

Received March 14, 2022, accepted April 29, 2022, date of publication May 10, 2022, date of current version May 31, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3174126

Predictability of Internet of Things Traffic at the Medium Access Control Layer Against Information-Theoretic Bounds

MERT NAKIP¹, (Student Member, IEEE), BARAN CAN GÜL²,
VOLKAN RODOPLU³, (Member, IEEE), AND CÜNEYT GÜZELİŞ³

¹Institute of Theoretical and Applied Informatics, Polish Academy of Sciences (PAN), 44100 Gliwice, Poland

²Department of Electrical Engineering, University of Stuttgart, 70550 Stuttgart, Germany

³Department of Electrical and Electronics Engineering, Yaşar University, 35100 İzmir, Turkey

Corresponding author: Volkan Rodoplu (volkan.rodoplu@yasar.edu.tr)

This work was funded by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement No. 846077, entitled "Quality of Service for the Internet of Things in Smart Cities via Predictive Networks".

ABSTRACT Most of the existing Medium Access Control (MAC) layer protocols for the Internet of Things (IoT) model the traffic generated by each IoT device via random arrivals such as those in a Poisson process. Under this model, since it is implied that IoT device traffic cannot be predicted, only reactive MAC-layer protocols in which the network responds to the current traffic are viable. In contrast, recent work has demonstrated that the traffic generated by an individual IoT device can be predictable, thus enabling predictive network protocols at the MAC layer. In this paper, we investigate information-theoretic bounds on the predictability of IoT traffic of individual devices. To this end, first, we compare the performance achieved by the following state-of-the-art forecasters on individual IoT device traffic: Logistic Regression, Multi-Layer Perceptron (MLP), 1-Dimensional Convolutional Neural Network (1D CNN), and Long Short Term Memory (LSTM) as well as MLP under feature selection based on Analysis of Variance (ANOVA) and Auto-Correlation Function (ACF). Second, we quantify the gap between the performance of these forecasters against information-theoretic bounds as follows: For IoT devices that generate a fixed number of bits at each generation instance, we measure the gap between the forecasting accuracy and the information-theoretic bound established by Fano's inequality on the probability of correct prediction. Our empirical results show that existing forecasting schemes perform close to the information-theoretic bound in this case. For IoT devices that generate a variable number of bits, we measure the gap between the Mean Square Error (MSE) and the estimation-theoretic counterpart to Fano's inequality. Our empirical results show that the performance of existing forecasting schemes is far from the information-theoretic bound in this case. This work motivates the machine learning community to develop forecasting schemes that approach information-theoretic bounds. Furthermore, this work is expected to impact the development of predictive MAC-layer protocols that exploit these bounds.

INDEX TERMS Internet of Things (IoT), network traffic, predictability, forecasting, machine learning.

I. INTRODUCTION

The application of Artificial Intelligence (AI) techniques to the Internet of Things (IoT) is increasingly playing a significant role in the design of next-generation wireless networks [1]. In particular, the application of AI techniques to the problem of predicting the traffic from IoT devices holds much promise in the new paradigm of "predictive networks",

The associate editor coordinating the review of this manuscript and approving it for publication was Wei-Wen Hu¹.

in which a network predicts the future input to the network, such as the traffic generation patterns of IoT devices, and allocates the network resources in advance. This contrasts sharply with the existing paradigm of network protocols, in which the network merely reacts to the current traffic demand on the network [2]. Significant advantages accrue when predictive network protocols are used in the place of reactive ones: The amount of control overhead that occurs due to handshaking and contention in reactive protocols is completely avoided in predictive network protocols, and a much higher network

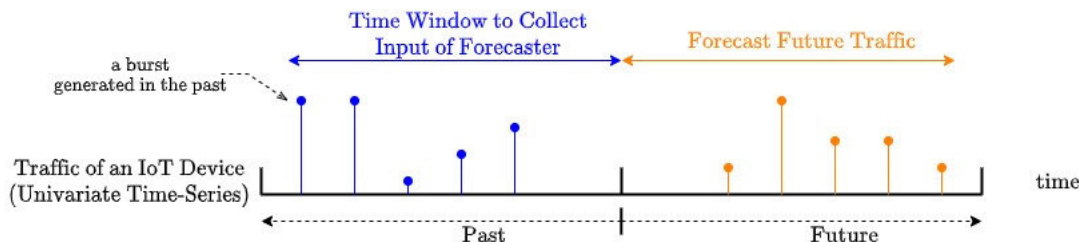


FIGURE 1. Visualization of the inputs and the outputs of a forecaster for an IoT device.

performance can potentially be achieved if accurate forecasts of the future inputs to the network can be formed.

The main motivation of this article is to demonstrate the predictability of the traffic from individual IoT devices and thus establish a firm foundation for the development of predictive network protocols in the near future. The demonstration of such predictability has the following two key benefits: (1) Predictability of individual IoT device traffic will enable the development of predictive MAC-layer protocols for IoT in next-generation networks. (2) The quantification of the gap between the performance of existing forecasting schemes and the information-theoretic bounds has the potential to motivate the machine learning community to develop novel forecasting schemes that approach the information-theoretic bounds.

In order to clarify the concept of forecasting future IoT traffic that is generated by an individual IoT device, we visualize the traffic forecasting as in Fig. 1. In this figure, the horizontal axis is the time axis where the left is the past, the right is the future, and the middle line is the current time (present). In addition, each stalagmite represents a burst in the traffic pattern of the IoT device under consideration; these bursts are either generated in the past (blue stalagmites) or predicted to be generated in the future (orange stalagmites). A forecaster is used to predict the future traffic generation based on the traffic that falls in a certain time window in the past.

This work stands in stark contrast with the current model of “random arrivals” (e.g. as in a Poisson process) that is typically used to model the traffic from IoT devices at the Medium Access Control (MAC) layer in most of the existing network protocols [3]–[8]. While the random arrivals model may be suitable for human-generated traffic, it fails to capture the predictability inherent in machine-generated traffic in Machine-to-Machine (M2M) communication. In this article, we show that the traffic from a variety of IoT devices in distinct traffic classes possesses inherent predictability.

In order to demonstrate this predictability, first, we present the selection of the optimal length of the time window that is considered as the input of a forecaster for the univariate time-series traffic data. To this end, we analyze the entropy of IoT traffic and compare it with that of a memoryless source as a function of increasing window length.

Second, we establish an information-theoretic framework by which we compare the empirical predictability that is achieved in forecasting the traffic of an IoT device against the information-theoretic bounds. For IoT devices that generate a

fixed number of bits at each generation instance, we compute the information-theoretic upper bound established by Fano’s inequality on the probability of forecasting error. For IoT devices that generate a variable number of bits, we compute a lower bound of the Mean Squared Error (MSE) via the estimation-theoretic counterpart to Fano’s inequality. Furthermore, we define two performance metrics: 1) Normalized Accuracy, which measures the gap between the accuracy that is realized by using a forecaster and the information-theoretic upper bound on accuracy; and 2) Normalized Reciprocal MSE, which measures the gap between the MSE realized using a forecaster and the information-theoretic lower bound on MSE.

Third, we compare the performance achieved by the following state-of-the-art forecasters on the traffic of individual IoT devices with respect to Normalized Accuracy and Normalized Reciprocal MSE metrics: Logistic Regression, Multi-Layer Perceptron (MLP), 1-Dimensional Convolutional Neural Network (1D CNN), Long-Short Term Memory (LSTM) as well as MLP combined with Analysis of Variance-based feature selection (ANOVA-MLP) and Auto-Correlation Function-based feature selection (ACF-MLP).

In this work, we demonstrate not only that the traffic of individual IoT devices is predictable but also give indications as to which forecasters achieve the best performance with regard to distinct IoT traffic classes. Our results also indicate the differences that may exist in the empirical predictability that is achieved across distinct IoT devices. Taking account of such differences will be important in designing next-generation predictive protocols based on the degree of predictability that can be achieved for each IoT device on the network.

In summary, the main contributions of this work are as follows:

- We investigate the predictability of the traffic of individual IoT devices by selecting the optimal length of the time window for the forecaster input.
- We perform a comparative analysis of the performance achieved by state-of-the-art forecasters on individual IoT device traffic.
- We quantify the gap between the performance of these forecasters against information-theoretic bounds.

The rest of this paper is organized as follows: In Section II, we contrast our work with the articles in the existing

TABLE 1. Summary of the recent related works on predictive network techniques for M2M/IoT in Section II-A.

Paper	Policy Name	Techniques Developed/Used	Forecasting Required
[10]	Predictive Pre-Allocation	Prediction of Device Activity	✓
[11], [12]	FUG	Binary Markovian process [11]	✓
		SVM & LSTM [12]	
[13]–[17]	JFS	Heuristic [13]	✓
		MSA [14]	
		MC Optimization [15]	
		RGT [16]	
		E-ASEF [17]	
[18]	DAFS	Throughput Estimation	✓
[19], [20]	QDTP	Queueing Analysis	✗
		Diffusion Approximations	
		Queueing Analysis	

literature. In Section III, we present high-level statistics of the considered IoT traffic data and the forecasting of IoT traffic including the hyper-parameter tuning of forecasting techniques. In Section IV, we analyze the predictability of IoT traffic that is generated by individual devices, and we present the comparison of the performance of forecasting techniques against the information-theoretic bounds. In Section V, we present our conclusions.

