 3B present the confusion matrix of the global model with an overall accuracy of approximately 0.985%. Moreover, the precision, recall and F-1 are recorded with the average value of 0.98, 0.989 and

TABLE 2 Comparative results for centralised learning and proposed VFWN for beam blockage predictions.

Technique	Accuracy	Precision	Recall	F1-Score
Centralised	0.99	0.98	0.99	0.99
FL	0.985	0.98	0.989	0.982

0.982, respectively. Although, accuracy is one of the commonly used metrics for a classification problem. However, the accuracy paradox in binary classification refers to the phenomenon where a model with high accuracy may not necessarily be the best model for the task. In such cases, the majority class predictions will dominate the overall accuracy. Therefore, to overcome the challenges, precision, recall, and F1-score are more informative metrics than accuracy. For instance, precision measures the proportion of positive predictions (blocked link) that are true positives, indicating that the model is not making many false positive predictions. Similarly, recall measures the proportion of actual positive examples (blocked links) that are correctly predicted by the model. High recall indicates that the model correctly identifies most of the positive classes. The result presented in Table 2 presents the comparative analysis of the proposed VFWN for beam blockage prediction. The model training is performed multiple times, and the results of each trial are averaged for both centralised and FL cases. The higher precision and recall scores indicate the generalised behaviour of the model with the ability to correctly identify the blocked and LOS links, resulting in fewer FP and FN.

From the results, it is evident that the FL is very effective for blockage prediction keeping the fact that no data sharing is done during model training.

4.2 Centralised vs. FL communication cost and latency analysis

After the accuracy comparison, we evaluate the performance of the proposed VFWN framework in the context of energy efficiency and latency. The cost of model training depends on multiple factors,

Communication cost comparison of centralised and FL setting for VFWN.

Latency comparison of centralised and FL setting for VEWN.

FIGURE 4
Comparision of communication cost **(A)** and latency analysis **(B)** for VFWN. The proposed scheme results in 81.31% reduction in communication cost and 6.77% reduction in inference latency.
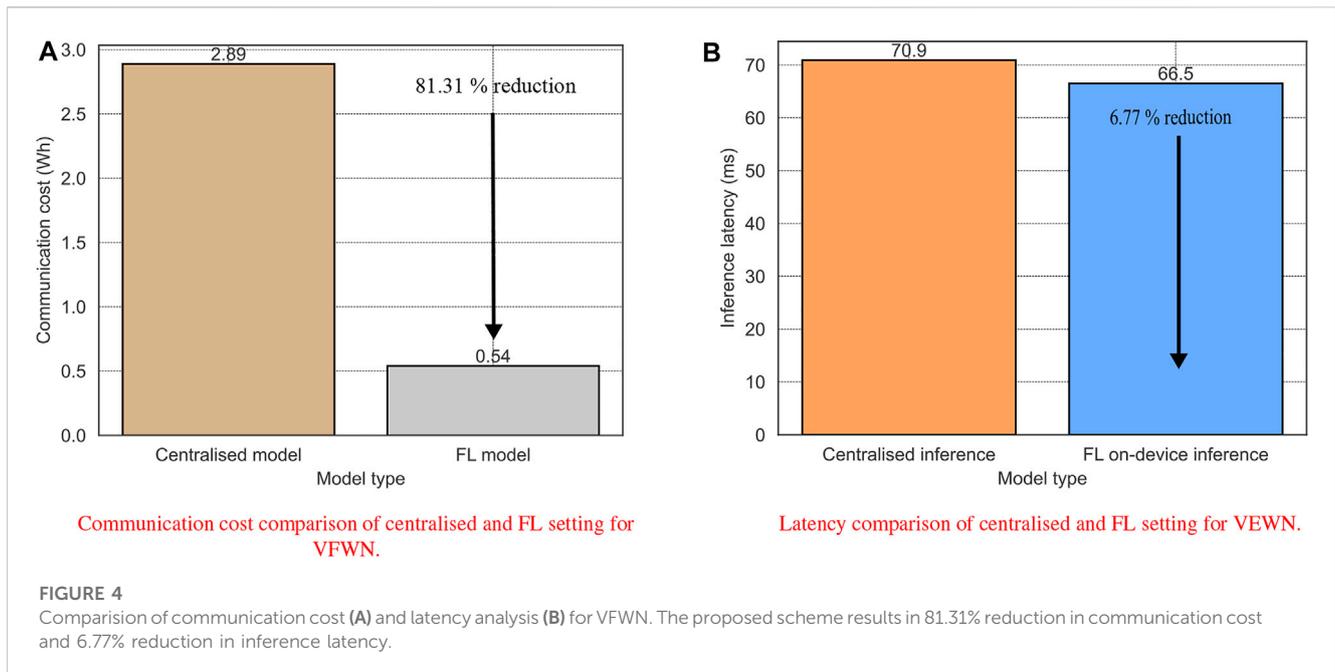
TABLE 3 Comparison of communication cost of centralised learning vs. proposed VFWN. The communication cost is computed in terms of average electrical energy consumption for wireless data transfer in kWh/GB.

