A Semantic-Oriented Federated Learning for Hybrid Ground-Aqua Computing Systems

Benedetta Picano, *Member, IEEE* and Romano Fantacci, *Fellow, IEEE,*

Abstract—Nowadays, an ambitious target of the next generation networks is to develop intelligent overarching space-air-ground-aqua computing systems, in order to provide a smart ecosystem able to efficiently operate computation in heterogeneous domains. In particular, in such a context, the underwater environment requires a special attention, since it is recognized as the most challenging domain, due to channel impairments and adverse propagation conditions. This paper proposes a self-intelligent system able to efficiently perform underwater environment monitoring or underwater survey of critical infrastructure, by resorting to the use of the semantic communication paradigm to lower the impairments due to the underwater channel propagation conditions. In particular, in our case, images sent by underwater devices are collected by shore small base stations (SSBSs) to form their training dataset to take part in a federated learning process with a ground base station. In particular, the paper considers a semantic communication scheme based on a deep-convolution neural networks encoder-decoder architecture for an efficient exploitation of the data transmission from underwater devices to the linked SSBSs. Performance analysis is provided to show the better behavior of the proposed system in comparison with the conventional alternative that does not involve the use of the semantic communications approach. Finally, a specific performance evaluation analysis is devoted to the investigation of the convergence behavior of the proposed federated learning procedure in reference to the cross ground-aqua system considered in order to highlight its advantages with respect to a classical implementation.

Index Terms—underwater communications, semantic communications, federated learning

I. INTRODUCTION

The emergence of the next generation networks is giving rise to a novel and wide class of applications, requiring to create an intelligent environment able to properly support disruptive new generation applications. During last decade, the emergence of these challenging aspects, alongside the ever increasing proliferation of mobile computing and Internet of Everything (IoE) devices, accordingly to which every single object on the Earth will be connected to the Internet for information exchange and communication, realizing intelligent and supervisory functions [1], has led to the need of developing seamlessly integration of systems operating in different domains, e.g., ground, air, space end even underwater [1]. Nowadays, this has generated an intense research efforts to push the artificial intelligence (AI) frontiers to the network edge regardless of the domain in which they operate [2], e.g., to perform communication channel conditions prediction [3]. This trend is in accordance with the emerging edge-intelligence (EI) paradigm which involves the use of edge

B. Picano and R. Fantacci are with University of Florence, 50139 Firenze, Italy, e-mails: {benedetta.picano, romano.fantacci}@unifi.it.

computing nodes (ENs), arranged in proximity of the source of data, to support data gathering and computations with the aim at hosting dedicated machine learning procedure to properly interpret and manipulate data stemmed from the network devices [4]. Nevertheless, EI is still at its infancy and the effective exploitation of AI techniques at network edges still represents a crucial problem. Recently, with the advent of 6G technology and the highfalutin applications which brings with it, has generated the imperative need to extend the unified AI-based network paradigm to space-air-ground-aqua integrated domains [1]. Such an innovative vision finds its roots in the fact that the about 71% of the Earth area is occupied by oceans, hosting critical infrastructures such as submarine pipelines, oil platforms to name a few, while large ocean areas are still unexplored. However, the effective realization of computations in the underwater domain is bound to data sensing, transmission, and forwarding, which makes the transmission of large volumes of data costly in terms of both time and power. In fact, such a novel computation paradigm implies that information is extracted under the water using embedded processors via data mining and/or data compression [5]. The main objective here is to include AI capabilities at the edge of a new generation integrated network (NGIN), to give rise to an EI-NGIN capable of information acquisition, processing, interpreting, and transmission in different domains and in relation to application purposes [1].

Whereas the interest in networks overlapping ground and underwater domains, to provide sustainable marine development and exploitation, is self-evident and it represents one of the most desideratum today's network challenges, there are still numerous issues to be addressed before it becomes a concrete reality. In particular, the limited communication conditions in the underwater environment dictates to decide between low transfer delay and short communication distance (expensive optical links), or longer communication delay with longer communication range (acoustic-based communications) [6]. The ambition of this paper is the proposal of an AI-driven framework to exploit *beyond-the-edge* computation capabilities, promoting the functional integration between machine learning and the device-centric approach, according to which end-devices play an active role in computing and communication. In this sense, the device-centric perspective is devoted to implement a novel AI based semantic communication scheme to mitigate the drawbacks due to the hostile behavior of underwater environments, then exploited to perform FL on the ground. This approach may enable important advancements in several application fields as ambient monitoring, conservation biology, and marine security. Note that, for the best of authors' knowledge, this is the first paper proposing a semantic

based communication framework to enable hybrid ground-aqua FL paradigm, which typically is exclusively confined to the ground domain.

In this picture, this paper considers the case of a proper ground-aqua EI-NGIN able to support the federated learning (FL) paradigm [7]–[9] to effectively train models on data (i.e., images) collected by underwater devices. In this context, the optimization of underwater data communications becomes crucial. Therefore, an AI-empowered semantic communications framework has been envisaged to extract and send only bits related to the semantic information of each collected images, instead of bits due to the statistic knowledge of source symbols. This involves a significant reduction in the amount of data to be transmitted and consequently a significant performance improvement in terms of transmission time reduction and data delivery reliability. Therefore, the contributions of the paper can be summarized as follows

- The design and development of a ground-aqua framework, applied to perform data gathering from underwater devices, to create image dataset at the SSBSs that perform FL with the ground SBS and, successively, to accomplish intelligent tasks. Image transmission among underwater devices and SSBSs is realized through the involvement of the semantic communications paradigm, in order to send only semantic information through the underwater channel, usually allowing slow transmission rates and subjected to severe communication impairments. It is important to highlight that existing frameworks devoted to realize collaborative federated analytics approaches, are typically confined to the ground domain exclusively, assuming the dataset on which they act yet gathered. Differently, this paper develops a cross-domain (i.e., aqua and ground) FL framework, considering and optimizing the underwater communication aspects involved in the data collection process;
- The proposal of a semantic communication scheme to overcome the criticalities of the underwater environment, enabling efficient and effective image transmission stemmed from the underwater devices, and devoted to cross-domain FL functioning. In this reference, a machine learning module based on the convolution neural networks (CNNs) has been applied, realizing the semantic encoder through the usage of a stack of Conv2D and max-pooling layer, whereas the decoder consists of a stack of Conv2D and Upsampling Layer;
- An in-depth performance analysis to test the proposed framework, in order to exhibit the validity of the approach adopted, especially focusing on performance measurements due to the underwater transmission, that represents the bottleneck of the whole system. Specifically, results investigate performance of the semantic encoder/decoder scheme, expressed in terms of loss function and underwater transmission time, and the performance of the whole FL-based framework, given in terms of convergence time and accuracy of the learning model trained.

