

Article

Non-Intrusive Load Monitoring Based on Novel Transient Signal in Household Appliances with Low Sampling Rate

Thi-Thu-Huong Le [†]  and Howon Kim ^{*,†}

School of Computer Science and Engineering, Pusan National University, Busan 609-735, Korea; lehuong7885@gmail.com

* Correspondence: howonkim@pusan.ac.kr

† Current address: Information Security & IoT Lab, Building A06, School of Computer Science & Engineering, Pusan National University, San-30, JangJeon-dong, Geumjeong-gu, Busan 609-735, Korea.

Received: 31 October 2018; Accepted: 3 December 2018; Published: 5 December 2018

Abstract: Nowadays climate change problems have been more and more concerns and urgent in the real world. Especially, the energy power consumption monitoring is a considerate trend having positive effects in decreasing affecting climate change. Non-Intrusive Load Monitoring (NILM) is the best economic solution to solve the electrical consumption monitoring issue. NILM captures the electrical signals from the aggregate energy consumption, feature extraction from these signals and then learning and predicting the switch ON/OFF of appliances used these feature extracted. This paper proposed a NILM framework including data acquisition, data feature extraction, and classification model. The main contribution is to develop a new transient signal in a different aspect. The proposed transient signal is extracted from the active power signal in the low-frequency sampling rate. This transient signal is used to detect the event of household appliances. In household appliances event detection, we applied to Decision Tree and Long Short-Time Memory (LSTM) models. The average accuracies of these models achieved 92.64% and 96.85%, respectively. The computational and result experiments present the solution effectiveness for the accurate transient signal extraction in the electrical input signals.

Keywords: NILM; energy disaggregation; MCP39F511; Jetson TX2; transient signature; decision tree; LSTM

1. Introduction

Developing countries with rapid urbanization in high buildings construction and the high power demand are a reason for the need for conservation and efficient energy program. The program requires monitoring of customer appliances energy consumption in real-time. Using smart meter had led to NILM enables estimation of individual power consumption used for aggregate power consumption in energy management recently.

In the factory field, researchers are working on Factory Energy Management System (FEMS) for efficient electric energy use. In recently, FEMS has been linked with to the Cyber-Physical Systems (CPS) of Industry 4.0. And the related researches will be more significant in this area [1]. At resident homes, NILM provides the households understand their consumption usage via a cost-effective real-time monitoring appliances system. The customers need giving up unwanted activities to avoid producing unnecessary energy consumption such as the appropriate appliances usage time and appliances usage optimization. These activities can be obtained showing to customers the consumption of each appliance in the sum of the total billing to detect excesses or malfunction [2]. Furthermore, it could be possible notify users of possible savings in their billing electricity. In contrary to this,

Kelly et al. [3] have argued that it is not proven yet that above activity and the additional feedback which become saving energy. In addition, there is an increase in micro-grids and renewable energy facility installation also continuous growth in recently. In purpose of increasing efforts these saving, automating energy measurement, energy monitoring, and the power management system are needed. In the load measurement of a power system, load monitoring has the main role that the process of acquiring and identifying the load [4]. This load monitoring will determine the status of appliances and their consumption. Besides, it supports to understand the behavior of each load in the whole power system. There are two types of the load monitoring including Intrusive Load Monitoring (ILM) and Non-Intrusive Monitoring (NILM).

In ILM, the term “intrusive” means that there is the meter device in the resident house and close to the appliances to monitor. ILM deploys a measurement of the energy consumption of one or more household appliances using meter devices. In the ILM ecosystem, more low-end meter devices are needed. This makes hard to install, maintain, expand as well as expensive. In contrary to ILM, the term “non-intrusive” in NILM means that no extra equipment is installed in the house. NILM is a process which gives data from whole house energy consumption. This process includes installing a sensor device at the panel level and then the appliances will be inferred with being used. NILM preferred using than ILM because it is cheaper and easier installation. Instead of at least one-meter device per room, this technique requires only one-meter device for each energy entrance to the house. Energy disaggregation is another synonym for NILM. This technique estimates the power demand for each appliance from a single meter which contains the overall demand for several appliances.

A NILM system has three roles including capturing the signals from the aggregate consumption, extracting the feature uniquely from the load signal and classifying to identify which appliances are turned ON based on these features extracted. To identify the individual signature of each device, the NILM system requires several steps such as signal sampling (data acquisition), feature extraction based on signal analysis from the electrical signals. This paper presents a NILM framework including data acquisition, data feature extraction, and classification model. The contributions in this NILM study are: (1) propose a NILM framework, (2) collect the household appliances energy consumption data set in low-frequency sampling rate, (3) propose the algorithm to extract a new transient signal in low-frequency sampling rate, (4) improve the performance of NILM model in event detection as well as load identification.

The main contribution to be presented in this paper is a new approach to extract the transient uniquely. The proposed transient signal is extracted from the active power signal in the low-frequency sampling rate. This transient signal is used to detect the event of household appliances. Besides, an embedded board, Jetson TX2, is used to build the proposed NILM application system. This board has two roles to use in this research. The first role is to connect with the sensor to request and storage energy data (data acquisition). The second one is to build machine learning and deep learning models with high-performance training and testing. It is integrated Graphic Processing Unit (GPU) related to deep learning and its performance such as LSTM model. Furthermore, NILM Web application can be built for visualizing NILM result at this board. On the other hand, it can be replaced Jetson TX2 by other hardware that integrated GPU or normal personal computer (PC) without accelerated computing. Hence, the roles of Jetson TX2 is the same role with PC that integrated GPU to perform deep learning model. In summary, Jetson TX2 can connect to the sensor to storage data. Besides, it can build the machine learning, deep learning models, training, predicting and visualizing the result Jetson board or the PC has GPU with support high computation in a deep neural networks.