II. RELATIONSHIP TO THE STATE OF THE ART

We now present the relationship between our work and the state-of-the-art works in three categories: 1) the works that develop techniques for predictive networks based on the predictability of IoT traffic, 2) the works that present algorithms to forecast IoT traffic, and 3) the works that use Fano’s inequality and its counterpart for predictability analysis targeting wireless communication systems. Tables 1, 2, and 3 present the summary of the recent related works for these categories respectively.¹

A. PREDICTIVE NETWORK TECHNIQUES

We present the works that develop proactive solutions to the massive access problem. Early works have developed proactive solutions targeting Human-to-Human (H2H) traffic in [21]–[23] and Machine-to-Human (M2H) traffic in [24]–[26]. For cognitive radio networks, a proactive MAC protocol was developed by Reference [27] in order to prevent the network against primary user emulation and data falsification attacks.

For M2M communication, Reference [28] has developed a resource allocation algorithm that gives access grant proac-

tively to the neighbors of activated devices. Reference [29] aims to minimize total service time for the transmission of IoT traffic packets by allocating resources proactively. Moreover, based on the prediction of the channel condition, the video quality and resource sharing have been determined in [30] while an energy-efficient stochastic predictive resource allocation technique has been developed in [31]. Reference [32] has developed a technique to drop a subset of packets based on their predicted impacts on the latency and the performance. Reference [10] has scheduled uplink access of Industrial IoT devices for delay-critical applications based on the prediction of activity of these devices. In addition, Reference [33] has recently developed a predictive network architecture based on the prediction of total generated traffic in the network.

A recent proactive solution technique, called Fast Uplink Grant (FUG) has been proposed by 3GPP in Release 14 to provide an uplink grant to traffic packets by scheduling those based on the predictions [34]–[36]. In addition, Reference [11] developed a method for FUG which uses binary Markovian process to model the packet transmissions of IoT devices, and Reference [12] predicts IoT traffic via each of Support Vector Machine (SVM) and LSTM techniques aiming to enable FUG.

Another recent proactive solution technique, called Joint Forecasting-Scheduling (JFS), is proposed to schedule the transmission of the traffic of individual devices based on forecasting in [13]. Reference [14] developed the Multi-Scale Algorithm (MSA) in which the forecasting of IoT traffic is performed on multiple time scales in order to enhance the scalability of JFS for a massive number of devices. Reference [15] extended JFS to the Multi-Channel (MC)

¹We took Reference [9] as an example for the styles of these tables.

case. In addition, Reference [16] developed the Randomization of Generation Times (RGT) preprocessing algorithm, which improves the performance of IoT traffic scheduling by using queueing theory techniques.

Furthermore, Reference [17] developed a technique, called Emulation of Application Specific Error Function (E-ASEF), to perform an analysis of the relationship between the forecasting error and the network performance under predictive network protocols. Reference [18] designed a meta-MAC framework, called Dynamic Automatic Forecaster Selection (DAFS), to select the best performing forecasting scheme with respect to the estimated network performance. For scalability improvement in IoT networks, diffusion analysis on the proactive scheduling of packet transmission is presented in Reference [19] and Quasi-Deterministic Transmission Policy (QDTP) is developed in References [19], [20]. The results of the works in this category show that the predictive network protocols (i.e. predictive scheduling of packet transmissions of IoT devices) are highly promising for the networks with a massive number of IoT devices. However, they also show that the performance of predictive protocols is very sensitive to the predictability of IoT traffic. In this paper, we analyze the predictability of the individual IoT device traffic, and we present a comparison of the performance of ML techniques against information-theoretic bounds.

B. IoT TRAFFIC FORECASTING

We compare our work to the previous studies on forecasting the traffic generation patterns of IoT devices. Reference [37] presented a comparative study of MLP, LSTM, 1D CNN, and ARIMA models along with the state-of-the-art feature selection algorithms, i.e. Auto-Correlation Function (ACF), embedding dimensions, and Analysis of Variance (ANOVA) to yield the best performing forecasters for each IoT device. Reference [38] developed an original neural network architecture, called Feature Selection Forecasting (FSF), specifically to predict traffic generation patterns of individual IoT devices. Reference [39] presented a deep learning approach as Gradient Boosting (GB) model with residual networks and stacked models to predict network traffic obtained by a mobile network operator. Reference [40] analyzed the sent and received stream packets to identify the 4 connected IoT devices by using 6 different machine learning models, including Decision Tree (DT) methods such as Random Forest (RF) and K-Nearest Neighbors (KNN). In addition, for security applications Reference [41] predicted IoT traffic on the edge network considering randomness and uncertainty via sample entropy. Reference [42] has developed an algorithm to predict the activation (i.e. the probability of traffic transmission) of IoT devices for each time slot.

Furthermore, Reference [43] predicted industrial IoT traffic, which is modeled as a Markovian process, via Reinforcement Learning (RL) for anomaly detection and network planning. Reference [44] predicted traffic of event-triggered IoT devices. Reference [45] used Recurrent Neural Networks (RNN) to predict the network traffic for the parameter

optimization of a random access control scheme. In order to forecast network traffic, linear predictors have been developed in [46], [47] to control sampling interval, while an LSTM based predictor has been used with deep learning in [48] and Nonlinear autoregressive exogenous (NARX) neural network has been used in [49].

In summary, the recent research uses various ML techniques such as neural networks, deep learning, and DT methods to forecast individual or aggregated IoT traffic. However, none of these works directly investigated the predictability of the traffic with respect to information-theoretic bounds. In contrast, in this work, we both analyze the predictability of IoT traffic and compute information-theoretic predictability bounds.

C. FANO'S INEQUALITY AND ITS COUNTERPART FOR PREDICTABILITY ANALYSIS

We use Fano's inequality [56] to set the information-theoretical limits for predicting IoT traffic generation. A set of previous studies used Fano's inequality primarily for mobility or location prediction. Reference [50] used three different measures of entropy, namely random entropy, uncorrelated entropy and actual entropy, with Fano's inequality to predict the mobility of individuals. Fano's inequality is also used to find the upper bound on predictability on the future locations of a person in [51] and to compute an upper bound for the prediction of an individual's location in [52]. Furthermore, Reference [53] analyzed the predictability of the radio spectrum state by using Fano's inequality and measuring statistical entropy. The works in this category used Fano's inequality to analyze the predictability of either position of individuals or radio spectrum state. Beside the works that use Fano's inequality and its counterpart, other information theory-based techniques are also developed and used in IoT networks. For example, Reference [54] developed information-aware traffic reduction to address network congestion. In addition, Reference [55] developed an anomaly detection method which captures the effects of abnormal/malicious traffic on information entropy. In contrast, in this paper, we use Fano's inequality and its counterpart to analyze the predictability of traffic generation patterns of IoT devices; to the best of authors' knowledge, there is no work that compares the predictability of IoT traffic against information-theoretical bounds based on Fano's inequality and its counterpart.

III. IoT TRAFFIC DATA AND ITS FORECASTING

A. CLASSIFICATION OF IoT TRAFFIC AND THE DATASET

Reference [37] has classified the IoT traffic in the MAC-layer into four categories with respect to its periodicity and bit representation. For a given IoT device, if the generation intervals of the bursts of traffic generated by that device are constant, such traffic is called "Periodic" traffic. Otherwise (if the generation intervals vary with time), the traffic generated by that device is called "Aperiodic". For example, if an IoT

TABLE 2. Summary of the related works on IoT traffic forecasting in Section II-B.

Paper	Aggregated Individual Traffic	Information Theory	ML Models										
			Linear	MLP ANN	LSTM RNN	CNN	Statistical (ARIMA etc.)	FSF	GB	RL	DT	SVM	
Our Study	Individual	✓	✓	✓	✓	✓	✓	✓					
Nakip et al. [37]	Individual			✓	✓	✓	✓	✓					
Nakip et al. [38]	Individual		✓	✓	✓	✓	✓	✓	✓				
Lopez-Martin et al. [39]	Aggregate			✓	✓	✓	✓	✓		✓			
Shahid et al. [40]	Aggregate			✓								✓	✓
Wan et al. [41]	Aggregate	✓											
Shoaei et al. [42]	Aggregate			✓									
Nie et al. [43]	Aggregate				✓						✓		
Ali et al. [44]	Aggregate	✓											
Jiang et al. [45]	Aggregate										✓		
Wu et al. [46], [47]	Aggregate		✓										
Abdellah et al. [48]	Aggregate				✓								
Abdellah et al. [49]	Aggregate			✓				✓					

TABLE 3. Summary of the related works in Section II-C.