The rest of the paper is organized as follows. In Section II an in-depth review of the related literature is presented. Section III details the system model and the problem considered, while Section IV presents the proposed overarching framework approach. Performance evaluations are presented in Section V and, finally, our conclusions are drawn in Section VI.

II. RELATED WORKS

The potentialities of the semantic communications in the next generation networks have been accurately discussed in [10], in which an in-depth survey about the application of machine learning techniques to the transmission systems to extract and retrieve the meaningful information has been proposed. Likewise, paper [11] investigates the applications of machine learning to the semantic transmission, considering the human-to-human, human-to-machine and machine-to-machine transmission modalities. The exploitation of EI to perform semantic transmission has been presented in [4], whereas a joint semantics-noise coding has been the object of the analysis addressed in [12], where a reinforcement learning approach has been applied. Furthermore, in [12], a critical discussion of the semantic approach has been provided, especially in relation to security and information overhead aspects. Differently, coexistence of heterogeneous tasks types has been the focus in [13], where an encoder architecture able to discern between classification and detection has been designed. Similarly, in [14] it was outlined the development of a dynamic context aware machine learning decoder to dynamically interpret the tasks type, aiming at performing proper semantic extraction.

A deep learning scheme has been applied in [15] to give rise to a semantic transmission system able to interpret and transmit natural language sentences. Moreover, in [15], the simultaneous maximization of both the cross-entropy and the Kullback-Leibler metric has been provided. Speech transmission has been the focus in [16], where a deep encoder-decoder neural network is developed minimizing the mean-square-error. In [17] a pruning redundancy method is propose to provide semantic communications in case of spectrum scarcity. In addition, fading channel effects have been considered, as well as channel state information aiming at mitigating the drawbacks of bad channel quality occurrence during transmission. The resource allocation problem has been analyzed in [18], where two novel semantic metrics have been proposed, namely the semantic rate and the semantic spectral efficiency. In addition, paper [19] presents a novel semantic networking concept, in which the semantic training has been performed thorough FL to support offloading strategies.

Similarly, the FL paradigm has been exploited in [20] to perform traffic classification, applying deep learning and the cross-silo horizontal technique to improve privacy and security. A FL has been tested in [21], aiming at evaluating the actual potentials of the FL technology having privacy as main focus. The FL has been also exploited in [22] to enable both the mobile intelligent surfaces reconfiguration and the users power allocation, in order to maximize channel quality, spectrum efficiency and users data rate. Authors in [23] focused on an edge computing landscape enabling tasks offloading. In that context, the FL has been applied to predict the execution time of the edge server considering an asymmetrical information

environment, aiming at optimizing the delay energy product metric, performing proper offloading policies.

Then, in [1], the hierarchical space-air-ground-underwater architecture is proposed, focusing on the protocol aspects of realizing a flexible and cross domain communications. Differently, a data-driven approach based on the usage of an echo-state-network has been provided in [24], in order to model the underwater channel conditions. The underwater domain is also the objective of paper [6] where Autonomous Underwater Vehicle (AUV)-assisted Underwater Wireless Networks are devoted to monitor and track underwater pollution. More in detail, authors propose a software-defined AUV networking system based on artificial potential field theory, devoted to network control to properly track underwater pollution. Underwater target tracking, along with target detection, is investigated also in [25], in which authors review several solutions to track unmanned underwater vehicles and they describe different ray tracing models, useful to perform tracking in underwater landscapes. A comprehensive survey about marine big data and corresponding data processing techniques, considering the problems typical of underwater environments, is represented by the paper [26]. The paper [27] focuses on AUV hybrid wireless network, where acoustic communications and magnetic induction coexist in the same network environment. The main contribution of that paper is the design and implementation of an alternating scheme to optimize AUV path planning and network data flow routing. Data collection problem within AUV sensor network is also the focus of paper [28], in which a data collection scheme integrating the AUV mobility is designed. In particular, the mobility model taken into account in [28] includes direction and velocity to produce a realistic 3D AUV mobility pattern. Exploiting AUVs as edge computing platform, authors propose a data collection and target node selection algorithm to effectively and efficiently visit all nodes in the network, balancing time and energy consumption. A cloud-based solution to real-time aquatic monitoring, involving underwater acoustic telemetry supported by edge computing paradigms, is proposed in [29]. The framework developed in [29] integrates a custom-designed, miniaturized, printed circuit board to compresse data and to increase transmission speed. A low-complexity algorithm to efficiently compute sound velocity, defined as the difference between the sound velocities at the transmitter and the receiver, for deep-sea Internet of Underwater Things (IoUT) networks, is developed in [30]. The paper [31] exploits stochastic geometry to optimize the densities of surface stations of a K-tier space–air–ground–sea underwater acoustic network, adopting realistic communication channel models, to maximize the coverage probability of the system. The design and realization of an efficient overarching Space–Air–Ground–Ocean network is also studied in [32], where a lightweight recurrent neural network is placed on board of low-consumption AUVs to self-navigate within the aquatic environment, safeguarding the limited AUV resources. The paper [33] applies contrastive learning within underwater networks to compress machine-friendly features under low bit-rates, devoted to underwater machine vision. Machine learning is also exploited in paper [34], whose main goal is to optimally set transmission parameters to avoid bandwidth
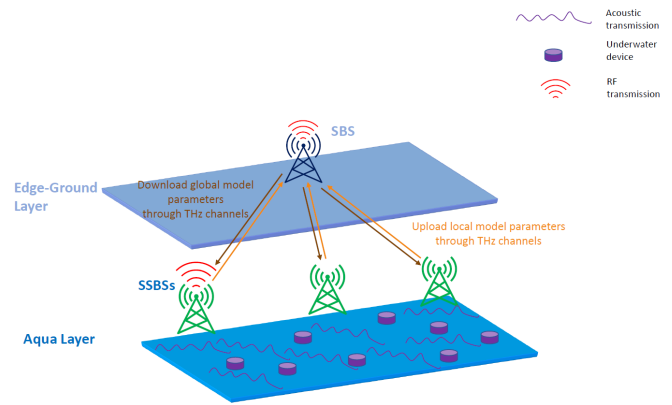

Fig. 1. FL over a Ground-Aqua Network Architecture

loss in underwater acoustic communications when compressed images are sent. In particular, the decision making selection strategy is achieved through reinforcement learning. Authors in paper [35] propose a comparison among machine learning techniques, properly tested on real underwater measurement gathered near the Gulf of Incheon, South Korea, to predict the most adequate communication parameters to mitigate the high propagation loss and drastic channel fluctuation problems.

