IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 10, 2020, accepted March 2, 2020, date of publication March 19, 2020, date of current version May 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2981973

Link Adaptation on an Underwater Communications Network Using Machine Learning Algorithms: Boosted Regression Tree Approach

ALAMGIR M S M, MST. NAJNIN SULTANA, (Member, IEEE), AND KYUNGHI CHANG, (Senior Member, IEEE)
Department of Electronic Engineering, Inha University, Incheon 22212, South Korea

Corresponding author: Kyunghi Chang (khchang@inha.ac.kr)

This work was supported in part by the Ministry of Oceans and Fisheries, South Korea, through the Project titled Development of Distributed Underwater Monitoring and Control Networks.

ABSTRACT Interest in the study of next-generation underwater sensor networks for ocean investigations has increased owing to developing concerns over their utilization in areas such as oceanography, commercial operations in maritime areas, and military surveillance. Underwater acoustic communications (UAC) network channels are fast-varying (spatially and temporally) according to environmental conditions. It is tempting to use adaptive modulation and coding (AMC) for UAC networks to improve the system efficiency by matching transmission parameters to channel variations. This paper focuses on analyzing a measured sea trial dataset by using a rule-based strategy (i.e., three-dimensional analysis, modulation-wise analysis, and a fixed-*SNR* strategy) to find the suitable link adaptation procedure depending on the channel quality. Hence, we plot a scenario of the measured UAC network *data rate* vs. *Signal to Noise Ratio* (*SNR*) and/or *Bit Error Rate* (*BER*) to pick the best AMC combinations in the context of adaptivity to the channel. Due to non-reversibility limitation of rule-based strategy, the work further extends to use machine learning (ML) algorithms to classify the MCS levels by investigating the channel characteristics. Boosted regression tree, from among the four ML algorithms we adopted for the analysis, shows formidable accuracy of 99.97% in classifying MCS levels. This ensemble of trees learns from the uplink data of the buoy and the base station and relates the MCS levels to channel metrics and signal characteristics especially subject to *SNR* and *BER* constraints.

INDEX TERMS Boosted regression tree analysis, K-nearest neighbors, link adaptation, machine learning, pseudo-linear discriminant analysis, support vector machine, UAC network.

I. INTRODUCTION

Next-generation underwater networks have great potential for observing and exploring the aquatic environment. Therefore, underwater acoustic communications (UAC) is broadly considered to be the only approach feasible for long-distance underwater communications, and it has been extensively adopted in various scenarios. The demand for quality of service arises in many militaries, scientific, and civilian applications, including communications between submarines, for underwater security surveillance, for scientific data collection at ocean-bottom stations, in offshore oil explorations by autonomous underwater vehicles, and for data exchange in underwater sensor networks for environmental monitoring [1]. To mitigate the distance problem work is done in the worst case, a Stackelberg game is modeled to analyze the behaviors between macrocell base station (MBS) and femtocell users (FUEs) [2]. The MBS sets interference prices to FUEs to maximize its utility and guarantee its transmission rate requirement. Based on the interference prices, FUEs determine their powers to optimize their utilities subject to the delay constraint. This proposed robust power allocation and pricing problem are to work with distance uncertainty. But for UAC, there is no such research work for solving this issue. Again, it is critical to deal with the underwater acoustic channel's challenging properties, such as long delay and large

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed Farouk.

Doppler spread, resulting in frequency and time-selective fading. Moreover, fading, Doppler spread, and multipath propagation severely impact UAC network performance [3], [4]. Channel with significant Doppler spread can effectively be handled by coupling a channel equalizer mitigating Intercarrier Interference (ICI). The compressed sensing algorithms in the form of Orthogonal Matching Pursuit (OMP) and Basis Pursuit (BP) algorithms perform better than the Least Squares Channel estimator when the channel is sparse. It also utilizes overcomplete dictionaries with an increased path delay resolution [5]. In one investigation [6], the authors analyze the energy efficiency and spectral efficiency of direct D2D, multi-hop D2D along with the traditional mobile communication system by utilizing the two-timeslot physical-layer network coding scheme under Rayleigh fading channels. They derive close approximations of the average energy efficiency and average spectral efficiency and the optimal power of UEs maximizes the energy efficiency. Unlike typical wireless channels, the transmission loss in a UAC network not only depends on the distance between transmitter and receiver but also on the signal frequency. Absorption loss, the transfer of acoustic energy into heat, is caused by the signal frequency [3]. Hence, the underwater channel presents formidable challenges when comparing it with the typical terrestrial channel model.

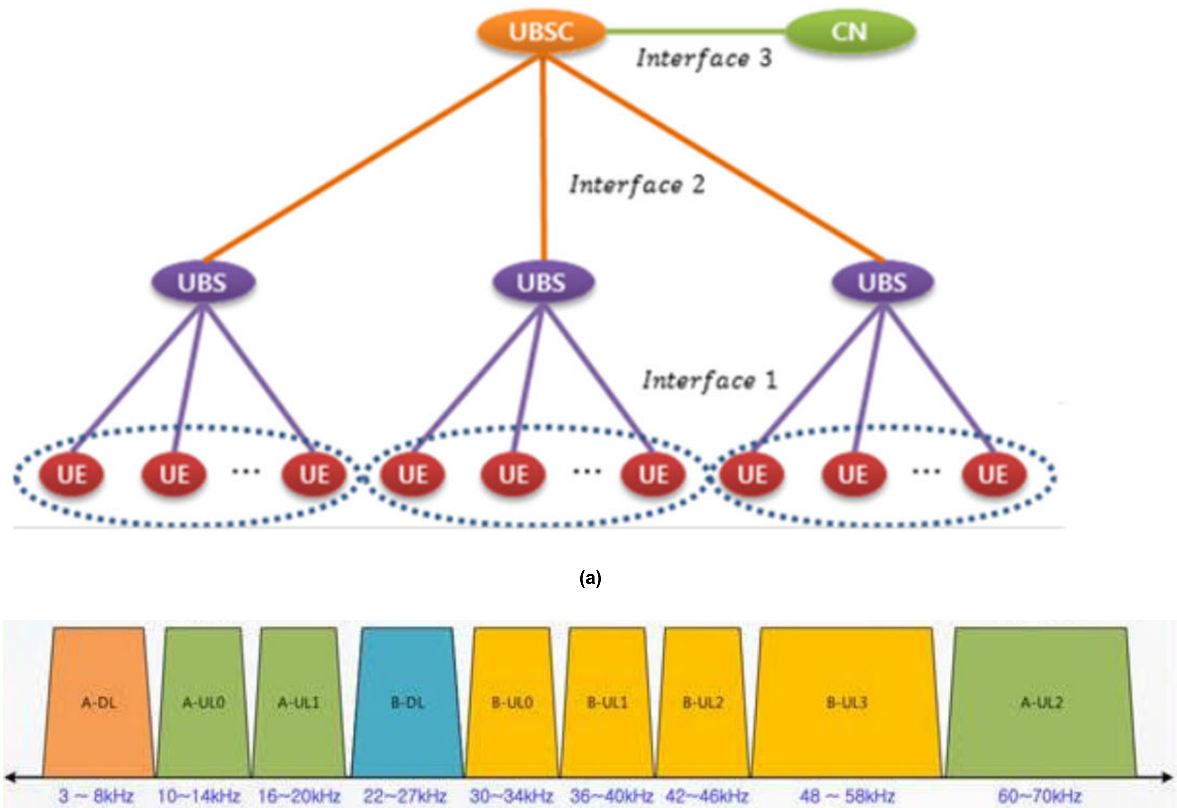
Due to the extreme limitations on the available bandwidth in a UAC network, frequency reuse and cellular concepts are more appealing to enhance the coverage and capacity for a UAC network [7]–[9]. A simulation-based link adaptation (LA) approach [10] has used a 12-path Rician fading channel for strategical LA. There are just a few parameters, i.e., modulation levels with a fixed coding rate and repetition patterns, considered in this approach for a widely varying UAC network channel. Also, the priority rules, have been formulated to discard the overlapping modulation levels from the feasible ones, are based on certain assumptions. This limits the applicability of that approach from a practical perspective. Also, the widely varying channel makes it very challenging to formulate a channel model for the UAC network, unlike terrestrial network scenarios, i.e., the knife-edge diffraction model, the Longley–Rice model, the Okumura model, etc. In [11], the authors propose a modem equipped with direct sequence spread spectrum signals of various coding rates and modulation orders. They estimate the channel through decision feedback equalizer and achieved high spectral efficiencies subject to a combination of *bit-error-rate* (*BER*) and *signal to noise ratio* (*SNR*) constraints. Adaptive selection of signals is achieved based on their *BER* prediction via boosted trees. It speeds up the communication 10-20 times than fixed-rate transmission. But this proposed receiver shows vulnerability due to the high time-variable (high Doppler spread) channel. A software-defined OFDM based underwater acoustic (UWA) communication system demonstrates the superiority of adaptive transmission where both modulation order/type and power on each subcarrier are selected based on channel conditions to maximize the throughput but it experimentally limited to short-distance transmission set up [12]. If the distance increases throughput decreases accordingly due to higher attenuation and it observes severe multipath effects. For the unknown and dynamically variable underwater channel, an optimal self-learning strategy (reinforcement learning) is implemented, which is a significant area of machine learning [13]. Due to slow sound velocity in an underwater scenario, acknowledgment from receiver to transmitter is a great concern, which deteriorates the convergence speed of the algorithm. So, the authors use a juggling-like ARQ mechanism with RL to minimize the long delay feedback problem.