Nodes	Data size (MB)	Data Tx cost (Wh)	Model size (MB)	Model energy cost (Wh)
EN-1	106	0.795	4.8	0.18
EN-2	193	1.447	4.8	0.18
EN-3	87.1	0.653	4.8	0.18
Total	386.1	2.89	14.4	0.54

including communication cost (data transfer), computation cost, and data storage cost. One of the effective metrics is energy consumption which mainly depends on communication (energy required to data or model transfer) and computation cost, which is proportional to computation time. The metric of energy consumption is given in Mian et al. (2022), mathematically represented as:

$$Energy = [\alpha \, (computation \, time) + \beta \, (data \, transfer)] \quad (6)$$

Where α is the computation constant and β is the communication constant. To simplify our analysis, we consider the communication cost only and take $\beta = 0.0075$. In energy efficiency comparison, the average electrical energy consumption for data transfer *via* a wireless network is considered measured as kilowatt-hours per gigabyte (kWh/GB). This study uses 0.0075 kWh/GB energy, which is the estimated using average energy value of 0.06 kWh/GB for UK in the year 2015, which is halved every 2 years Blenkinsop et al. (2021); Al-Quraan et al. (2021); Table 3 presents the extensive comparison of communication cost in terms of energy. In centralised model training, the raw data from each EN is shared with the centralised server, which incurs additional communication costs. For instance, the training dataset consists of 4,000 images with a

total size of 386.1 MB which resulted in a transmission cost of approximately 2.89 Wh using the average energy value of 0.0075 kWh/GB. In FL, the model parameters are shared during the training process instead of data sharing. In this case, the model size for each EN was approximately 300 KB which is shared for 40 communication rounds until the model converges. Therefore the model size for each EN for the FL training process was $300 \times 40 \times 2 =$ 24,000 (24 MB), where the factor of 2 is multiplied for bi-directional model sharing. The results in Figure 4A shows a significant improvement in energy efficiency with an overall 81.31% reduction in communication cost. Furthermore, the size of the local and global models is independent of the data samples. For instance, each EN has different training samples for model training; however, the model size for each EN was the same as each EN used the same model architecture shared by the FS.

The proposed VFWN framework uses on-device inference instead of centralised server inference for latency analysis. For instance, in the traditional vision-aided blockage prediction model, the EN captures the vision and wireless data and shares it with a centralised server for inference. The centralised inference incurs additional latency caused by the propagation delay of the communication link during data sharing. This study assumes a 10 Gbps mmWave back-haul link with a camera frame rate of