In reference to the existing literature summarized before, this paper, for the best of authors' knowledge, is the first paper proposing the use of the semantic communications paradigm in an underwater environment, as well as a combined ground-aqua FL process.

III. PROBLEM STATEMENT

A. System Model

The scenario considers the application of the FL to a cross domain network arranged to classify images stemmed from underwater devices. In reference to Figure 1, a two layers EI-network has been considered. Whereas the first layer represents the aqua network domain, the second layer consists of the ground network level, in which a SBS $\mathcal{S}$ coincides with a cloud/edge network node (ENN), for which terms SBS and ENN will be used interchangeably. In the aqua domain we have a set $\mathcal{B} = \{1, ..., B\}$ of SSBSs, equipped with processing and storage capability, having the ability to handle semantic communications with a set of underwater-devices $\mathcal{M} = \{1, ..., M\}$, that collect the images to be used at the linked SSBSs to locally training the FL model in cooperation with the ENN and the other SSBSs. After completion of the FL training process the ENN broadcasts the trained model to each SSBS. Then, the images subsequently sent by the underwater-devices to the SSBSs according to the semantic communications paradigm are used to take, whenever necessary, intelligent actions. In particular, in our case, each underwater device, after capturing a new image, $\boldsymbol{s}_u$, exploits a semantic encoding process as described in Section IV, to extract the corresponding semantic information ι_u and, then, send it out to the linked SSBS.

B. Federated Learning Remarks

The FL [36], [37] represents a cooperative learning paradigm in which two type of agents, i.e, participants, are typically involved: the low level devices (generally end-devices), and a more structured and powerful node, for example a central unit or more in general an aggregator. End-devices individually perform a learning task by only send the local learning model to the aggregator instead of transfer the whole training data base. The FL is an iterative process, and it consists of global epochs, each of which, in its turn, split in three further phases: 1) local computation; 2) model exchange; 3) central computation. Due to the to the very hostile underwater environment such as propagation, absorption, and so on, the FL in this framework is performed between the SSBSs and the ENN. More specifically, the SSBSs represent the framework level typically attributed to end-devices, where the data on which the SSBSs train their models are gathered collecting the images sent by the underwater devices belonging to $\mathcal{M}$. Then, the central aggregator is represented here by the ENN, linked to the SSBSs through radio frequency channels.

In step 1) the SSBSs provide local data training considering the shared model previously download from the ENN, and the dataset created collecting images from devices belonging to the aqua layer following the approach presented in Section IV. Therefore, we denote with D_b the dataset of the SSBS b, with $b \in \mathcal{B}$, composed by the images received by all the underwater devices linked to it[1]. Consequently, for each sample j in D_b, the main goal is the identification of a model parameter w that minimizes the loss function $L_j(w)$. Therefore, each SSBS b solves the minimization problem [7]

$$\min_w \mathcal{L}_b(w) = \frac{1}{|D_b|} \sum_{j \in D_b} L_j(w). \qquad (1)$$

The corresponding learning model is represented by the minimization of the global loss function given by

$$\min_{w \in \mathbb{R}^e} \mathcal{L}(w) = \sum_{b=1}^{B} \frac{|D_b|}{\sum_{b=1}^{B} |D_b|} \mathcal{L}_b(w), \qquad (2)$$

in which e is the input size. Furthermore, during each local computation round t of the FL framework, the SSBS b solves the local problem [7]

$$w_b^{(t)} = \arg\min_{w_b \in \mathbb{R}^e} F_b(w_b | w^{(t-1)}, \nabla \mathcal{L}^{(t-1)}), \qquad (3)$$

in which F_b represents the objective function of SSBS b, $w^{(t-1)}$ is the global parameter produced during the previous iteration, and $\mathcal{L}^{(t-1)}$ is the global loss function at time $(t-1)$. Once each SSBS b has completed the local model training, the SSBS b uploads w_b^t to the ENN during the second step, in which the ENN collects the weights received by the SSBSs belonging to $\mathcal{B}$. In its turn, in step 3), the ENN improves the global model by performing the weighted average of the local updates w_b^t previously uploaded by the SSBSs.

---

[1]We have assumed that each underwater device sends gathered images to the nearest SSBS.

Hence, the ENN aggregates the received information by performing the following computations

$$w^{(t+1)} = \frac{1}{B} \sum_{b=1}^{B} w_b^{(t)}, \qquad (4)$$

and

$$\nabla \mathcal{L}^{(t+1)} = \frac{1}{B} \sum_{b=1}^{B} \nabla \mathcal{L}_b^{(t)}, \qquad (5)$$

This procedure is iteratively repeated , until the desired accuracy is achieved or as a consequence of a termination criterion such as the reaching of the maximum number of iterations. In reference to the model architecture of the FL, we adopted the vanilla FL averaging, consisting of a CNN designed as follows [38]

- two 5x5 convolution layers (with 32 channels and 64, respectively, and in cascade two 2x2 max pooling, one for each layer),
- one fully connected layer with 512 units and ReLu activation,
- one softmax output layer.