Another research work has been made by taking *BER* as input and the implemented modem balances exploration of the search space against the exploitation of existing knowledge, optimizes for the average *data rate*, instead of searching for maximum possible *data rate*. It improves the average *data rates* than the previous rate [14].

Due to the complexity and instability of underwater acoustic communication systems, it is difficult to identify modulation during actual communication. There are some new methods of underwater acoustic communication such as sparse adaptive convolution cores, time-domain turbo equalization, and frequency-domain turbo equalization have been used, but these methods still have the problem of high computational complexity and low classification success rate. By applying Deep Learning, classification accuracy has found around 99% [15].

In this paper, by addressing the above-mentioned difficulties we apply machine learning (ML) to classify the modulation and coding scheme (MCS) levels. We focus on the analysis of two measured datasets [stored in one drive: link [16]] in a UAC network that has different larger distances than before, around 1 km, 2 km, and 3 km, between transmitter and receiver. Also, the datasets contained a wide range of measured channel parameters. This research aims to develop a link-adapted system that will be able to boost spectral efficiency in response to any channel conditions while retaining the desired level of covertness and reliability. We compare the throughputs for given two distances with combined effects of *SNR* and *BER* for different rule-based strategies such as 3-D analysis, modulation-wise analysis, and fixed-*SNR* strategy, etc. Then we classify the modulation type by applying several reliable ML algorithms including boosted regression tree in [11], because of its better performance, on our datasets subject to *SNR* and *BER* constraints and measure the accuracy.

In the rule-based analysis, we can see that, despite our best efforts, none of the strategies give any strong outcome due to the non-reversibility issue. Our measured sea trial datasets show that the channel has a colossal number of variations over the periods, e.g., morning to noon, day to night, summer to winter, etc. Besides, there are additional parameters, e.g., *coherence bandwidth* (BW_{coh}), *pilot spacing (PS)*, *maximum estimated delay (MED)*, and *BER*, to be considered and



(a)

(b)

FIGURE 1. (a) UAC network layout. (b) UAC network operating frequency and bandwidth.

measured during the sea trials, which contribute to channel quality measurement and our ML classification. So, the statistical ML algorithm–based analysis for LA prevailed over the rule-based LA approach.

In Section II, we discuss the UAC network architecture. In Section III, we provide the dataset overview. Section IV explains the rule-based strategy such as the three-dimensional analysis, followed by modulation-wise analysis, and the fixed-*SNR* strategy. Section V explains the reason why we need ML in a UAC network. Section VI provides insights into ML in the UAC network. MCS classification results using ML are shown in Section VII. Finally, conclusions are drawn in Section VIII.

In Figure 1(a), the underwater base station controller (UBSC) is connected to the core network (CN) through the interface 3 connection. Our main concern is UBSC, which is connected with three underwater base stations (UBSs) via interface 2, and these three UBSs are further connected with underwater equipment (UE) through interface 1. Interface 1 and 2 are wireless links where the acoustic wave of 100 Hz is used as carrier Interface 3 is LTE. Figure 1(b) shows the system operating frequency and bandwidth where A-DL to A-UL1, and A-UL2 connections are allocated to the UBSC and UBSs, and B-DL to B-UL3 are allocated to the UBSs and UE, respectively. The connection parameters' details are listed in Table 1.

II. UAC NETWORK ARCHITECHTURE

There are many contributions to cellular-based network topologies, as well as frequency reuse concepts in terrestrial networks [17]–[21]. The tremendous achievements of cellular networks are enough motivation to consider the cellular and frequency reuse concept in a UAC system. The UAC network is still in the development stages, and that is why a lot of attention and meaningful research is required in this area. Because of bandwidth limitations, the cellular-based network structure for a UAC network is appealing [22].

III. DATASET OVERVIEW

The analysis is based on two kinds of the measured datasets. In both cases, uplink data are measured (A-UL0) i.e. the transmitters are UBSs and the receiver is UBSC and FDD is used. The first dataset has been collected from August 17, 2018, to August 20, 2018. The experiment has been carried over two separate distances, i.e., 1 km, and 3 km between the transmitter (Tx) and receiver (Rx) in Mohang Port (Taean-gun).

The latitude and longitude coordinates for 1 km and 3 km distances were: Tx—36°54'43.13''N 126°12'21.53''E

TABLE 1. UAC network parameter details.

Channel	Frequency Band	Bandwidth	Connection Details
A-DL	3–8 kHz	5 kHz	Shared channel among 3 UBSs at max. of 10 km
A-UL0	10–14 kHz	4 kHz	Uplink channel at max. of 10 km
A-UL1	16–20 kHz	4 kHz	Uplink channel at max. of 7–8 km
A-UL2	60–70 kHz	10 kHz	Uplink channel at max. of 1 km
B-DL	22–27 kHz	5 kHz	Shared channel for UE at max. of 5 km
B-UL0	30–34 kHz	4 kHz	Uplink channel at max. of 5 km
B-UL1	36–40 kHz	4 kHz	Uplink channel at max. of 2–4 km
B-UL2	42–46 kHz	4 kHz	Uplink channel at max. of 1–2 km
B-UL3	48–58 kHz	10 kHz	Uplink channel at max. of 1 km

and Rx—36°54’11.83”N 126°11’27.90”E, and Tx—36°54’44.43”N 126°12’37.33”E and Rx—36°54’05.40”N 126°11’18.83”E, respectively. The second is from a 24-hour experiment from July 5, 2017, to July 6, 2017. The experiment has been carried out over a 2 km distance between the Tx and Rx in the Yellow Sea (Deokjeokdo, Incheon). The latitude and longitude coordinates were Tx—37°13’28.2”N 126°12’12.0”E and Rx—37°14’22.7”N 126°12’55.1”E.

There was a total of 288 combinations of different kinds of parameters, and 417 sets of experimental data in the Taean dataset [16], which is a good reason to take a step-by-step approach in selecting the suitable modulation and coding scheme (MCS) levels. The process to analyze the Taean dataset has been divided into three stages, which included three-dimensional analysis, modulation-wise analysis, and a fixed-*SNR* strategy. Unlike the Taean dataset, the Incheon dataset has 108 combinations of parameters and 10,105 sets of experimental data. This enormous dataset makes it impossible to analyze with any kind of strategy other than the ML approach.

A. TAEAN DATASET

The sea trial experiment has been conducted over two separate distances with the same parameters, 1 km, and 3 km. The dataset contains 11 parameters with different combinations, i.e., *pilot spacing* (*PS*), *repetition pattern* (*RP*), *coding rate* (*CR*), *modulation level* (*Mod*), *SNR*, *uncoded* & *coded bit error rate* (*BER*), *data rate*, *maximum estimated delay* (*MED*), *root mean square* (*RMS*) *delay spread*, and *coherence bandwidth* (BW_{coh}). By doing permutation and combination of *PS*, *RP*, *Mod*, and *CR*, we have measured the value of other parameters such as *SNR*, *uncoded* and *coded BER* and so on with different times in a day. The transmitted criterion parameter details are shown in Table 2.

The symbols ‘F’ and ‘T’ in *PS* and *RP* mean Frequency and Time. We have 417 sets of experimental data with different

TABLE 2. Taean dataset parameters.

Parameter	Values
Pilot Spacing (F, T)	(2, 1), (4, 2), (6, 3)
Repetition Pattern (F, T)	(1, 1), (2, 1), (4, 1), (8, 1), (1, 3), (2, 3), (4, 3), (8, 3), (1, 9), (2, 9), (4, 9), (8, 9)
Error Correction Coding	Turbo Coding
Coding Rate	(1/2), (1/3)
Modulation	BPSK, QPSK, 16QAM, 64QAM