Paper	Technique(s)	Focus
Our Study	Fano’s Inequality and its counterpart	Predictability bounds for IoT traffic
Song et al. [50]	Entropy with Fano’s Inequality	Predictability of human mobility
Lu et al. [51]	Entropy with Fano’s Inequality	Predictability of human mobility
Li et al. [52]	Fano’s Inequality and Clustering Link	Forecasting of human location
Ding et al. [53]	Statistical entropy with Fano’s Inequality	Predictability of the radio spectrum state
Ngai et al. [54]	Information-aware Traffic Reduction	Network congestion
Sun et al. [55]	Information Entropy	Anomaly detection

device generates traffic when triggered by an event, then that traffic is categorized as Aperiodic traffic. Furthermore, if each burst has a fixed number of bits, then such traffic is called “Fixed Bit”. In contrast, if the number of bits that in a single burst varies with time, then that traffic is called “Variable Bit”. As a result, the four categories are as follows: “Fixed Bit Periodic (FBP)”,² “Fixed Bit Aperiodic (FBA)”, “Variable Bit Periodic (VBP)”, and “Variable Bit Aperiodic (VBA)”. The traffic category in which an IoT device falls depends entirely on the characteristics of the traffic generated by that IoT device; hence, it is pre-determined and fixed for each device.³

²Note that the IoT device traffic that falls in the FBP class requires no forecasting: if the generation time and the amount are known at any one instant, they are known for the rest of the time. Hence, we focus on forecasting for only for the FBA, VBP, and the VBA classes.

³If there are multiple types of traffic generated, e.g. by distinct types of sensors on a device, we treat each distinct traffic stream separately. Hence, every traffic stream falls in only one of these four categories.

During the analysis in this paper, we use the publicly available dataset [57] whose collection and processing methodology has been presented in Reference [38]. This dataset contains the traffic generation patterns of 8 individual IoT devices. These devices and the classes in which they fall are as follows: FBA Class - Smart Home Energy Generation (SHEG), Non-Methane Hydrocarbon (NMHC) Gas Sensor, and Wind Speed; VBP Class - Light Dependent Resistor (LDR) and Relative Humidity (RH) Sensor; VBA Class - Elevator Button, NO₂ Gas Sensor, and Temperature Sensor. In addition, for each sensor i , we present the number of samples K_i and the number of unique burst sizes M_i in Table 4.

B. FORECASTING OF IoT TRAFFIC

We now describe the processing of time series IoT traffic data and the forecasting models that are used during the analysis in this paper.

TABLE 4. Properties of dataset.

Traffic Class	Device Name	Number of Samples (K_i)	Number of Unique Burst Sizes (M_i)
FBA	SHEG	8692	2
	NMHC Gas Sensor	7715	2
	Wind Speed	24999	2
VBP	Light Dependent Resistor	6834	6
	Relative Humidity Sensor	9357	6
VBA	Elevator Button	40080	15
	NO_2 Gas Sensor	2288	7
	Temperature Sensor	7371	10

TABLE 5. List of mathematical symbols in order of appearance.

Symbol	Definition
K_i	Number of samples for device i
M_i	Number of unique burst sizes for device i
x_k^i	Size of the burst that is generated by device i at discrete time k
W	Length of forecasting window
\hat{x}_{k+1}^i	The forecast size of the burst that is generated by device i at discrete time k
E	Number of MLP layers
n_e	Number of neurons at layer e of MLP
c_{CNN}	Number of filters in the convolution layer of 1D CNN
c_e	Number of neurons at hidden layer e of 1D CNN
E_{LSTM}	Number of fully connected layers in LSTM
h_{LSTM}	Number of LSTM units
h_e	Number of neurons at hidden layer e of LSTM
$T_i(W)$	Set of sequences of time series traffic of device i over time windows with duration W
$S_i(W)$	Entropy of $T_i(W)$
$P(\alpha)$	Probability of a sequence α
y_k^i	Traffic data that is generated by the memoryless source based on the traffic of device i
$\tilde{T}_i(W)$	Set of sequences for y_k^i
$\tilde{S}_i(W)$	Entropy of $\tilde{T}_i(W)$
W^*	The best value of the window size W
S_i^*	Entropy of the traffic generation of device i that is measured from real data under the optimal selection of window length, W^*
Π_i^{\max}	Maximum accuracy that any forecaster can achieve for the traffic of device i
$H(\Pi_i^{\max})$	Binary entropy function of Π_i^{\max}
$\Pi_{\text{normalized}}^{i,f}$	A metric that measures the ratio of the accuracy that is achieved by forecaster f to the upper bound of the accuracy for device i
Π_i^f	Accuracy achieved by forecaster f for device i
MSE_i^f	MSE that is measured for the predictions of forecaster f for device i
h_i^*	Differential entropy of traffic generation pattern of device i for W^*
MSE_i^{LB}	Lower bound for MSE_i^f
$\mathcal{E}_{\text{normalized}}^{i,f}$	Metric that measures how precise the predictions of forecaster f are compared to the lower bound of error for device i

First, we assume that the IoT traffic is generated in discrete time. Then, let x_k^i denote the burst that is generated by device i at discrete time k .⁴ We perform 1-step ahead forecasting based on the traffic generated in the past time window with duration W ; that is, at each discrete time k , any forecaster

predicts the value of the size of burst at $k + 1$, denoted by \hat{x}_{k+1}^i , based on $\{x_{k-m}^i\}_{m=0}^{W-1}$. As we shall present in Section IV-A, we analyze the predictability of IoT traffic for different values of W .

For the forecasting of IoT traffic at the MAC layer, in this paper, we use six different ML models, namely Logistic Regression, MLP, ANOVA-MLP, ACF-MLP, 1D CNN and

⁴Table 5 presents the list of mathematical symbols in order of appearance.

LSTM, where MLP-based models, 1D CNN and LSTM are the deep learning models. There are two main reasons that these models are selected as the forecasting schemes in this paper: 1) We aim to compare the performance of different types of (probabilistic, deterministic, static, recurrent) state of the art ML models. To this end, we represent the probabilistic and linear ML models with Logistic Regression, feed-forward multi-layer models with the widely used universal approximator MLP [58], the state of the art deep learning with 1D CNN [59], and recurrent models with LSTM (which is able to capture temporal dependence in time-series data). 2) References [37], [38] have shown that these models are able to achieve successful results for the forecasting of IoT traffic at the MAC-layer. For these models, the parameter settings that are used during this paper are determined as follows:

- **Logistic Regression:** In the implementation of the Logistic Regression model, we perform an exhaustive search for the local optimal value of the regularization term in the set $\{0.01, 0.1, 1, 10\}$; the method of penalty in the set $\{\text{none}, 11, 12, \text{elasticnet}\}$; the solver in the set $\{\text{newton's method}, \text{limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm}, \text{library for large linear classification}, \text{stochastic average gradient}, \text{stochastic average gradient with non-smooth penalty}\}$. For the implementation of Logistic Regression, we use the *scikit-learn* library [60] in Python.
- **Multi-Layer Perceptron (MLP):** In the design of the MLP model, the architectural hyperparameters are as follows: the number of layers E , the number of neurons n_e at each layer $e \in \{1, \dots, E\}$, and the activation function of each neuron. First, we set the value of E to 4 (which includes the output layer). Then, in order to make a 1-step ahead forecast, we set the $n_4 = 1$; that is, there is only one neuron in the output layer. Furthermore, we select the activation function of each neuron in each of the other three layers as *ReLU* (Rectified Linear Unit), and the activation function of each neuron at $e = 4$ (the output layer) as linear.⁵ Finally, we find the local optimal value of n_e for each layer $e \in \{1, 2, 3\}$ within the range $[2, 256]$ by using only integral powers of 2 for each hyperparameter by searching over 100 randomly generated points in the $E-1$ dimensional space, where $E-1$ is the number of hidden layers in the neural network. We shall refer to this particular search method as “random search” in the rest of this paper. We also use MLP with two different feature selection methods, namely ANOVA-based and ACF-based feature selection. In the ANOVA-based feature selection, we first compute the F-ratio [61] between each feature and the desired output. Then, we sort all features with respect to their F-ratios in descending order and select the first twelve features from this sorted sequence of fea-

tures. In the ACF-based feature selection, based on the ACF of each dataset, we select the following features: $\{1, 2, 3, 4, 5\} \cup \{12j\}_{j=1}^{10}$ for RH, the first 3 samples for NO_2 , $\{23j\}_{j=1}^5$, and the highest 12 values of the ACF for LDR, Temperature, Elevator Button, Wind Speed and SHEG.