C. Channels and Computation Modeling

In reference to Figure 1, we have to take into account that the underwater channel has different propagation conditions than the wireless channel used for communications between SSBSs and the ENN. Hence, due to the heterogeneity of the nature of the channels involved, the link capacity characterization has to be specific of the channel considered, i.e., underwater or wireless[2]. In particular, the wireless connections between each SSBS and the ENN are accomplished on individual no-interfering channels in the THz band, characterized by a channel capacity, i.e., maximum data rate $\mathcal{R}_{b,\mathcal{S}}$ given by

$$\mathcal{R}_{b,\mathcal{S}} = \mathcal{W}_1 \log_2 \left(1 + SNR_t \right), \qquad (6)$$

$$SNR_t = \frac{P A_0 d_0^{-2} e^{-K(f)d_0}}{N_0} \qquad (7)$$

in which P is the transmission power assumed equal for both the SSBSs and ENN , whereas $\mathcal{W}_1$ represents the bandwidth of the communication link, d_0 is the distance, between the source and it destination and N_0, considering both the molecular absorption noise and the Johnson-Nyquist noise at the receiving site, results in

$$N_0 = \frac{\mathcal{W}_1 \zeta}{4\pi} g_B T_0 + P A_0 d_0^{-2} (1 - e^{-K(f)d_0}), \qquad (8)$$

where g_B consists of the Boltzmann constant, T_0 is the temperature in Kelvin, ζ is the wavelength, $K(f)$ the global absorption coefficient of the physical medium and $A_0 = \frac{c^2}{16\pi^2 f^2}$ [40]. In reference to the underwater acoustic channels, it is widely recognized that they are related to one of the most hostile communication media. As a consequence of

---

[2]Note that similarly to [39], this paper exclusively considers the uplink channel. In addition, in this paper interference issues are not taken into account.

this, it is clearly certificated in the literature that acoustic underwater communications can be suitable exploited only at low frequencies, usually allowing a limited communication bandwidth. An in-depth discussion of the underwater acoustic channel behavior and related modelling approaches is anyhow out of the scope of this paper. Any interested reader can find more details on this issue in [41] and references therein. Here, we limit our discussion to characterize the underwater acoustic channel capacity according to [42], [43] as :

$$\mathcal{R}_{m,b} = \mathcal{W}_2 \log_2\left(1 + \frac{P_s}{P_n}\right). \tag{9}$$

where $\mathcal{W}_2$ is the signal bandwidth, P_s and P_n denote the mean received signal power and overall noise power, respectively, within $\mathcal{W}_2$, at the receiving side.

In reference to the computation model, each SSBS b has on board a CPU with working frequency f_b, given in number of CPU cycles per unit time. Therefore, the time needed by the SSBS b to perform the local model computation is

$$t_b = \log\left(\frac{1}{\epsilon_b}\right) f_b D_b, \tag{10}$$

where $\log\left(\frac{1}{\epsilon_b}\right)$ is the number of local iterations needed to achieve the local accuracy ϵ_b [44], [45] with respect to problem (1). Let v_b be the local parameter size expressed in bits associated to the SSBS b, each communication round exhibits a cost in time defined as

$$\tau_b = \frac{v_b}{\mathcal{R}_{b,\mathcal{S}}} \tag{11}$$

Therefore, considering the b-th SSBS, denoting with N the number of communications rounds, the total amount of time spent results in

$$T_b = N(t_b + \tau_b) + \gamma_b, \tag{12}$$

where γ_b represents the time needed to collect data from underwater devices. More in detail, defining with $\mathcal{K}$ the total number of steps necessary to gather images, the overall time spent by the SSBS to populate the whole dataset can be defined recursively as follows.

$$\gamma_b = \sum_{\chi=1}^{\mathcal{K}} \nu_\chi, \tag{13}$$

where

$$\nu_\chi = \begin{cases} \max\left\{\frac{u_{\chi,m}}{\mathcal{R}_{m,b}}\right\}, & \text{if } \chi = 1, \\ \max\{\nu_{\chi-1}, \nu_\chi\}, & \text{if } \chi > 1, \end{cases}$$

and $u_{\chi,m}$ is the size in bits of the data sent by the underwater device m towards the associated SSBS b, at step χ.

Note that terms γ_b and N are not related to each other. More in depth, γ_b refers to the time spent to gather data in the underwater domain, and it depends on factors such as the size of the semantic information to be transmitted, the rate of the underwater channel, the numerosity of the dataset we want to build, etc. Differently, N is the number of FL rounds which are related to both the local accuracy ϵ_b and the convergence accuracy of the FL we desire to reach. Note
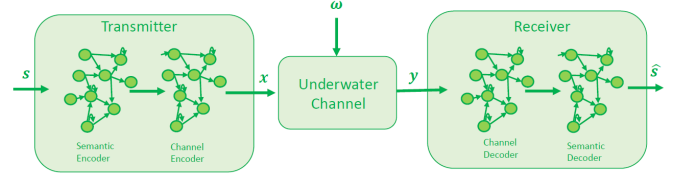

Fig. 2. Underwater semantic communications framework

that the number of SSBSs is determined in dependence on the number and position of underwater devices to which SSBSs have to provide coverage. In this respect, it is important to highlight that the SSBSs have a triple role:

- To provide coverage to underwater devices. Therefore, the choice about the number of SSBSs deployed is typically ascribed to a prior phase of network design.
- To gather data deriving from underwater device, in order to build the dataset exploited by the FL framework.
- To perform the FL with the ground SBS and the other SSBSs involved in the process. In this reference, note that the number of SSBSs does not directly impact on the convergence time of the FL. In fact, such a time is the result of multiple factors, for example data distribution, size of dataset, and so on.

D. Problem Formulation

With the aim at optimizing the FL framework for hybrid aqua-ground domain, the minimization of the mean overall time, i.e., the time needed to receive data from the underwater layer plus the time spent to actually training the model is crucial. Therefore, the main objective of the paper is to design an overarching framework able to optimize the following problem

$$\min \frac{1}{B} \sum_{b=1}^{B} T_b \tag{14}$$

The optimization problem given in (14) is very challenging, due to the strong influence provided by terms γ_b on T_b values[3], with $1 \leq b \leq B$. Therefore, in what follows, we will focus on the minimization of terms $\frac{u_{\chi,m}}{\mathcal{R}_{m,b}}$, by proposing a semantic communication framework that aims at the successful transmission of semantic information extracted from an image instead of a set of symbols or bits regardless of their meaning.