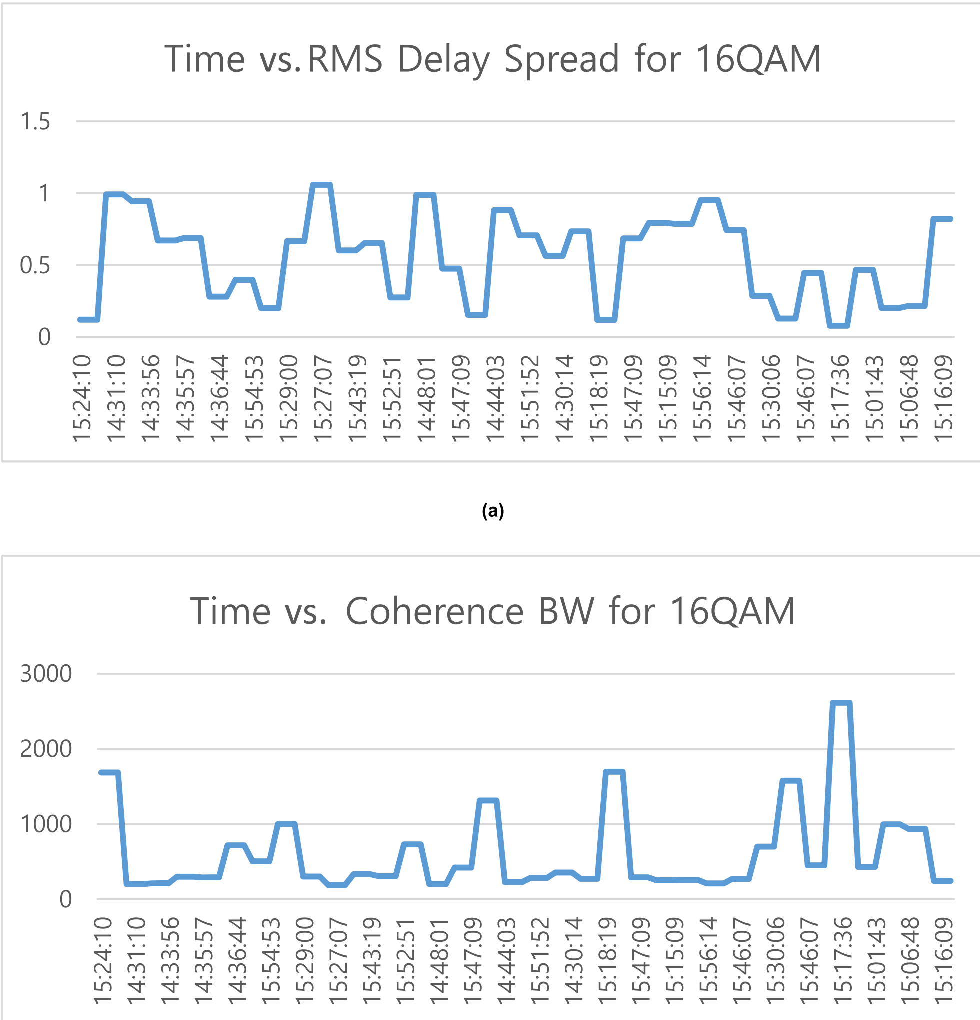
TABLE 3. Incheon dataset parameters.

Parameters	Values
Pilot Spacing (F, T)	(2, 1), (4, 2), (6, 3)
Repetition Pattern (F, T)	(1, 1), (2, 1), (4, 1), (8, 1), (1, 3), (2, 3), (4, 3), (8, 3), (1, 9), (2, 9), (4, 9), (8, 9)
Error Correction Coding	Turbo Coding, Convolutional Coding
Coding Rate	(1/3)
Modulation	QPSK, 16QAM, 64QAM

parameters [16]. The objective is to find the most suitable MCS level that provides the best performance according to channel quality among these 417 trials. Channel quality measurement can be obtained from the parameters, i.e., *MED*/*RMS delay spread*, BW_{coh}, *BER*, etc.

B. INCHEON DATASET

This dataset contained 12 parameters, i.e., *PS*, *RP*, *modulation level (Mod)*, *SNR*, *uncoded* & *coded BER*, *data rate*, *MED*, *RMS delay spread*, BW_{coh}, *Doppler spread* (*DS*), and *frequency shift*. The transmitted parameter details are shown in Table 3, which contains modulation level, error correction coding, and different combinations of frequency and time-domain values for *PS*, and *RP*. Depends on different combinations of these parameters over time we measured the received other parameters like *SNR*, *BER*, *RMS delay*, *frequency shift* and so on. Channel quality measurement can be obtained from the parameters, i.e., *MED*/*RMS delay spread*, BW_{coh}, *BER*, etc. The dataset has been organized into 10,105 rows and 18 columns, in which 17 columns contain numeric values and one column is for modulation level names. The rows characterize the number of experiments, and columns depict the number of features. Every row is populated with the input parameters: *PS*, *RP*, *SNR*, *uncoded BER*, *coded BER*, *data rate*, *MED*, *RMS delay spread*,



(a)

(b)

FIGURE 2. Statistically varying channel of Taean dataset: (a) Time vs. *RMS Delay Spread,* (b) Time vs. *Coherence BW* for 16QAM.

BW_{coh}, *DS*, and *frequency shift*, as well as the output parameter: *modulation level names*.

Besides, both datasets show that the channels are statistically varying over time. Figure 2 shows such shreds of evidence.

IV. RULE-BASED STRATEGY

A. THREE-DIMENSIONAL ANALYSIS

The goal is to compare three parameters at the same time in the three-dimensional analysis because for obtaining a link adaptation curve for all modulations, we need to analyze these parameters. We initially aimed to consider *SNR*, *data rate*, and *BER* as the deciding factors for the Taean 1 km dataset. We chose *BER* as the third parameter since the *RMS delay spread* causes *BER* to vary. So, *BER* could be a valid parameter for the analysis. Figure 3 visually represents these three parameters, where the X, Y, and Z axes represent *SNR* in decibels, the *data rate* in bits per second, and *BER*, respectively. Figure 3 shows that most MCS pairs are in the blue region, and those have a low *BER* at mostly less than a 1000 bps *data rate*. Then, the scattered points slowly build up, indicating fewer MCS pairs in the higher *BER* region. Figure 3 shows that *BER* keeps slowly increasing within a *data rate* range of 0–1000 bps. Then, it keeps fluctuating within the range of 1000–2000 bps. The fluctuation rate is slowly decreasing after 2000 bps, since there are fewer MCS pairs in that region, which means fewer MCS pairs with a higher *data rate*. From Figure 4, we can see that most MCS pairs that are in the *SNR* region higher than 16 have a lower *data rate*. The *SNR* region from 13.86 to 16.43 has MCS pairs with a higher *data rate*. Furthermore, analysis for binary phase-shift keying (BPSK) shows that most of the MCS levels are in the blue region, which means they have a low *BER*. Only those with higher *data rate*s are in a high *BER* region. There are a few MCS levels in the low *SNR* region that have higher *data rate*s. Higher *data rate*s are in the 15.59–17.59

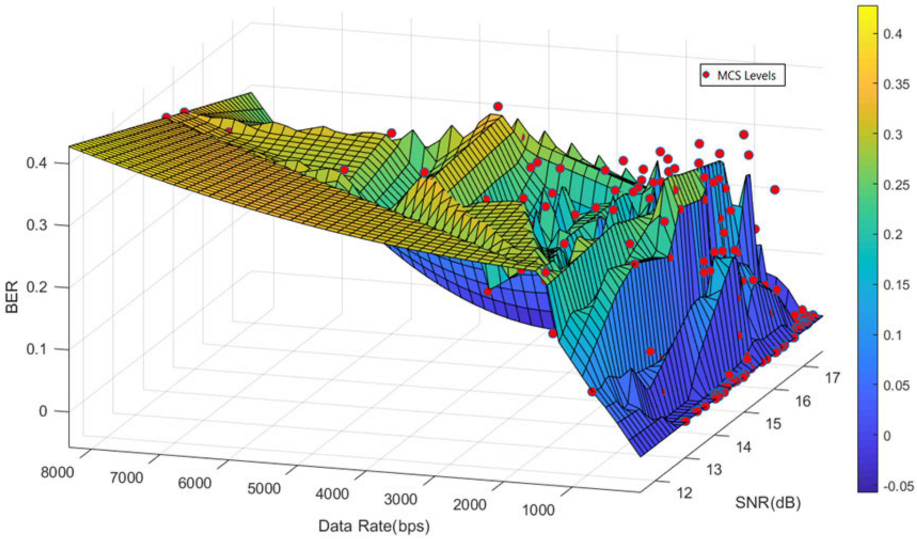


FIGURE 3. One-kilometer measured data representation in full 3-D.

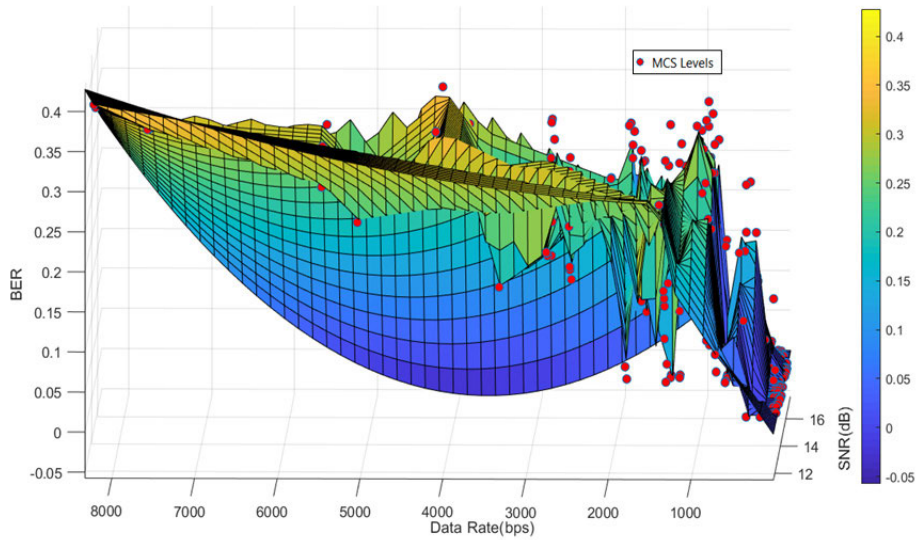


FIGURE 4. One-kilometer measured data representation in rotated 3-D.