- **1-Dimensional Convolutional Neural Network (1D CNN):** The internal architecture of the 1D CNN model in our studies is comprised of a convolution layer, a max pooling layer and four fully connected layers. We set the kernel size of each of the convolution layers and the max pooling layer to 3. We also set the stride of the convolution layer to 2. We set the activation function of the convolution layer and that of each neuron at each of the first three hidden layers to *ReLU*. We do not use any activation function for any neuron at the output layer. Furthermore, we set the number of neurons at the output layer to 1. Then, we perform random search in order to find the local optimal number c_{CNN} of convolution filters at the convolution layer and the local optimal number c_e of neurons at each hidden layer $e \in \{1, 2, 3\}$. We select the search intervals and the points in each search interval as follows: The interval for c_{CNN} is set to $[2, 256]$ sampled at integral powers of 2. The interval for c_e for each $e \in \{1, 2, 3\}$ is set to $[2, 256]$ and is sampled at multiples of 2.
- **Long-Short Term Memory (LSTM):** In the internal architecture of the LSTM, the following hyperparameters need to be determined: the number of LSTM layers, the number of LSTM units in each LSTM layer, the number E_{LSTM} of fully connected layers, the number of neurons in each hidden layer, and the activation function of each LSTM unit or neuron at each layer. We use one LSTM layer and three hidden layers, where the last hidden layer is the output layer. The activation function of each unit in the LSTM layer and of each neuron in the first two hidden layers is selected as *ReLU*. We use the *linear* activation function for all of the neurons at the output layer. Furthermore, there is only one neuron in the output layer to perform 1-step ahead forecasting. Then, we perform random search (as defined for MLP above) in order to find the local optimal number of LSTM units, which is denoted by h_{LSTM} , as well as the local optimal number of neurons, denoted by h_e in each of the layers $e \in \{1, 2\}$. The range of the search interval is $[2, 256]$. We use integral powers of 2 in this interval for the value of h_{LSTM} , and multiples of 2 for each $e \in \{1, 2\}$.

IV. ANALYSIS OF PREDICTABILITY AGAINST INFORMATION-THEORETIC BOUNDS

In this section, we aim to analyze the predictability of IoT traffic against information-theoretic bounds. To this end, we first present the calculation of the entropy of IoT traffic that is chopped up into sequential time windows, and we show the selection of the best value of the window size W . Then, using this entropy calculation, we analyze the predictability

⁵Since the traffic patterns in each of the VBA and the FBA classes include samples that correspond to zero traffic generation, we use *softmax* as the activation function at the output layer of all of the forecasting models.

of traffic in the FBA class based on Fano’s inequality [56] and the predictability of traffic in the VBP and VBA classes based on the estimation-theoretic counterpart to Fano’s inequality.

Specifically, we use Fano’s inequality to relate the randomness inherent in time-series binary data (as quantified by the entropy of the probabilistic model of those data) to the probability of predicting those data correctly. Intuitively, the more random the binary data, the lower the probability of correct prediction. Fano’s inequality gives the precise relationship between the entropy and the probability of correct prediction. This relationship is independent of the particular forecaster used and must be satisfied by all forecasters. We apply Fano’s inequality to FBA data, which is a binary-valued time series, in which the device either generates no bits or a constant number of bits at each discrete time.

The estimation-theoretic counterpart to Fano’s inequality generalizes this relationship from binary data to data that take values on a continuum. In this case, the probability of incorrect prediction is replaced appropriately by the MSE. Then, the counterpart to Fano’s inequality specifies the precise relationship between the differential entropy of the probabilistic model of the data and the lowest possible MSE that can be achieved by any forecaster on such data. We apply Fano’s inequality to VBP and VBA data. The multiple values taken on by data in these classes are approximated by a continuum, and the deviations are measured via the MSE.

A. ENTROPY OF IoT TRAFFIC WITH SEQUENTIAL TIME WINDOWS

First, let the time series IoT traffic be the collection of bursts that are generated by device i as $\{x_k^i\}_{k=0}^{K_i-1}$, where K_i denotes the total number of samples for device i in the considered dataset. Then, we split this time series traffic of device i into sequences over time windows of duration W , and we compute the set of sequences, denoted by $T_i(W)$ as

$$T_i(W) = \{\{x_k^i\}_{k=wW+1}^{(w+1)W}\}_{w=0}^{\lfloor (K_i-1)/W \rfloor} \tag{1}$$

Accordingly, the entropy of $T_i(W)$, denoted by $S_i(W)$, is computed as

$$S_i(W) = - \sum_{T' \in T_i(W)} P(T') \log_2 P(T') \tag{2}$$

where $P(\alpha)$ denotes the probability of sequence α .

Now, we aim to determine the best value of W by comparing the actual entropy $S_i(W)$ (calculated in (2)) with the entropy for the case where the generator of the traffic is a memoryless source that is distributed uniformly over the alphabet of the source at device i .⁶ We let y_k^i denote the traffic data that is generated by such a memoryless source that has the same distribution as the instantaneous traffic generated by device i . Then, the set of sequences for y_k^i , denoted by $\tilde{T}_i(W)$,

⁶In our implementation, we consider only those values of data that are actually observed to be generated by the source to belong to the source alphabet.

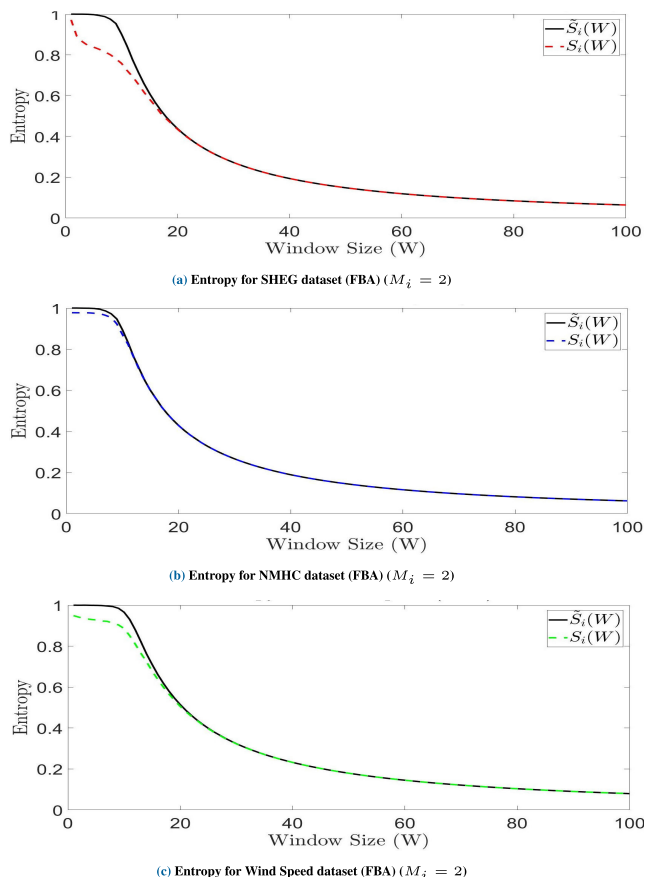


FIGURE 2. Determination of W^* for each device in the FBA class via the comparison of $S_i(W)$ and $\tilde{S}_i(W)$ for $W \in \{1, \dots, 100\}$ resulted in $W^* = 13$ for (a) SHEG, $W^* = 9$ for (b) NMHC and $W^* = 15$ for (c) Wind speed.

is computed as

$$\tilde{T}_i(W) = \{\{y_k^i\}_{k=wW+1}^{(w+1)W}\}_{w=0}^{\lfloor (K_i-1)/W \rfloor} \tag{3}$$

Accordingly, the entropy of the sequences for the memoryless source, denoted by \tilde{S}_i , is calculated as

$$\tilde{S}_i(W) = - \sum_{T' \in \tilde{T}_i(W)} P(T') \log_2 P(T') \tag{4}$$

Subsequently, we compare $S_i(W)$ and $\tilde{S}_i(W)$ for device i in each of FBA, VBA and VBP classes in Figures 2-4, respectively, and we select the best value of the window size W as

$$W^* = \arg \min_{W \in \{1, \dots, 100\}} (S_i(W) | S_i(W) \geq \tilde{S}_i(W) - \epsilon) \tag{5}$$

where W^* is selected to minimize the actual entropy $S_i(W)$ while keeping the difference between $S_i(W)$ and $\tilde{S}_i(W)$ greater than or equal to a threshold ϵ , and we set $\epsilon = \frac{3}{100} \max S_i(W)$. In this way, we aim to minimize the actual entropy while there remains a gap between the actual entropy and the entropy of the memoryless source.

Fig. 2 displays the comparison between $S_i(W)$ and $\tilde{S}_i(W)$ for $W \in \{1, \dots, 100\}$ for each of (a) SHEG, (b) NMHC and

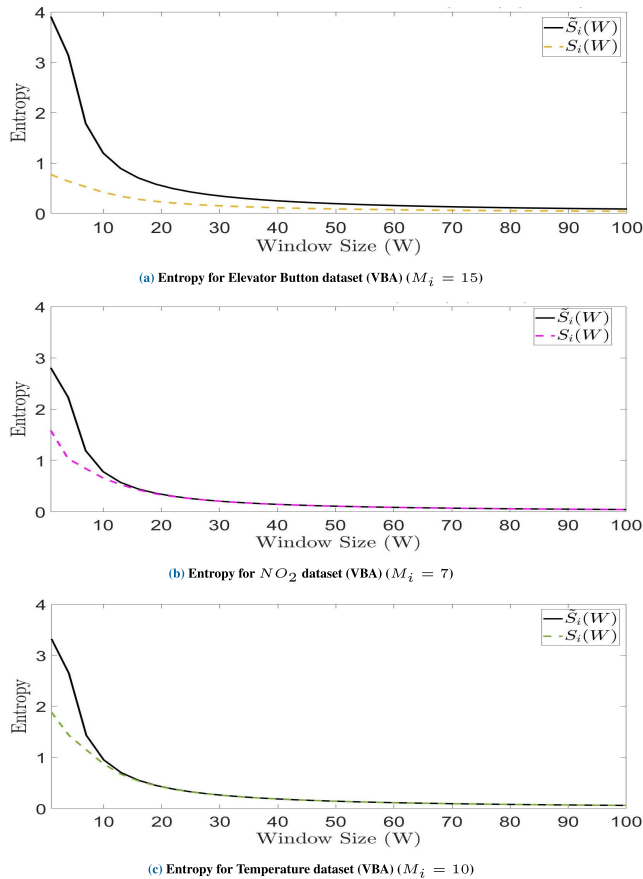


FIGURE 3. Determination of W^* for each device in the VBA class via the comparison of $S_i(W)$ and $\tilde{S}_i(W)$ for $W \in \{1, \dots, 100\}$ resulted in $W^* = 100$ for (a) Elevator button, $W^* = 12$ for (b) NO_2 and $W^* = 10$ for (c) Temperature.