26 frames per second (fps) Nawaratne et al. (2019). The overall latency is calculated using the following equation:

$$D = t_{cap} + t_{tran} + t_{inf}, \tag{7}$$

where t_{cap} is the time taken by the vision sensor to capture two consecutive images, t_{tran} is the transmission time for sharing data with the centralised server, and t_{inf} is the model inference time. The value of t_{cap} is 38.5 m as the camera frame rate is 26 fps, and the t_{tran} is 4.4 m as the transmission time is approximately 2.2 m for image resolution 1,280 × 720 using 10 Gbps mmWave back-haul link Nawaratne et al. (2019); Al-Quraan et al. (2022). The t_{inf} in our proposed scheme is about 28 m; therefore, the overall delay in centralised inference using Eq. 7 is 70.9 m. In FL, the optimal global model is available on EN for on-device inference. Hence, t_{tran} will be zero in the proposed scheme, and the overall delay will be 66.5 m which results in 6.77% reduction is inference latency as shown in Figure 4B. Therefore, the proposed VFWN utilises the on-device inference, providing the additional benefit of low latency for time-sensitive applications. Furthermore, the t_{inf} also depends on the model architecture. For instance, a more complex model converges early; however, more model parameters increase the t_{inf}. Therefore, there is always a trade-off between the model inference time and model complexity.

5 Conclusion and future work

This work proposed VFWN for beam blockage prediction to maintain seamless connectivity in high-frequency communications. The proposed solution adopts a distributed learning approach to train a CNN with highly decentralised data, capable of proactive beam blockage prediction using vision and wireless sensing (consecutive RGB images and mmWaves beams). This approach shares model parameters with a centralised server instead of raw data, ensuring privacy by design, on-device inference, low communication cost, and collaborative intelligence. To evaluate the effectiveness of our solution, the results are compared with a centralised model training approach using a publically available synthetic dataset. The proposed scheme predicts the potential beam blockage proactively and achieves the accuracy of approximately 98.5% compared to 99% in a centralised approach. These results show that FL is very effective and achieves a similar level of accuracy without data sharing. Furthermore, the performance of the proposed technique is also evaluated in the context of energy efficiency and latency. In energy efficiency, the communication cost for data transfer in the wireless network is considered, which is measured as average electrical energy (kWh/GB). The results

demonstrate an overall 81.31% reduction in communication cost compared to centralised model training. Similarly, the FL-based on-device inference also reduces the overall latency by 6.77%. Hence, this study demonstrates the effectiveness of FL in the future wireless network. The focus of this work was solely on FL-based blockage prediction; therefore, an interesting future research direction would be proactive handover using the proposed scheme. Furthermore, the existing framework is designed for a single user only, which could be extended for more complex environments with multiple users.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://viwi-dataset.net/scenarios.html.

Author contributions

AK: Methodology, validation, software, investigation, formal analysis, writing–original draft. IA: Software, data analysis, writing, and editing. LM, SH, MI, RR, AZ: Conceptualization, validation, review, and editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Al-Quraan, M., Khan, A., Mohjazi, L., Centeno, A., Zoha, A., and Imran, M. A. (2022). *Intelligent beam blockage prediction for seamless connectivity in vision-aided next-generation wireless networks*. IEEE Transactions on Network and Service Management.

Al-Quraan, M., Khan, A. R., Mohjazi, L., Centeno, A., Zoha, A., and Imra, M. A. (2021). "A hybrid data manipulation approach for energy and latency-efficient vision-aided udns," in 2021 Eighth International Conference on Software Defined Systems (SDS), Gandia, Spain, 06-09 December 2021 (IEEE), 1–7.

Alkhateeb, A., Beltagy, I., and Alex, S. (2018). "Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff," in IEEE Global conference on signal and information processing (GlobalSIP), Anaheim, CA, USA, 26-29 November 2018 (IEEE), 1055–1059.

Alrabeiah, M., and Alkhateeb, A. (2020). Deep learning for mmwave beam and blockage prediction using sub-6 ghz channels. *IEEE Trans. Commun.* 68, 5504–5518. doi:10.1109/tcomm.2020.3003670

Alrabeiah, M., Hredzak, A., and Alkhateeb, A. (2020a). "Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction," in 2020 IEEE 91st vehicular technology conference (VTC2020-Spring) (IEEE), 1–5.