SNR region. In quadrature phase-shift keying (QPSK), the maximum number of MCS levels are in the blue region, which means most MCS levels have a low *BER*. Also, QPSK MCS levels have more *BER* fluctuations, compared to BPSK. The higher *data rate*s are in the *SNR* region of 14.82–15.79, and most of the MCS levels that are in a higher *SNR* region, i.e., 16.34–17.23, have a low *data rate* for QPSK. Quadrature amplitude modulation with 16 possible signal combinations (16QAM) also has a lot of MCS levels in the blue region, which means a low *BER* and less than a 705 bps *data rate*. In the 705–2000 bps *data rate* range, *BER* keeps fluctuating, and then slowly goes down, and higher *data rate*s are in the 14.04–16.43 *SNR* region. Most of the MCS levels in the higher *SNR* region, i.e., 16.34–17.23, have a low *data rate*. In the case of 64QAM, we can see from the Y-axis view that the terrains start to build up early, compared to other modulations, which means most of the MCS levels have higher *BER* values. A lot of MCS levels under the *data rate* of 1000 bps have higher *BER* values, compared to those that are in a *data rate* range greater than 1000 bps. MCS levels with higher *data rate*s are in the 13.83–16.31 *SNR* region and have a higher *BER*. MCS levels in the *SNR* region greater than 16.31 have a low *data rate*.

The analysis steps for three-dimensional analysis is shown in Figure 5 by flowchart. Also, the finding steps of the threshold value for *BER*, *Data Rate* and *SNR* are shown in Table 4.

We consider *RP* to observe how it varies over different *CR*s and *PS*s. We can see that, as the *RP* goes higher, the *data rate*

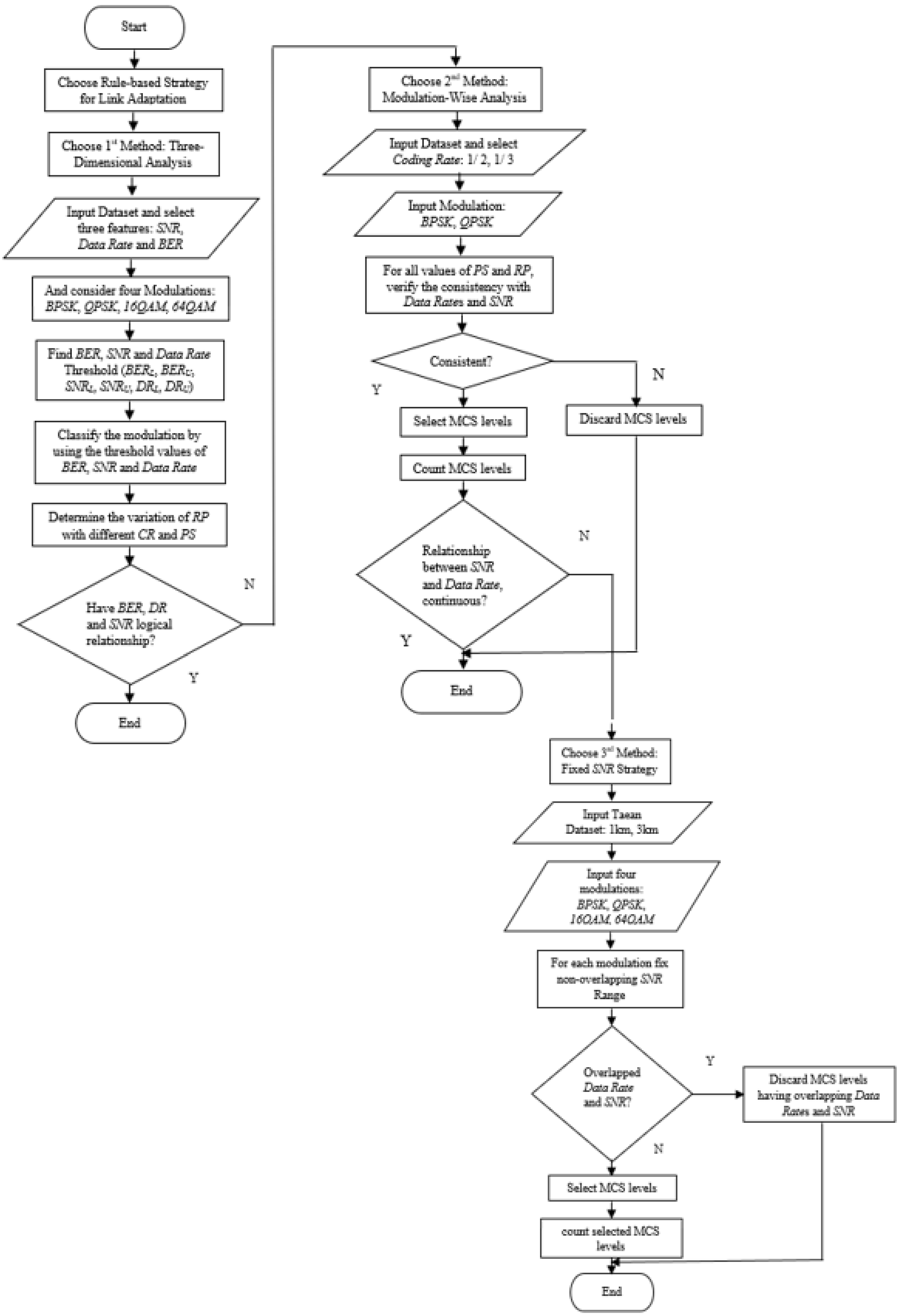


FIGURE 5. **Flowchart of the rule-based analysis.**

TABLE 4. Implementation steps of finding suitable link adaptation procedure through rule-based strategy (three-dimensional analysis).

❖ Three-Dimensional Analysis:

Initialization
Input x: *SNR*, y: *Data Rate* and z: *BER*
Input: Taean Dataset, 1 km
3-D scatterplot *Data Rate* (bps) vs. *SNR* (dB) vs. *BER*
(*BER* is selected as primary parameter because of its variation with delay spread, it's a channel quality parameter)

Step 1 *BER* Thresholds vs. *DR (Data Rate)* for MCS Pairs
- Find the *BER* Threshold according to the frequency of the MCS pairs based on 3-D scatterplot
[DR_L = lower limit of *Data Rate*, DR_U = upper limit of *Data Rate*]
count MCS pairs
for mostly occurred MCS pairs region
(blue region))
Determine DR_L
Assign BER_L
else
for fewer MCS pairs region (greenish region)
Determine the mid-level *Data Rate* range,
$DR_L < DR < DR_U$
Observe *BER* fluctuation
elseif
for fewer MCS pairs (yellow region) && decreasing *BER* fluctuation
Determine DR_U ($DR > DR_U$)
Assign BER_U
end
end
end

Step 2 *SNR* Thresholds vs. *DR* for MCS Pairs
- Find the *SNR* Threshold according to the frequency of the MCS pairs based on 3-D scatterplot
count MCS pairs
for mostly occurred MCS pairs region
(blue region)) && low *Data Rate*
Assign SNR_U (higher *SNR*)
else
for fewer MCS pairs && higher *Data Rate*,
Assign SNR_L
end
end

Step 3 Determine the Modulation Level
- Classify the Modulation level according to the relationship among *BER*, *Data Rate* and *SNR*
for different *BER*, *DR* and *SNR* region
Count the MCS levels
Find the threshold for *BER*, *Data Rate* and *SNR*
Classify the modulation
end

Step 4 *Data Rate* Threshold vs. *SNR* vs. *BER* for MCS Pairs
- Find the *Data Rate* Threshold according to the frequency of the MCS pairs

count MCS pairs
for mostly occurred MCS pairs region
(blue region)
for low *BER* && slowly increasing
&& $SNR > SNR_U$
Assign DR_L
else
for $SNR_L < SNR < SNR_U$ && fluctuating *BER*
Assign mid-level *Data Rate* Range →
$DR_L < DR < DR_U$
else
for fewer MCS levels
for decreasing fluctuation of *BER*
Assign DR_U
end
end
end
end
end

Step 5 Observe the Performance of *RP* with Different *CR* and *PS*
- Determine the variation of *RP* with different *CR* and *PS*

for *PS* = (6,3) && *CR* = 1/ 2
for *RP* = (1,1), (2,3), (8,1)
observation1 = observe *SNR*, *Data Rate* and *BER*
end
else
for *CR* = 1/ 3 and *RP* = (1,1), (2,3), (8,1)
observation2 = observe *SNR*, *Data Rate* and *BER*
end
compare (observation1, observation2)
end

goes lower. The same trend shows for *CR*s with different *PS*s. Also, *BER* goes lower as the *RP* goes higher. For *CR* (1/2), and *PS* (6,3), *RP* (8,1) has slightly higher *BER* compared to other cases. *RP* (2,3) has poor performance for *CR* (1/2) in terms of the *SNR*-to-*data-rate* ratio. *RP* (1,1) has acceptable performance for *CR* (1/3) in terms of the *SNR*-to-*data-rate*