(c) Wind Speed. The results in Fig. 2 (a) show that although $S_i(W)$ decreases as W increases, the gap between $S_i(W)$ and $\tilde{S}_i(W)$ also decreases with increasing W and becomes almost zero for $W > 20$. The reason is that the number of samples is too low for $W > 20$ that the actual data is very similar to the data of the memoryless source. In addition, we see that the gap between $S_i(W)$ and $\tilde{S}_i(W)$ decreases below 3% of $S_i(W)$ (which is the value of ϵ in (5)) after $W = 13$; thus, we select $W^* = 13$. Similarly, based on the results in Fig. 2 (b) and Fig. 2 (c), we select $W^* = 9$ for NMHC and $W^* = 15$ for Wind Speed.

Fig. 3 presents the comparison between $S_i(W)$ and $\tilde{S}_i(W)$ for $W \in \{1, \dots, 100\}$ for each of (a) Elevator Button, (b) NO_2 and (c) Temperature. Based on the results in Fig. 3 (b), we select $W^* = 12$ for NO_2 , and based on the results in Fig. 3 (c), $W^* = 10$ for Temperature. In addition, the results in Fig. 3 (a) show that the gap between $S_i(W)$ and $\tilde{S}_i(W)$ does not decrease below 3% of $S_i(W)$ until $W = 100$; thus, $W^* = 100$ for Elevator Button.

Finally, in Fig. 4, we present the comparison between $S_i(W)$ and $\tilde{S}_i(W)$ for $W \in \{1, \dots, 100\}$ for each of (a) LDR and (b) RH. For the VBP class, we select $W^* = 22$ for LDR

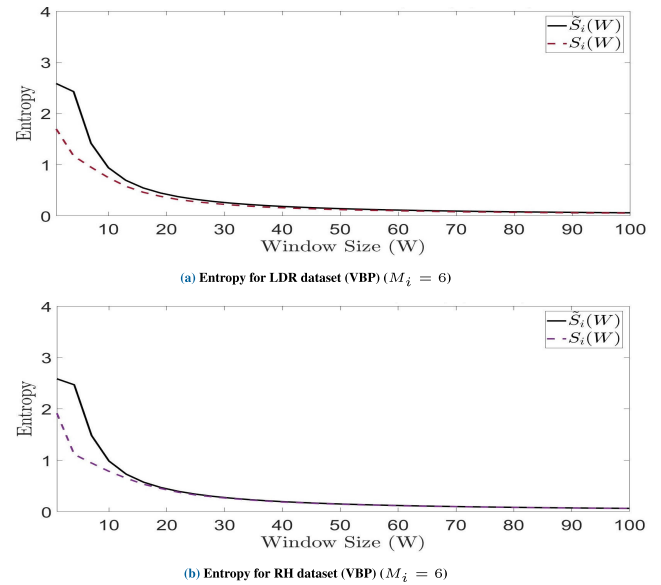


FIGURE 4. Determination of W^* for each device in the VBP class via the comparison of $S_i(W)$ and $\tilde{S}_i(W)$ for values of $W \in \{1, \dots, 100\}$ resulted in $W^* = 22$ for (a) LDR and (b) $W^* = 13$ for RH.

and $W^* = 13$ for RH respectively based on the results in Fig. 4 (a) and Fig. 4 (b).

B. ANALYSIS FOR FIXED BIT - EVENT-TRIGGERED TRAFFIC (FBA) BASED ON FANO'S INEQUALITY

We now present an analysis of the predictability of IoT traffic in the FBA class by calculating an upper bound for the accuracy achievable with any forecasting scheme.

First, we let S_i^* denote the entropy of the traffic generation of device i that is measured from real data under the optimal selection of window length, W^* . Also, let Π_i^{max} denote the maximum accuracy that any forecaster can achieve for the traffic of device i .

Then, the well-known Fano's inequality [56] is

$$S_i^* \leq H(\Pi_i^{max}) + (1 - \Pi_i^{max})\log_2(M_i - 1), \quad (6)$$

where $H(\Pi_i^{max})$ is the binary entropy function which is defined as

$$H(\Pi_i^{max}) \equiv -\Pi_i^{max}\log_2(\Pi_i^{max}) - (1 - \Pi_i^{max})\log_2(1 - \Pi_i^{max}). \quad (7)$$

Thus, in our case for the traffic generation patterns in the FBA class where $M_i = 2$, (6) simplifies to

$$S_i^* \leq H(\Pi_i^{max}). \quad (8)$$

Based on (8), we calculate the upper bound Π_i^{max} for the accuracy which is presented in Table 6 for each device i in the dataset [57]. The results in this table show that the upper bound for accuracy is slightly above 0.8 for SHEG and Wind Speed devices; that is, the traffic generation of these devices is more than 80% predictable. On the other hand, while Π_i^{max} for NMHC equals 0.66, the traffic patterns of SHEG and

TABLE 6. Value of Π_i^{\max} for each device i that falls in the FBA class in the dataset.

Datasets	SHEG	NMHC	Wind Speed
Π_i^{\max}	0.83	0.66	0.82

Wind Speed devices are significantly more predictable than that of NMHC.

Subsequently, we analyze the performance of distinct forecasting schemes which have been used for the forecasting of IoT traffic in recent works [37], [38]. To this end, we define ‘‘Normalized Accuracy’’, denoted by $\Pi_{\text{normalized}}^{i,f}$ as a metric that measures the ratio of the accuracy that is achieved by forecaster f to the upper bound of the accuracy for device i ; that is,

$$\Pi_{\text{normalized}}^{i,f} \equiv \frac{\Pi_i^f}{\Pi_i^{\max}}, \quad (9)$$

where Π_i^f is the accuracy achieved by forecaster f . Note that for the traffic of device i , $\Pi_{\text{normalized}}^{i,f}$ is the ratio of the accuracy that is achieved by forecaster f to the accuracy upper bound.

Fig. 5 displays the values of $\Pi_{\text{normalized}}^{i,f}$ for various forecasting schemes for each of the SHEG, NMHC and Wind Speed devices. In Fig. 5 (a), for the SHEG device, we see that Logistic Regression, MLP, LSTM and ANOVA-MLP forecasters achieve at least 85% of the accuracy upper bound, where Logistic Regression is the best performing model with $\Pi_{\text{normalized}}^{i,f} = 0.89$. In Fig. 5 (b), the results for NMHC device show that all of the forecasters achieves more than 80% of the upper bound, while the value of $\Pi_{\text{normalized}}^{i,f}$ is in the range [0.8, 0.89] for all forecasters, where ACF-MLP is the best and LSTM is the worst performing model. In Fig. 5 (c), for Wind Speed device, we see that LSTM significantly outperforms all of the other forecasters by achieving 89% of the accuracy upper bound, Π_i^{\max} .

C. ANALYSIS FOR VARIABLE BIT TRAFFIC (VBA & VBP) BASED ON COUNTERPART TO FANO’S INEQUALITY

We now extend our analysis for the case where the IoT devices generate varying numbers of bits in each burst. To this end, we use the estimation counterpart to Fano’s inequality (given by Theorem 8.6.6 in [62]), which is adapted for the predictability analysis of Variable Bit Traffic as

$$\text{MSE}_i^f \geq \frac{1}{2\pi e} e^{2h_i^*}, \quad (10)$$

where MSE_i^f is the Mean Squared Error (MSE) that is measured for the predictions of forecaster f for device i , and h_i^* denotes differential entropy of traffic generation pattern of device i for W^* .