IV. UNDERWATER SEMANTIC COMMUNICATIONS

In the semantic communications, the main goal is represented by the transmission of the semantic meaning of the source data instead of the whole data. The key difference with classical systems is the use of a semantic encoder (Figure 2) at the transmitter site able to extract the semantic features. As a consequence, semantic features exclusively are sent out, whereas data, once received at the destination, are processed on a semantic basis only by the receiver site, instead of at the bit level as in traditional systems. As depicted in

---

[3]As deeply certified by literature, the convergence time of the FL framework, once the dataset is built, is in the order of milliseconds [7], [46].
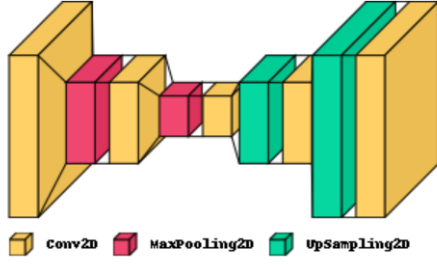
Fig. 3. Underwater semantic communications architecture

Figure 2, the semantic communications framework consists of two individual cascaded sub-modules: the semantic and the transmission part.

The first is responsible of semantic encoding and decoding performing information processing and semantic extraction, whereas the transmission level focuses on correctly transmitting the semantic information over the physical channel. The underwater channel is notoriously afflicted by numerous impairments and bad propagation conditions. For this reason, many efforts have been made during years to reduce such drawbacks by identifying suitable modulation procedures, in particular, to lower problems due to the low data-rate. However, in reference to this, we have to point out that an in-depth discussion of underwater communication techniques is out of the scope of this paper. As a consequence, in the following, we will refer to the multicarrier binary frequency shift keying (MC-BFSK) scheme, recently proposed in [42], just as an example of a promising scheme for handle underwater communications without pretending to propose a definitive solution for this issue. Hence, in order to summarize the main features of the MC-BFSK scheme[4] we can say that, accordingly to [42], $\frac{M}{2}$ parallel and independent subcarriers from a single transducer has been considered, resulting into a composite signal where each signaling instance carries $\frac{M}{2}$ channel coded bits [42]. From results reported in [42], it emerges that by a suitable selection of the modulation scheme parameters an error-free data transmission may be guaranteed for signal-to-noise ratio (SNR) about greater than 2 dB [42]. Consequently, the errors due to channel impairments in our paper have been assumed negligible. Before transmission, the picture $\mathbf{s}$ is mapped into symbols $\mathbf{x}_m$ in order to be sent over the physical channel, experiencing transmission impairments due to sea environment, which implies that the symbol received at the receiver $\mathbf{y}_{m,b}$ is subjected to wireless channel impairments [16], [17].

The framework of interest here focuses on the image-transmission and consists of an encode-decoder architecture based on deep CNNs. More in depth, the semantic encoder is realized through the usage of a stack of Conv2D and max-pooling layer, whereas the decoder consists of a stack of Conv2D and Upsampling Layer, whose architecture is described in Figure 3. As depicted in Figure 2, accordingly to literature, the encoder part is formed by a semantic and

[4]Any interested readers can refer to [42] and refrences therein for a more insightful discussion of this topic.

a physical coding modules. Let $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ be the neural network parameters set for the channel and the semantic encoder modules, respectively. Therefore, the encoded symbol $\mathbf{x}_m$ can be expressed as

$$\mathbf{x}_m = \mathcal{C}_{\boldsymbol{\alpha}_1}(\mathcal{S}_{\boldsymbol{\alpha}_2}(\mathbf{s})), \tag{15}$$

where $\mathcal{C}_{\boldsymbol{\alpha}_1}(\cdot)$ and $\mathcal{S}_{\boldsymbol{\alpha}_2}(\cdot)$ express the channel and the semantic encoder functions, respectively. Then, $\mathbf{y}_{m,b}$ is decoded by the receiver, with the aim at retrieving the original $\mathbf{s}$. As a consequence, the retrieved copy of $\mathbf{s}$, $\hat{\mathbf{s}}$, results in

$$\hat{\mathbf{s}} = \mathcal{S}_{\boldsymbol{\beta}_1}^{-1}(\mathcal{C}_{\boldsymbol{\beta}_2}^{-1}(\mathbf{y}_{m,b})), \tag{16}$$

in which $\mathcal{S}_{\boldsymbol{\beta}_1}^{-1}(\cdot)$ is the semantic decoder with parameter $\boldsymbol{\beta}_1$, whereas $\mathcal{C}_{\boldsymbol{\beta}_2}^{-1}(\cdot)$ is the channel decoder having parameter $\boldsymbol{\beta}_2$. Since the sea environment usually exhibits deep channel fluctuations and adverse propagation conditions, the primary objective here is to design an integrated channel-semantic coding able to maximize the similarity between $\hat{\mathbf{s}}$ and $\mathbf{s}$. Referring to [4], [15], a common loss function adopted is represented by the binary cross-entropy metric that, with the Adam optimizer, aims at giving insight about the similarity between $\hat{\mathbf{s}}$ and $\mathbf{s}$. In order to have a measurement of the goodness of the encoding-decoding procedure provided, for a underwater device m, with $m \in \mathbf{M}$, transmitting towards the SSBS b, with $b \in \mathcal{B}$, the cosine images similarity metric [47] has been exploited. Hence, the similarity between the image sent by underwater device m toward the SSBS b results in

$$\zeta_{m,b} = \frac{\boldsymbol{s} \cdot \hat{\boldsymbol{s}}}{||\boldsymbol{s}||||\hat{\boldsymbol{s}}||}, \tag{17}$$

that is the ratio between the dot product of the images expressed as vectors, and the product of L2-norm of both the vectors. The meaning of parameter $\zeta_{m,b}$ is of a feedback metric to catch the level of validity and accuracy provided by the transmission system. In fact, a $\zeta_{m,b}$ decrease due to hash propagation conditions, has a negative impact on the semantic communication quality. Note that the semantic framework exhibits a worst case complexity driven by the presence of the convolution operation. Therefore, its computational complexity is mainly in the order of $\mathcal{O}(U \cdot V \cdot u \cdot v)$, where $U \cdot V$ and $u \cdot v$ express the size of the original image and the size of the kernel applied during convolution.