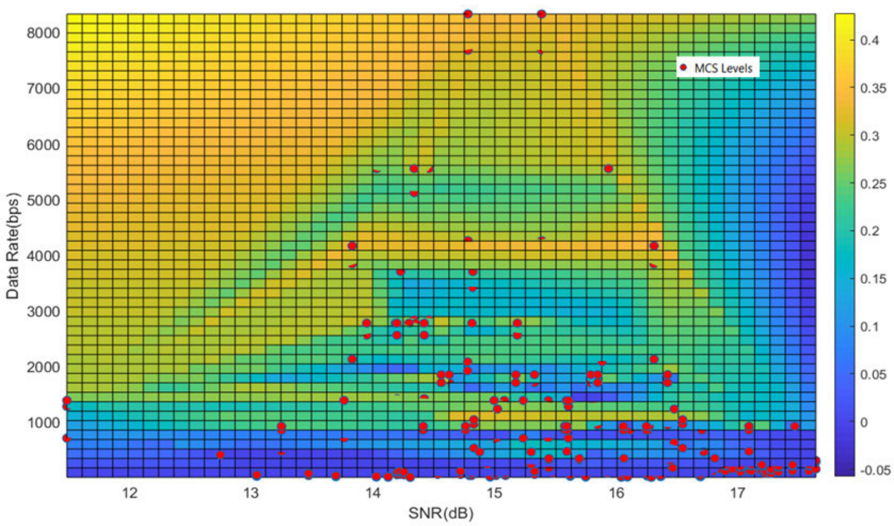


FIGURE 6. One-kilometer measured data representation from a 2-D top view.

ratio, even though the *BER* is higher. Also, *RP* (1,1) has a higher *data rate* for a low *BER* under *CR* (1/2), whereas it is the opposite for *CR* (1/3). From the observations in three-dimensional analysis, it is noticeable that the *data rate* is lower for a high *BER* at all MCS levels, which should not be the case for LA. This led us to analyze the dataset for each modulation to achieve a consistent *RP* across different *PS*s. Figure 6 shows the 2-D top view of the same dataset.

B. MODULATION-WISE ANALYSIS

We have separated the Taean data for different *CR*s for different modulations, e.g., BPSK (1/2), BSPK (1/3), QPSK (1/2), QSPK (1/3), etc. *BER*, BW_{coh}, or *RMS delay spread* have been considered to discard MCS levels for different combinations of *PS* and *RP*. We have found consistent *RP*s are (1, 1), (2, 1), (4, 1), (8, 1), (1, 3), (2, 3), (4, 3), and (1, 9). We have discarded the MCS levels that have overlapping *data rate*s and *SNR* (after combining those that have consistent *RP*s) and have got the 21 finalized MCS levels from this process. The flowchart of selection process of MCS levels is provided in Figure 5 and the implementation steps are described in Table 5.

Figure 7 shows that the *SNR* is still discontinuous, which does not follow the LA scheme. This leads to our final trial, which we call the fixed-*SNR* strategy, before the analysis using ML.

C. FIXED-SNR STRATEGY

1) TAEAN DATASET, 1 km SCENARIO

Since the *SNR* value has gone randomly in our previous strategies, we have decided to fix the *SNR* for specific modulation levels, e.g., BPSK (11,12), QPSK (13,14), 16QAM (15), 64QAM (16,17). The dataset for the 1 km sea trials contained a narrow range for *SNR*, i.e., 11–17, for four of the modulation levels. We have used our previous discarding strategies to

TABLE 5. Implementation steps of finding suitable link adaptation procedure through rule-based strategy (modulation-wise analysis).

❖ Modulation-Wise Analysis

```
Find out the MCS Levels with Consistent RPs
Input: Taean Dataset
(Separate all parameters according to modulation)
for CR = 1/2 to 1/3
    for modulation = BPSK, QPSK
        for PS = (2,1), (4,2), (6,3)
            for RP = (1, 1), (1, 3), (1, 9), (2, 1), (2, 3),
                     (2, 9), (4, 1), (4, 3), (4, 9), (8, 1),
                     (8, 3), (8, 9)
                discard MCS levels having overlapping
                Data Rates and SNR
                select MCS levels with consistent RPs
               (i.e. Data rates are proportional to SNR)
            end
        end
        count selected MCS levels
    end
    plot the SNR vs. Data Rate curve
end
```

discard MCS levels after assigning the *SNR* to each modulation level, and we have finalized five total MCS levels, which is a much less than from the previous strategies. The flowchart of discarding and selecting MCS levels is given in Figure 5 and the implementation steps of fixed-*SNR* strategy to select relevant MCS levels are described in Table 6. Even though this selection of MCS levels technically follows an LA scheme, as seen in Figure 8, the choices for selecting MCS levels depending on channel quality are fewer. Also, we cannot utilize the highest *data rate* due to the fixed-*SNR* strategy.

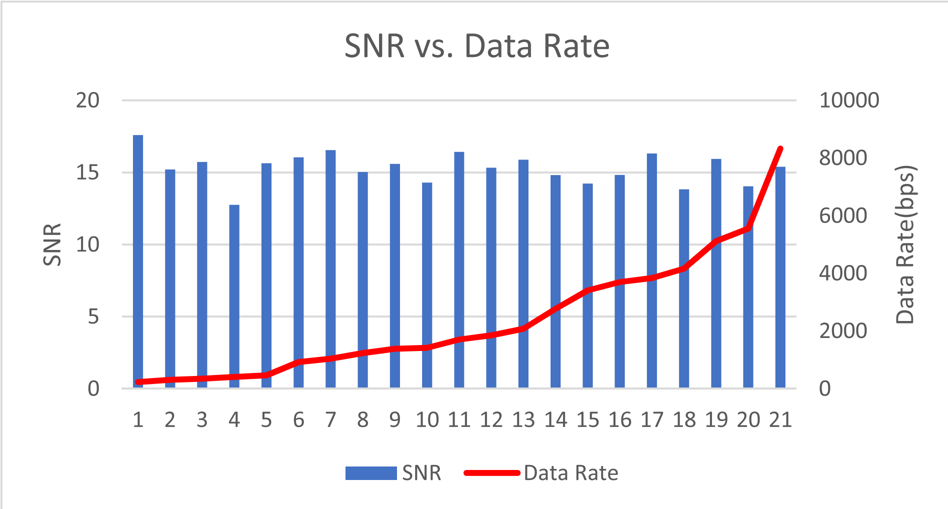


FIGURE 7. ***SNR*** **vs.** ***data rate*** **(modulation-wise analysis).**

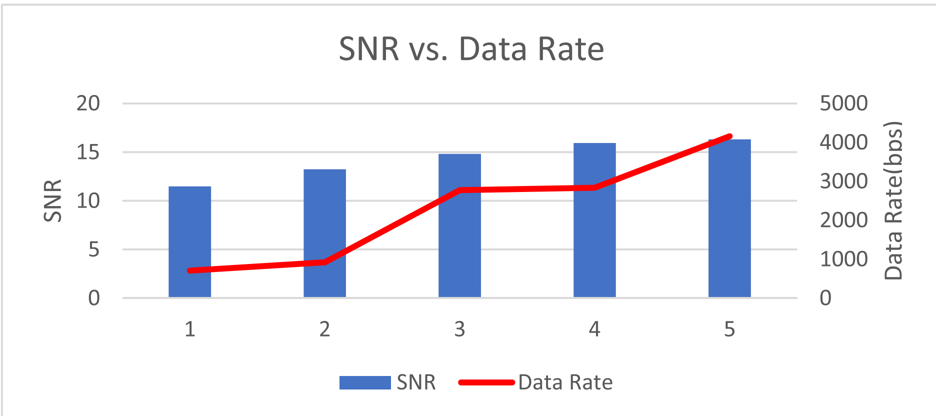


FIGURE 8. ***SNR*** **vs.** ***Data Rate*** **for the fixed-*SNR* strategy.**

TABLE 6. **Implementation steps of finding suitable link adaptation procedure through rule-based strategy (fixed-*SNR* strategy).**

❖ Fixed-*SNR* Strategy

Find out the MCS Levels with Fixed-*SNR* Range
Input: Taean Dataset
for 1 km and 3 km
 for each modulation (*BPSK, QPSK, 16QAM, 64QAM*)
  fix nonoverlapping *SNR* range
  for fixed-*SNR* range
   discard MCS levels having overlapping *Data Rates* and *SNR*
  end
  count selected MCS levels
 end
 plot the *SNR* vs. *Data Rate* curve
end

2) TAEAN DATASET, 3 km SCENARIO

The *SNR* range is 2–17, which is relatively higher than the 1 km scenario. Our desired *SNR* ranges for each modulation level are 2–5 for BPSK, 6–9 for QPSK, 10–13 for 16QAM, and 14–17 for 64QAM. But, BPSK starts from *SNR* 7, QPSK from 4, 16QAM from 2, and 64QAM from 4 in the dataset. This prevents the application of the strategy since we must skip the BPSK modulation level completely to apply the desired *SNR* assignment strategy. Also, it is impossible to combine the 3 km LA strategy with the 1 km scenario due to the difference in *SNR* ranges. Apart from that, the *data rate* goes down as the *SNR* increases in the 3 km dataset, which contradicts the LA scheme.