Alrabeiah, M., Hredzak, A., Liu, Z., and Alkhateeb, A. (2020b). "Viwi: A deep learning dataset framework for vision-aided wireless communications," in 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring) (IEEE), 1–5.

Blenkinsop, S., Foley, A., Schneider, N., Willis, J., Fowler, H. J., and Sisodiya, S. M. (2021). Carbon emission savings and short-term health care impacts from telemedicine: An evaluation in epilepsy. *Epilepsia* 62, 2732–2740. doi:10.1111/epi.17046

Charan, G., and Alkhateeb, A. (2022). *Computer vision aided blockage prediction in real-world millimeter wave deployments*. arXiv preprint arXiv: 2203.01907.

Charan, G., Alrabeiah, M., and Alkhateeb, A. (2021a). Vision-aided 6g wireless communications: Blockage prediction and proactive handoff. *IEEE Trans. Veh. Technol.* 70, 10193–10208. doi:10.1109/tvt.2021.3104219

Charan, G., Alrabeiah, M., and Alkhateeb, A. (2021b). "Vision-aided dynamic blockage prediction for 6g wireless communication networks," in 2021 IEEE International Conference on Communications Workshops (ICC Workshops) (IEEE), 1–6.

Choi, J.-S., Lee, W.-H., Lee, J.-H., Lee, J.-H., and Kim, S.-C. (2017). Deep learning based nlos identification with commodity wlan devices. *IEEE Trans. Veh. Technol.* 67, 3295–3303. doi:10.1109/tvt.2017.2780121

Khan, L. U., Saad, W., Niyato, D., Han, Z., and Hong, C. S. (2022). Digital-twin-enabled 6g: Vision, architectural trends, and future directions. *IEEE Commun. Mag.* 60, 74–80. doi:10.1109/mcom.001.21143

Lu, Y., Huang, X., Zhang, K., Maharjan, S., and Zhang, Y. (2020). Low-latency federated learning and blockchain for edge association in digital twin empowered 6g networks. *IEEE Trans. Industrial Inf.* 17, 5098–5107. doi:10.1109/tii.2020. 3017668

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2017). "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics* (Fort Lauderdale, Florida USA: PMLR), 1273–1282.

McMahan, H. B., Moore, E., Ramage, D., and y Arcas, B. A. (2016). *Federated learning of deep networks using model averaging*. arXiv preprint arXiv: 1602.05629 2.

Mian, A. N., Shah, S. W. H., Manzoor, S., Said, A., Heimerl, K., and Crowcroft, J. (2022). A value-added iot service for cellular networks using federated learning. *Comput. Netw.* 213, 109094. doi:10.1016/j.comnet.2022.109094

Nawaratne, R., Alahakoon, D., De Silva, D., and Yu, X. (2019). Spatiotemporal anomaly detection using deep learning for real-time video surveillance. *IEEE Trans. Industrial Inf.* 16, 393–402. doi:10.1109/tii.2019.2938527

Nishio, T., Koda, Y., Park, J., Bennis, M., and Doppler, K. (2021). When wireless communications meet computer vision in beyond 5g. *IEEE Commun. Stand. Mag.* 5, 76–83. doi:10.1109/mcomstd.001.2000047

Nishio, T., Okamoto, H., Nakashima, K., Koda, Y., Yamamoto, K., Morikura, M., et al. (2019). Proactive received power prediction using machine learning and depth images for mmwave networks. *IEEE J. Sel. Areas Commun.* 37, 2413–2427. doi:10.1109/jsac.2019.2933763

Wang, S., Tuor, T., Salonidis, T., Leung, K. K., Makaya, C., He, T., et al. (2018). "When edge meets learning: Adaptive control for resource-constrained distributed machine learning," in IEEE INFOCOM 2018-IEEE conference on computer communications, Honolulu, HI, USA, 16-19 April 2018 (IEEE), 63–71.