According to Equation (8.94) in [62], differential entropy can be used to bound the discrete entropy. The application of

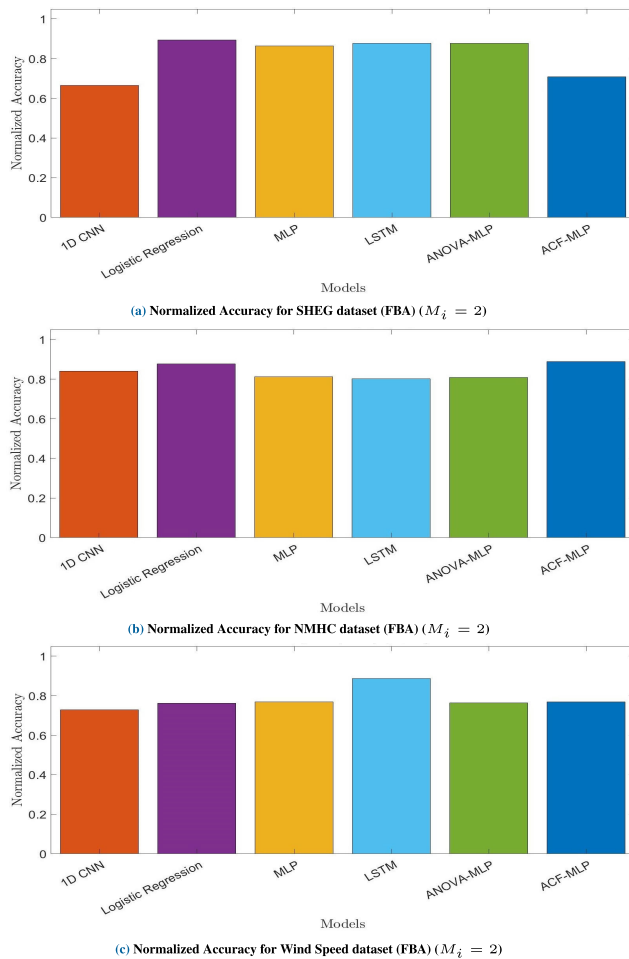


FIGURE 5. Comparison of various forecasters with respect to $\Pi_{\text{normalized}}^{i,f}$ i.e. normalized accuracy, for SHEG, NMHC and wind speed devices whose traffic fall in FBA class.

this bound to our case gives $h_i^* \geq S_i^*$. Hence, we set a lower bound to the forecasting error as

$$\text{MSE}_i^f \geq \frac{1}{2\pi e} e^{2h_i^*} \geq \frac{1}{2\pi e} e^{2S_i^*}, \quad (11)$$

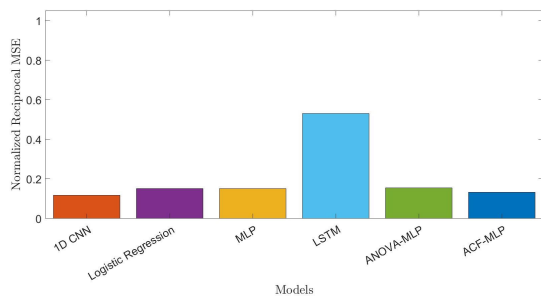
and we compute the lower bound for MSE_i^f , denoted by MSE_i^{LB} , as

$$\text{MSE}_i^{\text{LB}} = \frac{1}{2\pi e} e^{2S_i^*}. \quad (12)$$

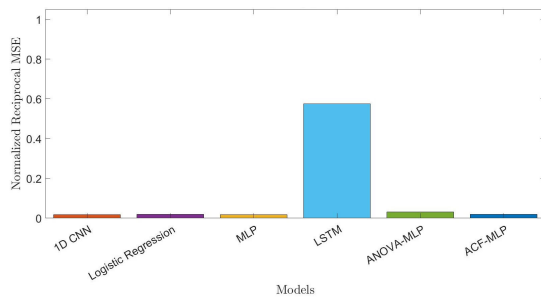
First, in Table 7, we present the lower error bound MSE_i^{LB} for each device i whose traffic generation pattern falls in the VBA or VBP class. For the VBA class (the first three rows of this table), we see that the traffic generated by NO_2 device is the most predictable with $\text{MSE}_i^{\text{LB}} = 0.18$ compared with the traffic generated by Elevator Button and that by Temperature. Furthermore, for the VBP class (in the lower part of this table), we see that the value of MSE_i^{LB} for the LDR device is lower than that for RH, which shows that the traffic generation pattern of LDR is more predictable than that of RH. Finally, based on the results for both VBA and VBP

TABLE 7. Value of MSE_i^{LB} for each device i that falls in the VBA and VBP classes in the dataset.

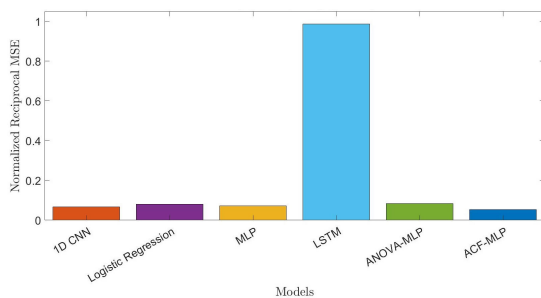
VBA			
Datasets	Elevator Button	NO_2	Temperature
\mathcal{E}_i^{\min}	0.63	0.18	0.33
VBP			
Datasets	LDR	RH	
\mathcal{E}_i^{\min}	0.11	0.21	



(a) Normalized Reciprocal MSE for Elevator Button dataset (VBA) ($M_i = 15$)



(b) Normalized Reciprocal MSE for NO_2 dataset (VBA) ($M_i = 7$)

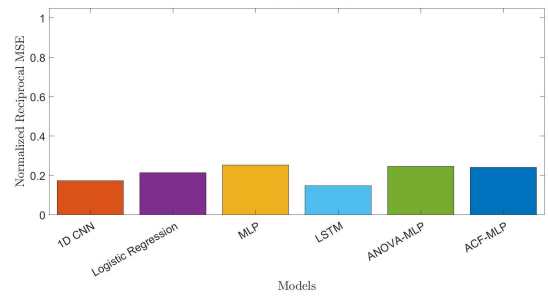


(c) Normalized Reciprocal MSE for Temperature dataset (VBA) ($M_i = 10$)

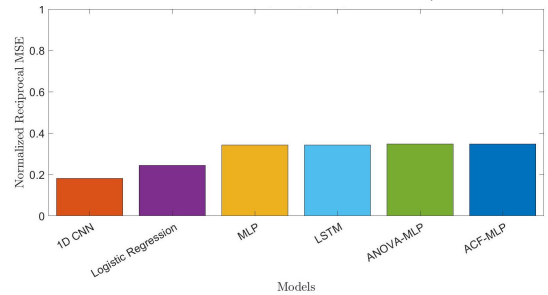
FIGURE 6. Comparison of various forecaster with respect to $\mathcal{E}_i^{i,f}$ normalized, i.e. normalized reciprocal MSE, for (a) Elevator button, (b) NO_2 and (c) Temperature devices whose traffic generation patterns fall in VBA class.

classes, the order of devices with respect to predictability is as follows: LDR, NO_2 , RH, Temperature, and Elevator Button.

Next, we analyze the performance of different forecasting schemes which are presented in Section III-B. To this end, we define ‘‘Normalized Reciprocal MSE’’, denoted by $\mathcal{E}_i^{i,f}$ normalized, as a metric that measures how precise the predic-



(a) Normalized Reciprocal MSE for LDR dataset (VBP) ($M_i = 6$)



(b) Normalized Reciprocal MSE for RH dataset (VBP) ($M_i = 6$)

FIGURE 7. Comparison of various forecaster with respect to $\mathcal{E}_i^{i,f}$ normalized, i.e. normalized reciprocal MSE, for (a) LDR and (b) RH devices whose traffic generation patterns fall in VBP class.

tions of forecaster f are compared to the lower bound of error for device i ; that is,

$$\mathcal{E}_{\text{normalized}}^{i,f} \equiv \frac{MSE_i^{LB}}{MSE_i^f}. \tag{13}$$

We present $\mathcal{E}_{\text{normalized}}^{i,f}$ for each of the Elevator Button, NO_2 , and Temperature devices in Fig. 6. For all of the devices whose traffic generation patterns fall in the VBA class, we see that LSTM significantly outperforms other forecasters. In addition, LSTM achieves almost the maximum normalized reciprocal MSE with 0.99 (i.e. MSE under LSTM is almost equal to the lower error bound) for Temperature, and it also achieves 0.6 Normalized Reciprocal MSE for Elevator Button and NO_2 devices.

Finally, Fig. 7 displays $\mathcal{E}_{\text{normalized}}^{i,f}$ for the LDR and RH devices whose traffic belong to the VBP class. For the LDR device, our results in Fig. 7 (a) show that the MLP-based forecasters (MLP, ANOVA-MLP and ACF-MLP) slightly outperform the other forecasters with $\mathcal{E}_{\text{normalized}}^{i,f} \approx 0.25$. On the other hand, for this set of results, we see that none of the forecasters can perform better than $4 \times MSE_i^{LB}$. Moreover, for the RH device in Fig. 7 (b), we see that none of the forecasters can perform better than almost $3 \times MSE_i^{LB}$ while LSTM and the MLP-based forecasters perform better than 1D CNN and Logistic Regression with $\mathcal{E}_{\text{normalized}}^{i,f} \approx 0.34$.

V. CONCLUSION

In this paper, we have presented the analysis on the predictability of IoT traffic at the MAC layer and have defined information-theoretic performance bounds (i.e. upper bound for accuracy and lower bound for MSE) by using Fano’s

inequality and the counterpart to Fano's inequality. We also have compared the performance of well-known ML techniques against these bounds. We have performed the analysis as well as the performance comparison on a publicly available dataset contains the raw traffic generation patterns of various IoT devices.