V. PERFORMANCE ANALYSIS

This section provides a performance analysis of the proposed system whose focus is the functional integration of a FL process with an underwater semantic communications system. As stated before in Sec.IV, being the investigation of the data underwater transmission schemes out of the scope of this paper in its present form, without loss of generality, we have referred here, as an example, to the MC-BFSK underwater data communication technique proposed in [42] and referring for the system parameter values to those reported therein in Table I. In particular, the information data transmission rate assumed in our computer simulations is derived in accordance with Eq. (6) of [42], that depends on the modulation profile selected in accordance with the procedure outlined in Sec. II of the same
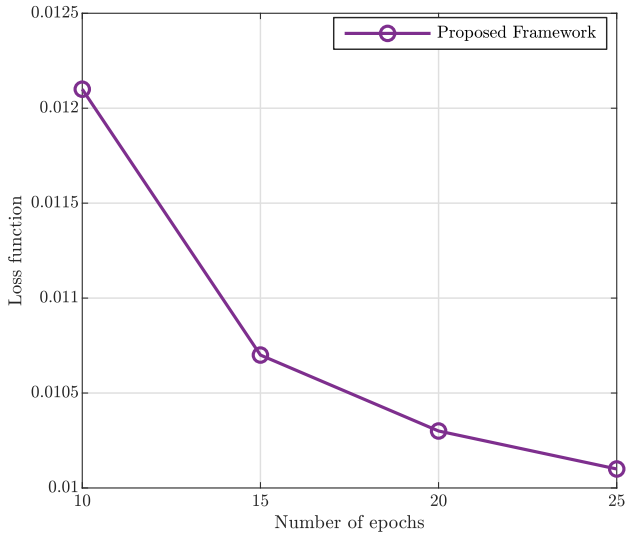
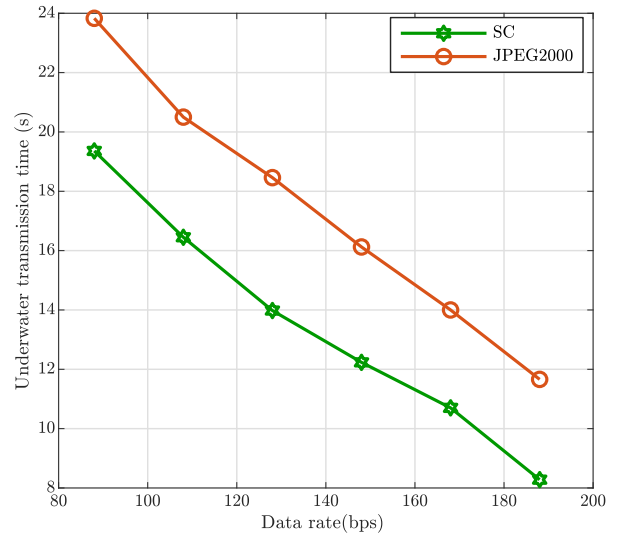Fig. 4. Loss function as a function of the number of epochs.

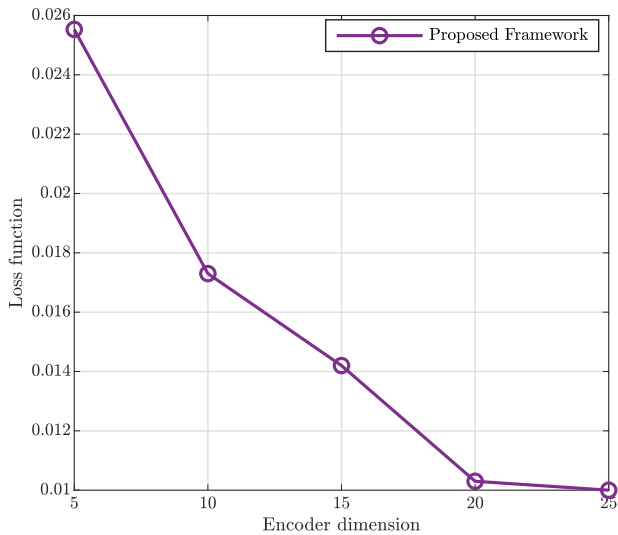Fig. 6. Underwater transmission time as a function of the data rate.

Fig. 5. Loss function as a function of the encoder dimension.
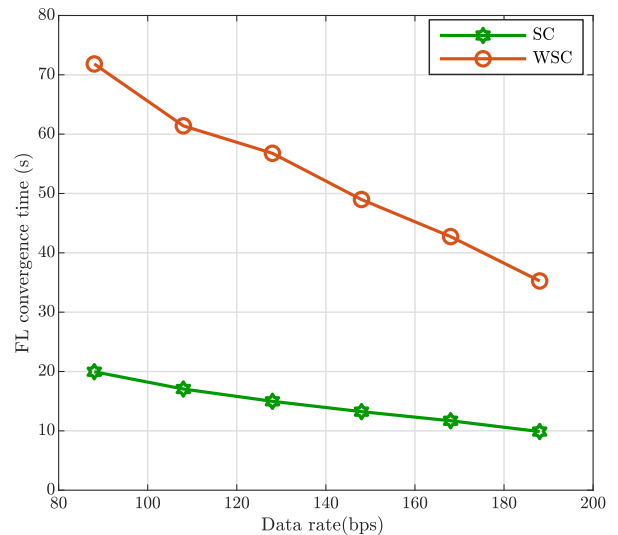

Fig. 7. FL convergence time as a function of the data rate.

paper. Furthermore, in order to properly test the performance of the proposed framework, we have resorted to the *MNIST handwritten digits* dataset [48], considering images resolution of 28 x 28 pixels, where one pixel has been assumed composed by 8 bits. The 70% of the dataset is devoted to the training phase, and the 30% is used for testing. We have assumed a number of SSBSs involved in the FL process equal to 15. Such an assumption is justified by the fact that, in our scenario, it represents the best trade-off between the accuracy reached performing the FL and the number of SSBSs taking part in the process (a greater number of SSBSs does not remarkably improve accuracy, whereas a lower number of SSBSs gets worse accuracy values). Aiming at validating the proposed scheme (SC curve), Figure 4 shows the loss function value as

the number of learning epochs grows. As it is straightforward to note in this figure, the loss function decreases, i.e., the performance of the learning model improves, considering high values of the number of epochs. The performance behavior depicted in Figure 4 is also confirmed by the increase of the encoder dimension considered in Figure 5 in relation to the achieved loss function performance. Also in this case, the loss trend grows as the number of neural unit increases. This is clearly due to the fact that as the encoder dimension grows, the intelligence capability of the considered deep-CNNs network increases. Then, Figure 6 shows the ability of the SC approach in performing image size reduction, in comparison with the JPEG2000 compression technique, as a function of data rates reported in [42]. As it is evident to