V. WHY ML FOR LA IN A UAC NETWORK?

If we summarize the results of the rule-based strategy, we can state that for incremental *data rate*, the *SNR* is slightly decreasing which is opposite to the LA scheme. For high *BER*, the *data rate* is lower. *BER* is overall incremental but not consistent. For each modulation level, when *SNR* is increasing, the *data rate* is decreasing. For a fixed-*SNR* strategy, even though it looks like the LA curve, still a small percentage of MCS levels following this trend of LA. In the case of 3 km, Taean Dataset, a fixed-*SNR* strategy is contradictory to the LA scheme. So, the hugely varying channel makes it more challenging to formulate a channel model for a UAC network, unlike a terrestrial network scenario, i.e., the knife-edge diffraction model, the Longley–Rice model, the Okumura model, etc. There is also a timing mismatch between the channel measurement and the application of adapted MCS levels. For the real-time application of the classification results, the measured data from the receiver must be delivered to the transmitter to be processed and classified at the adapted MCS level. But there is an approximate two-second frame delay between the transceivers, which causes an issue for applying the adapted MCS levels in real-time for LA. The solution to these problems can be a rule-based strategy at the receiver. But from measured data to get link adaptation may be possible but the reverse is not possible, what is the actual case in LTE or any other known wireless channel. Besides, our measured sea trial datasets show that the channel has a colossal number of variations over the periods, e.g., morning to noon, day to night, summer to winter, etc. So, due to day-night timing and location, the temperature, current velocity, salinity, depth of the ocean and other properties of fluid, underwater communication channel characteristics inconsistently vary over time. So, by rule-based analysis, despite our best efforts, we don't have noticeable outcomes to develop a link adaptation threshold. The relationship between *SNR* vs. *data rate* behaves peculiarly, too many inconsistencies. So, MCS selection is very difficult still many MCS levels discarding due to overlapping *data rate*s and *SNR* values. Due to this limitation, ML has a good opportunity to test and validate a huge number of measured training data and classify the modulation more accurately.

Besides, there are additional parameters, e.g., BW_{coh}, PS, MED, and BER, to be considered and measured during the sea trials, which contributes to channel quality measurement and our ML classification. So, the statistical ML algorithm–based analysis for LA prevailed over the rule-based LA approach.

VI. MACHINE LEARNING APPLICATION FOR UAC NETWORK ANALYSIS

ML algorithms can be grouped into three categories (supervised, unsupervised, and reinforcement learning), which are based on how the algorithms are trained for the dataset. In supervised learning, as the name indicates, the algorithms must be trained with a dataset same to the test data, e.g., the training data are already tagged correctly prior to the final classification. Unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here, the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training with other data. Reinforcement learning is suitable for taking action to maximize the reward in a certain scenario, and it finds the best possible behavior, or path, it should take in a specific situation. Unlike supervised learning, there is no answer, but the reinforcement agent decides what to do to perform the given task. In the absence of a training dataset, it is bound to learn from its experience.

The ML approach for our analysis starts with preparing the dataset and finding the features that contain the most variations. Those features are likely to contribute to the classification decisions. Principal component analysis (PCA) is a well-known unsupervised ML method to reduce the dimension of a test dataset and extract the important features. In our scenario, we use PCA to extract the features from the dataset, but it reduces the classification accuracy as a result. Our best guess is that, since every channel parameter contributes to the selection of MCS levels for the UAC network, unlike signal processing (in which the noise is easily separable, since the voice contains the major features), reducing the dimension of the dataset results in less accuracy in classifications. So, we have discarded the results produced by PCA. Choosing the right ML algorithm is based on trial and error, and one algorithm cannot solve all ML problems. Selecting the optimal algorithm for a certain problem is dependent on its features, such as speed, classification accuracy, training time, amount of data required to train it, implementation complications, etc.

The Incheon dataset is laid out in a format such that we could use each feature from the dataset. The ML analysis is organized by using a support vector machine (SVM) [23], k-nearest neighbor (KNN) [24], pseudo-linear discriminant analysis [25], and boosted regression tree analysis to classify the modulation levels. Every ML algorithm has its advantages and disadvantages for different types of dataset. The algorithms we decided to use usually show favorable classification accuracy with a non-linear dataset like ours. The algorithm backgrounds are briefly discussed in the following sections, although our flagship algorithm (boosted regression tree analysis) is elaborated extensively due to its higher classification accuracy.

A. SVM

Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other. An SVM model aims to separate the categories as widely as possible from the points in space by the hyperplane. New examples are then identified from the feature space based on which side of the hyperplane they fall into. In an SVM, classification can be viewed as the task of separating classes in feature space, and the data can be classified by the hyperplane.

B. KNN

This algorithm keeps all the available cases from the dataset and classifies a new case based on similarity measurements using distance functions (e.g., Euclidean distance, Hamming distance) among the samples. Each case is assigned to the class most common amongst its K nearest neighbors and is classified by a majority vote of its neighbors. If $K = 1$, then the case is simply assigned to the class of its nearest neighbor. K can be selected by $\sqrt{n}$, where n is the number of data points.

C. PSEUDO-LINEAR DISCRIMINANT ANALYSIS

Linear discriminant analysis (LDA) focuses on maximizing separability among known categories. LDA takes information from the features to create a new axis by following two criteria simultaneously:

- Maximize the distance between the means (μ) of the categories.
- Minimize the variation within each category, and the variation is called scatter (s^2).

The algorithm can be mathematically expressed as follows:

$$\frac{(\mu_1-\mu_2)^2}{s_1^2+s_2^2} \tag{1}$$

In LDA, all classes have the same covariance matrix, but it will be inverted using the pseudo inverse.

D. BOOSTED REGRESSION TREE ANALYSIS

Regression trees bisect the input parameters to create binary partitions, which are called nodes, and terminal nodes are called leaf nodes. Within each node, the regression tree estimates the modulation as the average of the modulations corresponding to the transmissions. This process minimizes the mean squared error (MSE) between predicted and measured value. To optimally bisect the input parameter space in each iteration, the training algorithm selects the input parameter and its associated value to maximally reduce the overall MSE for a given training dataset. This node-splitting procedure is repeated recursively until the desired MSE is achieved for the tree, or until a desired maximum tree depth is reached. Regression trees can be used with heterogeneous datasets, e.g., numerical, categorical, and ordinal inputs. It can also handle missing inputs transparently. These two properties make trees one of the most versatile statistical learning methods currently in use. Yet, like other statistical learning techniques, regression trees can produce low-bias estimates, but the estimates may be susceptible to high variance.

To resolve the high-variance problem, in statistical learning, researchers have used ensembles of trees and achieved more robust estimates [26]–[29], compared to single trees. In boosted trees [30], the general learning technique AdaBoost [31] is applied to regression trees. Many trees are trained on the entire training data iteratively in such a way that in each iteration the training samples and the predictions are assigned weights adaptively, depending on the accuracy of the predicted values. The aggregated, final prediction from boosted trees is a combination of weighted predictions from each tree, and the result is that the overall boosted tree classifier produces successively more accurate predictions as a function of the number of trees:

$$\boldsymbol{modulations} = \sum\nolimits_{i=1}^{N} w_i T_i \tag{2}$$

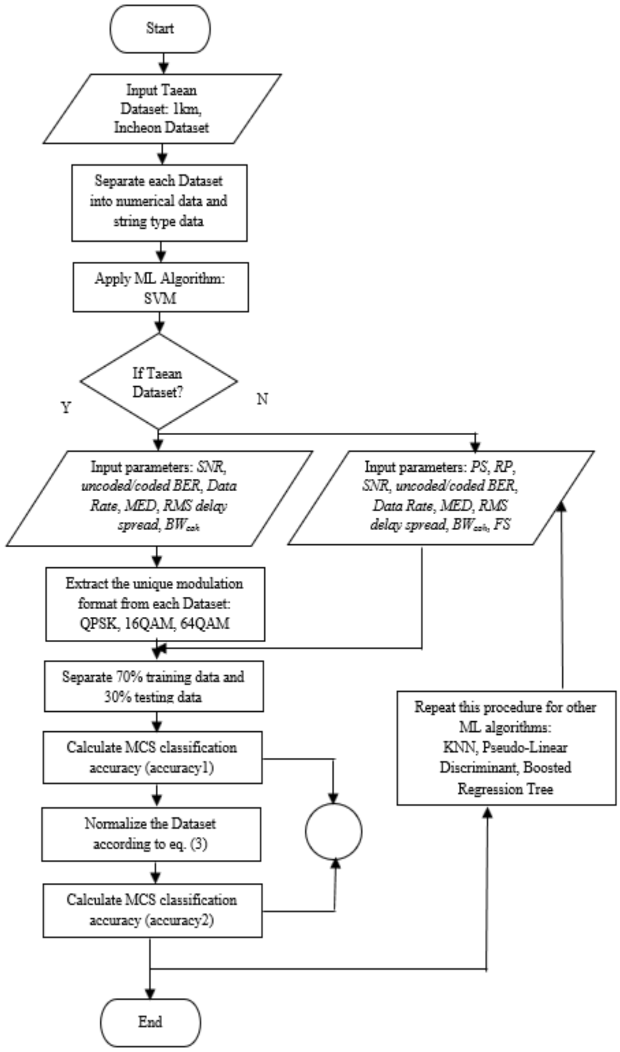


FIGURE 9. Flowchart for Machine Learning Algorithms.

where i = 1, 2,.......N is the number of trees, and w_i is the AdaBoost weight of T_i, which is the ith tree in the boosted ensemble.