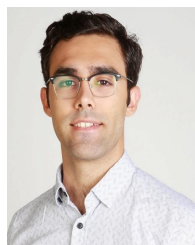
The results of our analysis show that the traffic generation pattern of the majority of considered IoT devices is predictable with more than 80% accuracy or less than 0.4 MSE. On the other hand, the comparison results in this paper show that the ML techniques used to forecast IoT traffic in the current literature perform very close to the upper bound for FBA traffic class while the performance of the majority of these techniques is far from the performance bound for VBA or VBP traffic classes.

In our future work, we shall (1) develop advanced Neural Network architectures as well as ML algorithms specific to each traffic class (FBA, VBA, and VBP); (2) investigate the development of novel forecasting schemes that approaches information-theoretic bounds, especially in the case of Variable Bit traffic; (3) develop predictive MAC-layer protocols that exploit the information-theoretic bounds on forecasting performance presented in this paper.

REFERENCES

- [1] R. Sekaran, S. N. Goddumarri, S. Kallam, M. Ramachandran, R. Patan, and D. Gupta, "5G integrated spectrum selection and spectrum access using AI-based frame work for IoT based sensor networks," *Comput. Netw.*, vol. 186, Feb. 2021, Art. no. 107649. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1389128620312676>
- [2] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.
- [3] H. Jin, W. Toor, B. C. Jung, and J.-B. Seo, "Recursive pseudo-Bayesian access class barring for M2M communications in LTE systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8595–8599, Sep. 2017.
- [4] L. Liang, L. Xu, B. Cao, and Y. Jia, "A cluster-based congestion-mitigating access scheme for massive M2M communications in Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2200–2211, Jun. 2018.
- [5] L. Tello-Quendo, I. Leyva-Mayorga, V. Pla, J. Martinez-Bauset, J.-R. Vidal, V. Casares-Giner, and L. Guijarro, "Performance analysis and optimal access class barring parameter configuration in LTE-A networks with massive M2M traffic," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3505–3520, Apr. 2018.
- [6] L. Tello-Quendo, D. Pacheco-Paramo, V. Pla, and J. Martinez-Bauset, "Reinforcement learning-based ACB in LTE-A networks for handling massive M2M and H2H communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–7.
- [7] N. Jiang, Y. Deng, A. Nallanathan, and J. Yuan, "A decoupled learning strategy for massive access optimization in cellular IoT networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 668–685, Mar. 2021.
- [8] A. O. Almagrabi, R. Ali, D. Alghazzawi, A. AlBarakati, and T. Khurshaid, "A Poisson process-based random access channel for 5G and beyond networks," *Mathematics*, vol. 9, no. 5, p. 508, Mar. 2021.
- [9] F. Amin, A. Ahmad, and G. S. Choi, "Towards trust and friendliness approaches in the social Internet of Things," *Appl. Sci.*, vol. 9, no. 1, p. 166, 2019.
- [10] M. Li, X. Guan, C. Hua, C. Chen, and L. Lyu, "Predictive pre-allocation for low-latency uplink access in industrial wireless networks," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2018, pp. 306–314.
- [11] M. Shehab, A. K. Hagelskjar, A. E. Kalor, P. Popovski, and H. Alves, "Traffic prediction based fast uplink grant for massive IoT," in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Aug. 2020, pp. 1–6.
- [12] E. Eldeeb, M. Shehab, and H. Alves, "A learning-based fast uplink grant for massive IoT via support vector machines and long short-term memory," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3889–3898, Mar. 2022.
- [13] M. Nakip, V. Rodoplu, C. Guzelis, and D. T. Eliyi, "Joint forecasting-scheduling for the Internet of Things," in *Proc. IEEE Global Conf. Internet Things (GCloT)*, Dec. 2019, pp. 1–7.
- [14] V. Rodoplu, M. Nakip, D. T. Eliyi, and C. Guzelis, "A multiscale algorithm for joint forecasting-scheduling to solve the massive access problem of IoT," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8572–8589, Sep. 2020.
- [15] V. Rodoplu, M. Nakip, R. Qorbanian, and D. T. Eliyi, "Multi-channel joint forecasting-scheduling for the Internet of Things," *IEEE Access*, vol. 8, pp. 217324–217354, 2020.
- [16] M. Nakip and E. Gelenbe, "Randomization of data generation times improves performance of predictive IoT networks," in *Proc. IEEE 7th World Forum Internet Things (WF-IoT)*, Jun. 2021, pp. 350–355.
- [17] M. Nakip, A. Helva, C. Guzelis, and V. Rodoplu, "Subspace-based emulation of the relationship between forecasting error and network performance in joint forecasting-scheduling for the Internet of Things," in *Proc. IEEE 7th World Forum Internet Things (WF-IoT)*, Jun. 2021, pp. 247–252.
- [18] M. Nakip, E. Çakan, V. Rodoplu, and C. Güzelis, "Dynamic automatic forecaster selection via artificial neural network based emulation to enable massive access for the Internet of Things," *J. Netw. Comput. Appl.*, vol. 201, pp. 103360–103379, May 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1084804522000273>
- [19] E. Gelenbe, M. Nakip, D. Marek, and T. Czachorski, "Diffusion analysis improves scalability of IoT networks to mitigate the massive access problem," in *Proc. 29th Int. Symp. Model., Anal., Simul. Comput. Telecommun. Syst. (MASCOTS)*, Nov. 2021, pp. 1–8.
- [20] E. Gelenbe and K. Sigman, "IoT traffic shaping and the massive access problem," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Seoul, South Korea, May 2022, pp. 1–6.
- [21] Y. Edalat, J.-S. Ahn, and K. Obraczka, "Smart experts for network state estimation," *IEEE Trans. Netw. Service Manage.*, vol. 13, no. 3, pp. 622–635, Sep. 2016.
- [22] V. Petkov and K. Obraczka, "The case for using traffic forecasting in schedule-based channel access," in *Proc. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2011, pp. 208–212.
- [23] V. Petkov and K. Obraczka, "Collision-free medium access based on traffic forecasting," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2012, pp. 1–9.
- [24] L. Ruan, M. P. I. Dias, and E. Wong, "Machine learning-based bandwidth prediction for low-latency H2M applications," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3743–3752, Apr. 2019.
- [25] H.-Y. Kim and J.-M. Kim, "A load balancing scheme based on deep-learning in IoT," *Cluster Comput.*, vol. 20, no. 1, pp. 873–878, 2017.
- [26] D. Raca, A. H. Zahran, C. J. Sreenan, R. K. Sinha, E. Halepovic, R. Jana, and V. Gopalakrishnan, "On leveraging machine and deep learning for throughput prediction in cellular networks: Design, performance, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 11–17, Mar. 2020.
- [27] M. Patnaik, V. Kamakoti, V. Matyas, and V. Rehak, "PROLEMus: A proactive learning-based MAC protocol against PUEA and SSDF attacks in energy constrained cognitive radio networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 2, pp. 400–412, Jun. 2019.
- [28] J. Brown and J. Y. Khan, "A predictive resource allocation algorithm in the LTE uplink for event based M2M applications," *IEEE Trans. Mobile Comput.*, vol. 14, no. 12, pp. 2433–2446, Dec. 2015.
- [29] C. Yao, J. Guo, and C. Yang, "Achieving high throughput with predictive resource allocation," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 768–772.
- [30] R. Atawia, H. S. Hassanein, and A. Noureldin, "Optimal and robust QoS-aware predictive adaptive video streaming for future wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.
- [31] R. Atawia, H. S. Hassanein, H. Abou-zeid, and A. Noureldin, "Robust content delivery and uncertainty tracking in predictive wireless networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2327–2339, Apr. 2017.
- [32] N. Khambari, B. Ghita, and L. Sun, "QoE-driven video enhancements in wireless networks through predictive packet drops," in *Proc. IEEE 13th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2017, pp. 355–361.
- [33] J. Wen, M. Sheng, J. Li, and K. Huang, "Assisting intelligent wireless networks with traffic prediction: Exploring and exploiting predictive causality in wireless traffic," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 26–31, Jun. 2020.