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3325289
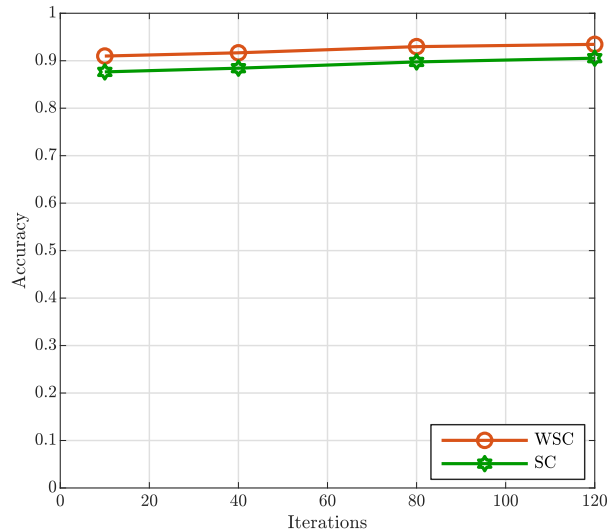
8

Fig. 8. Accuracy reached by the FL model as a function of the number of iterations.

note, the SC curve remarkably lowers the transmission time, representing a useful solution to reduce the impact of the underwater environment on the FL framework. Differently, Figure 6 illustrates the trend of the overall FL convergence time as a function of the data rates, comparing the scheme applying the SC and that in which the SC are not involved (WSC curve). When the semantic communications are not applied (WSC curve), the image is sent in the conventional method, without extracting and sending exclusively the se-mantic information. As it is evident to note from the Figure 7, the time required to send data when the semantic commu-nications (SC curve) approach is used, is significantly lower than the time spent with the conventional method, implying a significant impact on the overall FL completion time, which means an improvement on the practical applicability of the training framework. The behavior due to the results provided in Figure 7 becomes clear evident when cross-analyzed them with those given in Figure 8, dealing with the accuracy gap in comparison with the ideal case achieved by resorting to the considered FL scheme, i.e., by performing the semantic extraction/reconstruction process (WSC curve). As it is evident from this Figure, the accuracy is not significantly afflicted by the semantic extraction/reconstruction process, and, as a consequence, the FL process reaches almost the same accuracy values of the scheme exploiting original images. Consequently, jointly considering Figure 7 and Figure 8, it is easy to note that the convergence time improvements due to the semantic extraction-reconstruction process does not significantly afflict the accuracy of the overall framework, while it introduces significant benefits in terms of lowering the amount of data to be transmitted, thus speeding up the FL convergence process. Hence, as a final remark, we can say that the results provided in this section highlight the concrete possibility to implement an EI-ground-aqua network based on the exploitation of the semantic communications to counteract the hostile behavior of

the underwater communication channels without a significant impact on the accuracy of the retrieved data and allowing a strong reduction of the FL convergence time that is of paramount important in several application cases.

VI. CONCLUSION

This paper has investigated the problem of enabling the he FL paradigm execution in an integrated ground-aqua environ-ment by exploiting data stemmed from underwater devices. In this reference, a ground-aqua network, mainly devoted to monitoring activities based on images transmissions, has been considered, in order to properly perform the FL framework. In addition, the semantic communication approach has been in-troduced to face with the hostile sea propagation environment and long data transmission delays. For this purpose, a deep-CNNs architecture has been designed, and its performance are validated by assuming a proper underwater channel model. Finally, performance analysis has been provided, in order to exhibit the validity of the overarching framework proposed, especially in terms of the accuracy and overall time needed by the considered FL process. Future works may include the integration of AI-based techniques to forecast underwater channel behaviour in order to adapt to it the selection of the most suitable data transmission scheme.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] J. Liu, X. Du, J. Cui, M. Pan, and D. Wei, "Task-oriented intelligent networking architecture for the space–air–ground–aqua integrated net-work," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5345–5358, 2020.
[2] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1738–1762, 2019.
[3] B. Picano and R. Fantacci, "An intelligent radio buffer-aided relaying scheme with adaptive link selection," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3677–3684, 2021.
[4] W. Yang, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Cao, and K. B. Letaief, "Semantic communication meets edge intelligence," 2022.
[5] H. Albarakati, R. Ammar, and R. Elfouly, "Real-time underwater com-puting system," in *2018 IEEE Symposium on Computers and Commu-nications (ISCC)*, 2018, pp. 00 890–00 896.
[6] C. Lin, G. Han, J. Jiang, C. Li, S. B. H. Shah, and Q. Liu, "Underwater pollution tracking based on software-defined multi-tier edge computing in 6g-based underwater wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 2, pp. 491–503, 2023.
[7] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019, pp. 1387–1395.
[8] Y. Xiao, G. Shi, and M. Krunz, "Towards ubiquitous ai in 6g with federated learning," 2020.
[9] J. Konecny, H. B. McMahan, F. X. Yu, P. Richtarik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2017.
[10] Q. Lan, D. Wen, Z. Zhang, Q. Zeng, X. Chen, P. Popovski, and K. Huang, "What is semantic communication? a view on conveying meaning in the era of machine intelligence," *Journal of Communications and Information Networks*, vol. 6, no. 4, pp. 336–371, 2021.

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3325289

9

[11] ——, "What is semantic communication? a view on conveying meaning in the era of machine intelligence," *Journal of Communications and Information Networks*, vol. 6, no. 4, pp. 336–371, 2021.

[12] K. Lu, Q. Zhou, R. Li, Z. Zhao, X. Chen, J. Wu, and H. Zhang, "Rethinking modern communication from semantic coding to semantic communication," *IEEE Wireless Communications*, pp. 1–13, 2022.

[13] Q. Wu, F. Liu, H. Xia, and T. Zhang, "Semantic transfer between different tasks in the semantic communication system," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 566–571.

[14] Y. Zhang, H. Zhao, J. Wei, J. Zhang, M. F. Flanagan, and J. Xiong, "Context-based semantic communication via dynamic programming," *IEEE Transactions on Cognitive Communications and Networking*, pp. 1–1, 2022.

[15] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, 2021.

[16] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," 02 2021.