VII. CLASSIFICATION OF MCS LEVELS USING ML ALGORITHMS

We split the datasets into two groups, e.g., 70% of the data for training the model, and 30% for testing the classification accuracy, for all the algorithms used. We use 2000 ensemble constituent trees for boosted regression tree analysis, in which the minimum leaf size of each tree is 20, and the minimum parent size is 20. The simulation run is once for each algorithm. For SVM we test both Incheon and Taean dataset (1 km) and for the other three algorithms, we only test the Incheon dataset, because classification performance with Incheon dataset shows better than the Taean dataset. For all four ML algorithms, SVM, KNN, Pseudo-Linear Discriminant and Boosted Regression Tree analysis with both Taean and Incheon Datasets, the flowchart shows the working procedure in Figure 9. Also finding out the

TABLE 7. Implementation steps of finding suitable link adaptation procedure through machine learning.

❖ Machine Learning Algorithms: SVM, KNN, Pseudo-Linear Discriminant, Boosted Regression Tree

Step 1 Find out the Classification Accuracy for SVM
Input: Taean 1 km Dataset, Incheon Dataset
Separate the Dataset into numerical data (features) and string type data (modulation)
for ML algorithm: SVM
Input parameters:
if
Taean Dataset: *SNR, uncoded/coded BER, data rate, MED, RMS delay spread,* BW_{coh}
else
Incheon Dataset: *PS, RP, data rate, MED, RMS delay spread,* BW_{coh}*, FS, SNR, uncoded/code BER*
extract the unique modulation format (*QPSK, 16QAM, 64QAM*)
separate 70% training data and 30% testing data
calculate MCS classification accuracy (accuracy1)
normalize the dataset according to eq. (3)
calculate MCS classification accuracy (accuracy2)
compare (accuracy1, accuracy2)
plot confusion matrix (for showing individual modulation classification accuracy)
measure execution time
end
Step 2 Find out the Classification Accuracy for Other ML Algorithms
for ML algorithm: KNN, Pseudo-Linear Discriminant, Boosted Regression Tree
Input: Incheon Dataset
do same as before
end

classification accuracy, the implementation steps are described in Table 7. Besides, for comparison purposes, we measure execution run time which is enlisted in Table 8. Also, we generate a confusion matrix of the target class and predicted class from test data to show the classification accuracy of each class separately in case of boosted regression tree analysis.

A. CLASSIFICATION USING AN SVM FOR THE TAEAN DATASET

The first analysis is using the following parameters: *SNR*, uncoded/coded *BER*, *data rate*, *MED*, *RMS delay spread*, and BW_{coh}. The classification accuracy is 46.4%. Next, we normalize the Taean dataset so that each variable contributes equally to the analysis. We use the default SVM template in MATLAB. The kernel function is empty consideration. If there are large differences between the ranges of initial variables, those variables with larger ranges will dominate

TABLE 8. Classification accuracy of different algorithms.

Algorithm	Dataset	Max. Classification Accuracy	Elapsed Time
SVM	Taean	46.4%	0.18s
SVM	Incheon	83.31%	5.7s
KNN	Incheon	75.75%	0.88s
Pseudo-linear Discriminant Analysis	Incheon	81.99%	0.31s
Boosted Regression Tree Analysis	Incheon	99.97%	6.92s

those with small ranges. The formula to normalize the dataset is as follows:

$$z = \frac{value - mean}{standad\ deviation} \tag{3}$$

After normalization, the accuracy is found as 41.6%, but it is still less than from using all the parameters.

B. CLASSIFICATION USING AN SVM FOR THE INCHEON DATASET

First analysis of the Incheon dataset is using all the features available in the dataset, i.e., *PS*, *RP*, *data rate*, *MED*, *RMS delay spread*, BW_{coh}, *frequency shift* (*FS*), *SNR*, and *uncoded/ coded BER*, and the classification accuracy is 82.448% which is quite a bit higher than what we achieved with the Taean dataset. The reason is that better channel parameters contributed to better results. The next analysis is after decoding the data, which means the dataset contained all the features of turbo coding data. The classification accuracy slightly increased to 83.3111%. We also normalize the Incheon dataset using equation (3), and find the accuracy is increased to 69.8779%, which is still less than from using all the parameters. The trend in classification accuracy continued to be the same as the Taean dataset.

C. CLASSIFICATION USING KNN

The classification accuracy results from the previous analysis show that the Incheon channel is better, compared to the Taean channel, thus helping classifiers to have better accuracy. Also, adopting all the features from the dataset for ML classification produced a better result, compared to using a normalized dataset. So, we decide to use the Incheon dataset with all the features for further ML analysis from here on. Using all the features, as conducted in SVM classification, we obtain 75.75% accuracy with KNN classification.

D. CLASSIFICATION USING PSEUDO-LINEAR DISCRIMINANT ANALYSIS

As discussed in the KNN classification analysis, the Incheon dataset with all the parameters is also used for pseudo-linear discriminant analysis classification, and the accuracy is 81.99%.

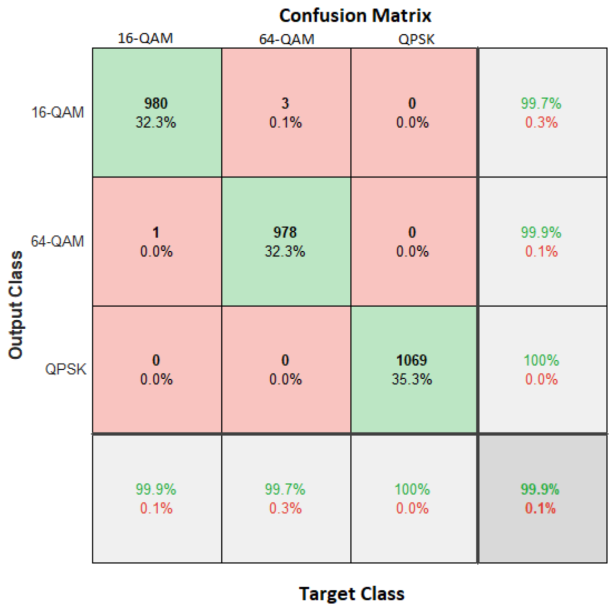


FIGURE 10. **Target classes and Predicted classes using all the parameters for the Incheon dataset.**

E. CLASSIFICATION USING BOOSTED REGRESSION TREE ANAYSIS

For the previously mentioned Incheon dataset, we achieve the highest (and nearly perfect) classification accuracy from among all the ML algorithms by using boosted tree regression analysis, which was 99.967%. We use boosted trees to learn the relationship between the measured *BER* with the signal parameters, the received *SNR*, and other related channel metrics which characterize the signal distortion [12]. To alleviate the high variance of data, boosted tree option is a very good option. Figure 10 shows that the algorithm successfully predicted all the modulation labels/ classes separately as same as the corresponding target labels/ classes for all test data approximately. The boosted regression tree algorithm can be used for heterogeneous datasets, i.e., numerical, categorical, or ordinal inputs. It can also transparently handle missing inputs. The Incheon dataset is well organized for the boosted regression tree algorithm to classify the modulation level, thus producing the near-perfect result. The only tradeoff is having numerous branch trees, which divided the dataset into hundreds of groups, and caused overhead, compared to other algorithms. The classification accuracy results of each algorithm are rearranged in Table 8.

There are a few complications in applying machine learning, such as the reciprocity issue in the media access control (MAC) layer, and frame delay between transceivers. Our UAC network is a frequency division duplexing (FDD) system, and thus, it uses separate channels for uplink and downlink transmissions, which means the frequency is different between transmitter and receiver. The Incheon dataset contains only uplink information on different parameters; thus, the classification result is only applicable to the uplink channel. So, the system being FDD creates a complication in applying the classifying result to the transmitter end.

VIII. CONCLUSION

In this paper, we extensively analyzed the UAC network measured data using three-dimensional analysis, modulation-wise analysis, a fixed-*SNR* strategy, and machine learning analysis. The obtained results aim towards developing a link-adapted system that will be able to boost spectral efficiency in response to any channel conditions while retaining the desired level of covertness and reliability. We have investigated a few reliable machine learning algorithms for classification of MCS levels and boosted tree regression produced the best result, boasting 99.97% accuracy in classification. Like 3GPP LTE, the *SNR* threshold for the UAC network must be set on the receiver side to differentiate each MCS level. We have several parameters in our case, i.e., *SNR*, *DR*, *BER*, BW_{coh}, etc. The *BER* plot can be used to see how the channel affects those parameters over a certain period. This could help us to calibrate the *SNR* vs. Throughput by assigning more adequate MCS levels for the improvement of link adaptation in the UAC network.

REFERENCES

[1] F. Qu, Z. Wang, L. Yang, and Z. Wu, "A journey toward modeling and resolving Doppler in underwater acoustic communications," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 49–55, Feb. 2016, doi: 10.1109/MCOM.2016.7402260.

[2] Z. Liu, L. Gao, Y. Liu, X. Guan, K. Ma, and Y. Wang, "Efficient QoS support for robust resource allocation in blockchain-based femtocell networks," *IEEE Trans. Ind. Informat.*, early access, Sep. 3, 2019, doi: 10.1109/TII.2019.2939146.

[3] T. Melodia, H. Kulhandjian, and E. Demirors, "Advances in underwater acoustic networking," in *Mobile Ad Hoc Networking: Cutting Edge Directions*, 2nd ed. Hoboken, NJ, USA: Wiley, 2013, pp. 504–854. [Online]. Available: https://doi.org/10.1002/9781118511305.ch23.