- [34] E. Soltanmohammadi, K. Ghavami, and M. Naraghi-Pour, "A survey of traffic issues in machine-to-machine communications over LTE," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 865–884, Dec. 2016.
- [35] D. Astely, M. Stattin, G. Wikstrom, J. Cheng, A. Hoglund, M. Frenne, and F. Gunnarsson, "LTE release 14 outlook," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 44–49, Jun. 2016.
- [36] S. Ali, N. Rajatheva, and W. Saad, "Fast uplink grant for machine type communications: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 97–103, Mar. 2019.
- [37] M. Nakip, B. C. Gül, V. Rodoplu, and C. Güzelis, "Comparative study of forecasting schemes for IoT device traffic in machine-to-machine communication," in *Proc. 4th Int. Conf. Cloud Comput. Internet Things*, Sep. 2019, pp. 102–109.
- [38] M. Nakip, K. Karakayali, C. Guzelis, and V. Rodoplu, "An end-to-end trainable feature selection-forecasting architecture targeted at the Internet of Things," *IEEE Access*, vol. 9, pp. 104011–104028, 2021.
- [39] M. Lopez-Martin, B. Carro, and A. Sanchez-Esguevillas, "Neural network architecture based on gradient boosting for IoT traffic prediction," *Future Gener. Comput. Syst.*, vol. 100, pp. 656–673, Nov. 2019.
- [40] M. R. Shahid, G. Blanc, Z. Zhang, and H. Debar, "IoT devices recognition through network traffic analysis," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 5187–5192.
- [41] Y. Wan, K. Xu, F. Wang, and G. Xue, "Characterizing and mining traffic patterns of IoT devices in edge networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 1, pp. 89–101, Jan. 2020.
- [42] A. D. Shoaie, D. T. Nguyen, and T. Le-Ngoc, "Traffic prediction for reconfigurable access scheme in correlated traffic MTC networks," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 953–958.
- [43] L. Nie, Z. Ning, M. S. Obaidat, B. Sadoun, H. Wang, S. Li, L. Guo, and G. Wang, "A reinforcement learning-based network traffic prediction mechanism in intelligent Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 2169–2180, Mar. 2021.
- [44] S. Ali, W. Saad, and N. Rajatheva, "A directed information learning framework for event-driven M2M traffic prediction," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2378–2381, Nov. 2018.
- [45] N. Jiang, Y. Deng, and A. Nallanathan, "Traffic prediction and random access control optimization: Learning and non-learning-based approaches," *IEEE Commun. Mag.*, vol. 59, no. 3, pp. 16–22, Mar. 2021.
- [46] S. Wu, W. Mao, C. Liu, and T. Tang, "Dynamic traffic prediction with adaptive sampling for 5G HetNet IoT applications," *Wireless Commun. Mobile Comput.*, vol. 2019, pp. 1–11, Jun. 2019.
- [47] S. Wu, W. Mao, T. Hong, C. Liu, and M. Kadoch, "Compressed sensing based traffic prediction for 5G HetNet IoT video streaming," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 1901–1906.
- [48] A. R. Abdellah and A. Koucheryavy, "Deep learning with long short-term memory for IoT traffic prediction," in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. New York, NY, USA: Springer, 2020, pp. 267–280.
- [49] A. R. Abdellah, O. A. K. Mahmood, A. Paramonov, and A. Koucheryavy, "IoT traffic prediction using multi-step ahead prediction with neural network," in *Proc. 11th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT)*, Oct. 2019, pp. 1–4.
- [50] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [51] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, "Approaching the limit of predictability in human mobility," *Sci. Rep.*, vol. 3, no. 1, pp. 1–9, Dec. 2013.
- [52] F. Li, Q. Li, Z. Li, Z. Huang, X. Chang, and J. Xia, "A personal location prediction method based on individual trajectory and group trajectory," *IEEE Access*, vol. 7, pp. 92850–92860, 2019.
- [53] G. Ding, J. Wang, Q. Wu, Y.-D. Yao, R. Li, H. Zhang, and Y. Zou, "On the limits of predictability in real-world radio spectrum state dynamics: From entropy theory to 5G spectrum sharing," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 178–183, Jul. 2015.
- [54] E. C.-H. Ngai, E. Gelenbe, and G. Humber, "Information-aware traffic reduction for wireless sensor networks," in *Proc. IEEE 34th Conf. Local Comput. Netw.*, Oct. 2009, pp. 451–458.
- [55] Y. Sun, J. Yu, J. Tian, Z. Chen, W. Wang, and S. Zhang, "IoT-IE: An information-entropy-based approach to traffic anomaly detection in Internet of Things," *Secur. Commun. Netw.*, vol. 2021, pp. 1–13, Dec. 2021.
- [56] R. M. Fano, "Transmission of information: A statistical theory of communications," *Amer. J. Phys.*, vol. 29, no. 11, pp. 793–794, 1961.
- [57] (Dec. 2021). *EU-QoSIoTSmartCities846077DataSet1*. [Online]. Available: <https://www.kaggle.com/mariecurieeu/euqosiotsmartcities846077dataset1>
- [58] F. Scarselli and A. C. Tsoi, "Universal approximation using feedforward neural networks: A survey of some existing methods, and some new results," *Neural Netw.*, vol. 11, no. 1, pp. 15–37, Jan. 1998.
- [59] Y. Chen, B. Zheng, Z. Zhang, Q. Wang, C. Shen, and Q. Zhang, "Deep learning on mobile and embedded devices: State-of-the-art, challenges, and future directions," *ACM Comput. Surv.*, vol. 53, no. 4, pp. 1–37, Jul. 2021, doi: 10.1145/3398209.
- [60] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Dec. 2011.
- [61] D. L. Hahs-Vaughn and R. G. Lomax, *Statistical Concepts: A Second Course*. Evanston, IL, USA: Routledge, 2020.
- [62] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley Series in Telecommunications and Signal Processing), 2nd ed. Hoboken, NJ, USA: Wiley, Jul. 2006.



MERT NAKIP (Student Member, IEEE) received the B.Sc. (Hons.) and M.Sc. degrees in electrical electronics engineering from Yaşar University, İzmir, Turkey, in 2018 and 2020, respectively, and the Ph.D. degree from the Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Gliwice, Poland. His design of a multi-sensor fire detector via machine learning methods was ranked #1 nationally with the Industry-Focused Undergraduate Graduation Projects Competition organized by TÜBİTAK (Turkish Scientific and Technological Research Council). His thesis focused on the application of machine learning methods to IoT and was supported by the National Graduate Scholarship Program of TÜBİTAK 2210C in High-Priority Technological Areas. He is currently a Research Assistant with the Institute of Theoretical and Applied Informatics, Polish Academy of Sciences. He also participates in European Commission H2020 Program under the IoTAC Research and Innovation Action as a Researcher.



BARAN CAN GÜL received the B.S. degree (Hons.) in electrical electronics engineering from Yaşar University, İzmir, in 2019. He is currently pursuing the M.Sc. degree in electrical engineering with the University of Stuttgart, Stuttgart. He has worked as a part-time Research Assistant with the University of Stuttgart, from 2019 to 2021, and working as a Student at Robert Bosch GmbH, Stuttgart, Germany, from 2020 to 2021. His bachelor's thesis was supported by TÜBİTAK (Turkish Scientific and Technological Research Council) under the 2209-B Program. His current research interests include the Internet of Things, forecasting, artificial intelligence, and industrial automation systems. His M.Sc. thesis focuses on communication protocols for smart autonomous vehicles.



VOLKAN RODOPLU (Member, IEEE) received the B.S. degree (*summa cum laude*) in electrical engineering from Princeton University, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, in 1998 and 2003, respectively. He has worked with the Wireless Research Division of Texas Instruments, Dallas, TX, USA, and Tensilica Inc., Santa Clara, CA, USA. He worked as an Assistant Professor in electrical engineering with the University of

California at Santa Barbara, from 2003 to 2009, where he was promoted to the position of an Associate Professor with tenure. He is currently a Professor in electrical engineering at Yaşar University, İzmir, Turkey, and a Marie Skłodowska-Curie Fellow of the European Commission. His current research interests include the Internet of Things, predictive networks, smart cities, and visible light communication. He is the winner of the TUBITAK (Scientific and Technological Research Council of Turkey) Achievement Award, the Research Achievement Award at Yaşar University, the National Science Foundation CAREER Award, USA, the University of California Regents' Junior Faculty Fellowship, the Stanford Department of Electrical Engineering Outstanding Service Award, the Stanford Graduate Fellowship (as an Andreas Bechtolsheim Fellow), the Stanford Department of Electrical Engineering Fellowship, the John W. Tukey Award from the American Statistical Association, the G. David Forney Award, and the George B. Wood Legacy Prize.



CÜNEYT GÜZELİŞ received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Istanbul Technical University, Istanbul, Turkey, in 1981, 1984, and 1988, respectively. He was a Visiting Researcher and a Lecturer with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA, USA, between 1989 and 1991. He worked as a full-time Faculty Member at Istanbul Technical University, from 1991 to 2000,

where he became a Full Professor, in 1998. He was a Professor in electrical and electronics engineering at Dokuz Eylül University, İzmir, Turkey, from 2000 to 2011, where he has worked as the Dean of the Faculty of Engineering, and the İzmir University of Economics, İzmir, from 2011 to 2015, where he has worked as the Director of the Graduate School of Natural and Applied Science. Since 2015, he has been a Professor in electrical and electronics engineering at Yaşar University, İzmir, where he has worked as the Director of the Graduate School. He has supervised 17 M.S. and 14 Ph.D. students and published over 50 SCI indexed journal articles, six peer-reviewed book chapters, and more than 80 peer-reviewed conference papers. He has participated in over 20 scientific research projects funded by national and international institutions, such as the British Council and the French National Council for Scientific Research. His research interests include artificial neural networks, biomedical signal and image processing, nonlinear circuits-systems and control, and educational systems.

...