[17] H. Xie and Z. Qin, "A lite distributed semantic communication system for internet of things," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 142–153, 2021.

[18] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Y. Li, "Resource allocation for text semantic communications," *IEEE Wireless Communications Letters*, pp. 1–1, 2022.

[19] G. Shi, Y. Xiao, Y. Li, and X. Xie, "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *IEEE Communications Magazine*, vol. 59, no. 8, pp. 44–50, 2021.

[20] U. Majeed, S. S. Hassan, and C. S. Hong, "Cross-silo model-based secure federated transfer learning for flow-based traffic classification," in *2021 International Conference on Information Networking (ICOIN)*, 2021, pp. 588–593.

[21] S. K. Das and S. Bebortta, "Heralding the future of federated learning framework: Architecture, tools and future directions," in *2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, 2021, pp. 698–703.

[22] R. Zhong, X. Liu, Y. Liu, Y. Chen, and Z. Han, "Mobile reconfigurable intelligent surfaces for noma networks: Federated learning approaches," *IEEE Transactions on Wireless Communications*, pp. 1–16, 2022.

[23] F. Sun, Z. Zhang, S. Zeadally, G. Han, and S. Tong, "Edge computing-enabled internet of vehicles: Towards federated learning empowered scheduling," *IEEE Transactions on Vehicular Technology*, pp. 1–17, 2022.

[24] O. Onasami, M. Feng, H. Xu, M. Haile, and L. Qian, "Underwater acoustic communication channel modeling using reservoir computing," *IEEE Access*, vol. 10, pp. 56 550–56 563, 2022.

[25] H. Ghafoor and Y. Noh, "An overview of next-generation underwater target detection and tracking: An integrated underwater architecture," *IEEE Access*, vol. 7, pp. 98 841–98 853, 2019.

[26] M. Jahanbakht, W. Xiang, L. Hanzo, and M. Rahimi Azghadi, "Internet of underwater things and big marine data analytics—a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 904–956, 2021.

[27] D. Wei, C. Huang, X. Li, B. Lin, M. Shu, J. Wang, and M. Pan, "Power-efficient data collection scheme for auv-assisted magnetic induction and acoustic hybrid internet of underwater things," *IEEE Internet of Things Journal*, vol. 9, no. 14, pp. 11 675–11 684, 2022.

[28] S. Cai, Y. Zhu, T. Wang, G. Xu, A. Liu, and X. Liu, "Data collection in underwater sensor networks based on mobile edge computing," *IEEE Access*, vol. 7, pp. 65 357–65 367, 2019.

[29] Y. Yang, R. Elsinghorst, J. J. Martinez, H. Hou, J. Lu, and Z. D. Deng, "A real-time underwater acoustic telemetry receiver with edge computing for studying fish behavior and environmental sensing," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 821–17 831, 2022.

[30] T. Zhang, G. Han, L. Yan, and Y. Peng, "Low-complexity effective sound velocity algorithm for acoustic ranging of small underwater mobile vehicles in deep-sea internet of underwater things," *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 563–574, 2023.

[31] J. Xu, M. A. Kishk, and M.-S. Alouini, "Coverage enhancement of underwater internet of things using multilevel acoustic communication networks," *IEEE Internet of Things Journal*, vol. 9, no. 24, pp. 25 373–25 385, 2022.

[32] Z. Li, J. Wen, J. Yang, J. He, T. Ni, and Y. Li, "Energy-efficient space–air–ground–ocean-integrated network based on intelligent autonomous underwater glider," *IEEE Internet of Things Journal*, vol. 10, no. 11, pp. 9329–9341, 2023.

[33] Z. Fang, L. Shen, M. Li, Z. Wang, and Y. Jin, "Prior-guided contrastive image compression for underwater machine vision," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 6, pp. 2950–2961, 2023.

[34] W. Su, J. Tao, Y. Pei, X. You, L. Xiao, and E. Cheng, "Reinforcement learning based efficient underwater image communication," *IEEE Communications Letters*, vol. 25, no. 3, pp. 883–886, 2021.

[35] J. Byun, Y.-H. Cho, T. Im, H.-L. Ko, K. Shin, J. Kim, and O. Jo, "Iterative learning for reliable link adaptation in the internet of underwater things," *IEEE Access*, vol. 9, pp. 30 408–30 416, 2021.

[36] V. Smith, C. Chiang, M. Sanjabi, and A. Talwalkar, "Federated multi-task learning," vol. abs/1705.10467, 2017. [Online]. Available: http://arxiv.org/abs/1705.10467

[37] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," vol. abs/1902.04885, 2019. [Online]. Available: http://arxiv.org/abs/1902.04885

[38] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," 2023.

[39] T. H. Thi Le, N. H. Tran, Y. K. Tun, M. N. H. Nguyen, S. R. Pandey, Z. Han, and C. S. Hong, "An incentive mechanism for federated learning in wireless cellular networks: An auction approach," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 4874–4887, 2021.

[40] C. Chaccour, M. Soorki, W. Saad, M. Bennis, and P. Popovski, "Can terahertz provide high-rate reliable low latency communications for wireless vr?" 04 2020.

[41] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 84–89, 2009.

[42] P. Zetterberg, F. Lindqvist, and B. Nilsson, "Underwater acoustic communication with multicarrier binary frequency shift keying," *IEEE Journal of Oceanic Engineering*, vol. 47, no. 1, pp. 255–267, 2022.

[43] P. van Walree, R. Otnes, and T. Jenserud, "The watermark manual and user's guide,."

[44] L. U. Khan, M. Alsenwi, Z. Han, and C. S. Hong, "Self organizing federated learning over wireless networks: A socially aware clustering approach," in *2020 International Conference on Information Networking (ICOIN)*, 2020, pp. 453–458.

[45] D. Chen, C. S. Hong, L. Wang, Y. Zha, Y. Zhang, X. Liu, and Z. Han, "Matching-theory-based low-latency scheme for multitask federated learning in mec networks," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11 415–11 426, 2021.

[46] C. S. Hong, L. U. Khan, M. Chen, D. Chen, W. Saad, and Z. Han, *Federated Learning for Wireless Networks*, 01 2021.

[47] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," in *2016 4th International Conference on Cyber and IT Service Management*, 2016, pp. 1–6.

[48] L. Deng, "The mnist database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.