[4] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 84–89, Jan. 2009, doi: 10.1109/MCOM.2009.4752682.

[5] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708–1721, Mar. 2010.

[6] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Energy efficiency and spectrum efficiency of multihop device-to-device communications underlaying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 367–380, Jan. 2016, doi: 10.1109/TVT.2015.2389823.

[7] S.-U. Kim, H.-S. Cheon, S.-B. Seo, S.-M. Song, and S.-Y. Park, "A hexagon tessellation approach for the transmission energy efficiency in underwater wireless sensor networks," *J. Inf. Process. Syst.*, vol. 6, no. 1, pp. 53–66, Mar. 2010, doi: 10.3745/JIPS.2010.6.1.053.

[8] M. Stojanovic, "On the design of underwater acoustic cellular systems," in *Proc. OCEANS*, Aberdeen, U.K., Jun. 2007, pp. 1–6, doi: 10.1109/OCEANSE.2007.4302226.

[9] B. Sirinivasan, "Capacity of underwater acoustic OFDM cellular networks," M.S. Thesis, Univ. California, Santa Barbara, Santa Barbara, CA, USA, 2008, doi: 10.1109/OCEANSSYD.2010.5603911.

[10] I. Ahmad and K. Chang, "Effective SNR mapping and link adaptation strategy for next-generation underwater acoustic communications networks: A cross-layer approach," *IEEE Access*, vol. 7, pp. 44150–44164, 2019, doi: 10.1109/ACCESS.2019.2908018.

[11] K. Pelekanakis and L. Cazzanti, "On adaptive modulation for low SNR underwater acoustic communications," in *Proc. OCEANS*, Charleston, SC USA, Oct. 2018, pp. 1–6.

[12] M. Sadeghi, M. Elamassie, and M. Uysal, "Adaptive OFDM-based acoustic underwater transmission: System design and experimental verification," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, İstanbul, Turkey, Jun. 2017, pp. 1–5, doi: 10.1109/BlackSeaCom.2017.8277660.

[13] J. Wang, C. Yuen, Y. L. Guan, and F. Ge, "Reinforcement learning-based link adaptation in long delayed underwater acoustic channel," in *Proc. MATEC Web Conf.*, vol. 283, 2019, p. 07001, doi: 10.1051/matecconf/201928307001.

[14] S. Shankar and M. Chitre, "Tuning an underwater communication link," in *Proc. MTS/IEEE OCEANS*, Bergen, Norway, Jun. 2013, pp. 1–9, doi: 10.1109/OCEANS-Bergen.2013.6607956.

[15] Y. Wang, H. Zhang, Z. Sang, L. Xu, C. Cao, and T. A. Gulliver, "Modulation classification of underwater communication with deep learning network," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–12, Apr. 2019, doi: 10.1155/2019/8039632.

[16] *Onedrive Link for Dataset*. [Online]. Available: https://1drv.ms/u/s!AswxJeiLN4eNjxEgvdcXN06Pna4S?e=Nc7b6c and https://ieee-dataport.org/documents/taean-and-incheon-measured-data, doi: 10.21227/4x41-7146.

[17] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1996, p. 736.

[18] I. Ahmad, Z. Kaleem, and K. Chang, "Uplink power control for interference mitigation based on users priority in two-tier femtocell network," in *Proc. Int. Conf. ICT Converg. (ICTC)*, Jeju, South Korea, Oct. 2013, pp. 474–475, doi: 10.1109/ICTC.2013.6675399.

[19] I. Ahmad, W. Chen, and K. H. Chang, "Co-channel interference analysis using cooperative communication schemes for the coexistence of PS-LTE and LTE-R networks," in *Proc. IEEE Commun. Electron. Special Session LTE Technol. Services*, Hạ Long Bay, Vietnam, Jul. 2016, pp. 181–182.

[20] I. Ahmad, W. Chen, and K. Chang, "LTE-railway user priority-based cooperative resource allocation schemes for coexisting public safety and railway networks," *IEEE Access*, vol. 5, pp. 7985–8000, May 2017, doi: 10.1109/ACCESS.2017.2698098.

[21] Z. Kaleem, Y. Li, and K. Chang, "Public safety users' priority-based energy and time-efficient device discovery scheme with contention resolution for ProSe in third generation partnership project long-term evolution-advanced systems," *IET Commun.*, vol. 10, no. 15, pp. 1873–1883, Oct. 2016, doi: 10.1049/iet-com.2016.0029.

[22] M. Stojanovic, "Design and capacity analysis of cellular-type underwater acoustic networks," *IEEE J. Ocean. Eng.*, vol. 33, no. 2, pp. 171–181, Apr. 2008, doi: 10.1109/JOE.2008.920210.

[23] Z. Wu, X. Wang, Z. Gao, and G. Ren, "Automatic digital modulation recognition based on support vector machines," in *Proc. Int. Conf. Neural Netw. Brain*, Beijing, China, Oct. 2005, pp. 1025–1028, doi: 10.1109/ICNNB.2005.1614792.

[24] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statist.*, vol. 46, no. 3, pp. 175–185, Aug. 1992, doi: 10.2307/2685209.

[25] J. Liu, S. Chen, X. Tan, and D. Zhang, "Efficient pseudoinverse linear discriminant analysis and its nonlinear form for face recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 21, no. 8, pp. 1265–1278, Dec. 2007, doi: 10.1142/S0218001407005946.

[26] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2001.

[27] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[28] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[29] L. Breiman, "Arcing classifiers," *Ann. Statist.*, vol. 26, no. 3, pp. 801–849, 1998.

[30] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.

[31] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.*, no. 5, pp. 197–227, Jun. 1990, doi: 10.1023/A:1022648800760.

ALAMGIR M S M received the B.S. degree in Electronics Engineering from Kongju National University, South Korea, in 2016, and the M.S. degree from the Mobile Telecommunication Research Laboratory (MTRL), Department of Electronics Engineering, Inha University, South Korea.

He is currently a Chief Research Engineer with Entec Electric and Electronic Company Ltd. His research interests include mobile ad-hoc networks for UAV and anti-drone technology using AI, and link adaptation on underwater communication using machine learning.

Mr. Alamgir was a recipient of the Korean Government Scholarship Program (KGSP) and the Jungseok International Scholarship to pursue his B.S. and M.S. degrees, due to his excellent academic career.

MST. NAJNIN SULTANA (Member, IEEE) was born in Noakhali, Bangladesh, in 1978. She received the B.S. and M.S. degrees in Applied Physics and Electronics from the University of Dhaka, Bangladesh, in 2003 and 2005, respectively. She is currently pursuing the Ph.D. degree with the Mobile Telecommunication Research Laboratory (MTRL), Electronic Engineering Department, Inha University, South Korea.

From 2008 to 2012, she was a Research Assistant with the UWB-ITRC Laboratory, Inha University. She has been serving as an Assistant Professor (study leave) with the Computer Science and Engineering Department, Uttara University, Bangladesh, since 2016. Since 2014, she has been an Assistant Professor with the Electronics and Telecommunication Department, Daffodil International University, Bangladesh. She is the author of one book, a coauthor of another book chapter, and the author of eight articles. Her research interests include deep learning for LTE, 5G link adaptation, MIMO beamforming, and massive MIMO.

Ms. Sultana is an Associate Member of SEEK. Her honors include the First Grade Comilla Board Scholarship (Bangladesh) to carry out the B.S. degree, the Scholarship from the University of Dhaka (Bangladesh) for pursuing her M.S. degree, and the Jungseok Memorial International Scholarship (South Korea) to pursue the Ph.D. degree, due to her excellent academic career.

KYUNGHI CHANG (Senior Member, IEEE) received the B.S. and M.S. degrees in Electronics Engineering from Yonsei University, Seoul, South Korea, in 1985 and 1987, respectively, and the Ph.D. degree in Electrical Engineering from Texas A&M University, College Station, TX, USA, in 1992.

From 1989 to 1990, he was with the Samsung Advanced Institute of Technology (SAIT) as a Member of the Research Staff and was involved in digital signal processing system design. From 1992 to 2003, he was with the Electronics and Telecommunications Research Institute (ETRI) as a Principal Member of the Technical Staff, where he led the design teams involved in the WCDMA UE modem and 4G radio transmission technology (RTT). He is currently with the Electronic Engineering Department, Inha University. His research interests include radio transmission technology in 3GPP LTE-A and 5G systems, public safety and mobile ad-hoc networks (especially for UAV), cellular-V2X technology, maritime and underwater communications, and applications of AI technologies.

Prof. Chang was a recipient of the LG Academic Awards, in 2006, the Haedong Best Paper Awards, in 2007, the IEEE ComSoc Best Paper Awards, in 2008, the Haedong Academic Awards, in 2010, and the SKT SafeNet Best Idea Awards, in 2015. He is currently the Chairman of the Expert Committee in SafeNet Forum, and Mobile and Automotive Convergence Committee in 5G Forum. He has served as the Editor-in-Chief and the Executive Director for the *Journal of Korean Institute of Communications and Information Sciences* (KICS), from 2010 to 2012 and in 2013, respectively, and the Vice President at the KICS, from 2017 to 2018. He has also served as an Editor for ITU-R TG8/1 IMT.MOD.