This paper is organized as follows. The next Section introduces the NILM system concept and presents a related literature works. In Section 3, a NILM framework is proposed. The core of this NILM framework is to propose the algorithm to develop a new transient signal from power signal in low-frequency sampling rate. The experimental setup is described and experimental results are presented in Section 4. Finally, conclusions and future works are drawn in Section 5.

2. Related Works

The initial NILM approach was proposed by Hart in 1990s [5]. The author introduced the Non-Intrusive Appliance Load Monitor (NALM) is a software which was able to analyze single point electrical data and then obtained energy used correspond to individual electrical appliances. Figure 1 shows the first concept of general NILM.

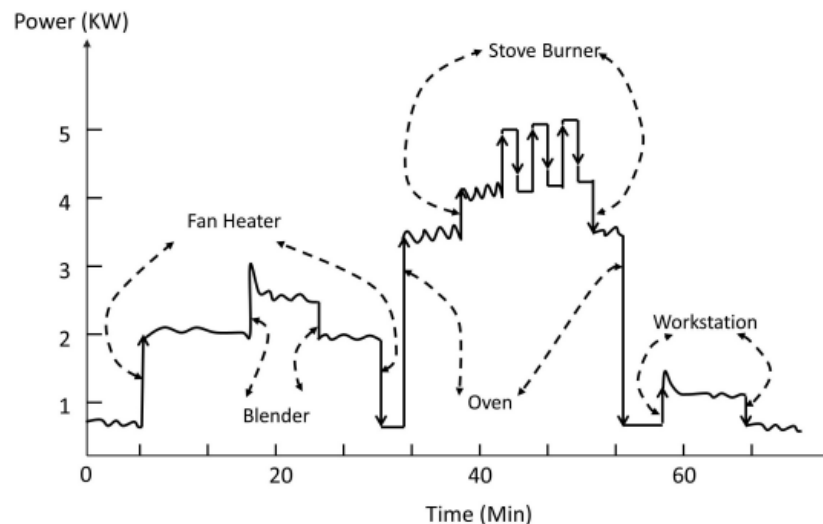


Figure 1. The first concept of NILM based on a single point measurement by Hart [5].

In recently, there are a number of studies have applied and extended this approach by using directly sampled which is increased the resolving power of the $\Delta P - \Delta Q$ space such in [6,7]. This approach has a limitation which cannot distinct appliances has similar in power signal and their operational principles, for example, hairdryer and iron. Hence, Laughman et al. [8] proposed another advanced solution. The capable of this idea based on the transient shapes of appliances to recognize individual load. Besides, they proposed the analysis of the spectral envelopes for continuously variable loads. However, they have not solved the electric noise of appliance usage yet. Therefore, Patel et al. [9] proposed a method to avoid the electric noise via combination software and hardware in household-level current sampling task at 1 MHz. They applied Support Vector Machine (SVM) model to trained the data to achieve accuracy at 90%. However, this technique has a limitation which requires high sampling rate with more kHz. Because of the meter device limitation, it is hard to apply in the real world environment. Furthermore, the need for an adjusting the prediction models on different data on training data is another drawback. The NILM algorithm needs to detect the appliance operation status (ON and OFF) from the power measurements. The NILM approaches can be classified as *event-based* or *state-based* based on different event detection strategies.

Event-based methods generated the state transient edges of appliances. This approach uses a change detection algorithm to determine the start and end of an event such in [10,11]. The significant information needs to extract to identification the event has occurred, for example, appliance signatures such as active power, rising or falling edge etc. These extracted features are analyzed to detect the event based on the appliances and their power consumption estimated. The researcher [12] has used different classification methods such as K-Means, K-Nearest Neighbour (KNN), SVM, Hidden Markov Model (HMM) and HMM's variations to detect event appliance. However, the performance of the event-based approach is not high because of the fixed threshold in the change detection algorithm, the large noise, and the similar among steady state signatures. Furthermore, false positive rate or false alarm rate may arise in the detection of edges methods.

State-based methods do not base on event detection. The idea is how to determine each appliance operation based on a state machine which is a different state transient using usage pattern of

appliances [13]. When the appliance is turned ON/OFF or is changed running states, the method creates new state power signal through a probability distribution to match to the original power signal of the appliance. HMM and its variants [14–17] are used in state-based NILM. However, state-based approaches have several limitations. Firstly, the need for expert knowledge to set a prior value for each appliance by long training periods. Secondly, they have the complexity to compute [13,18]. Finally, there is not a good way to handle states may stable unchanged for long time intervals [19]. The requirement for an effective NILM algorithm is unique features or signatures have to characterize appliance behavior. Appliance signatures are a unique energy consumption pattern of all appliances. Ahmed Z., et al. [17] used appliance signatures to uniquely identify and recognized appliance operations from the aggregated load measurements. In feature selection, two main appliance features are used by NILM research to identify loads including *steady-state* and *transient state* [20].

Steady state is extracted when an appliance changes its running states related to sustained changes in power characteristics. The factors are used in this method including active power, reactive power, current, and voltage. The steady state signatures extraction of current and voltage do not demand high-end meter devices. Features at low frequency are used in the most commonly in steady state features in advance researches. However, the performance of this approach is limited by similarities among steady state signatures. In recently, deep neural network (DNN) in deep learning field becomes more attractive and widely system recognition applied in several areas such as handwritten recognition, speed recognition etc. Specifically, LSTM model is a kind of DNN model which is applied to classification applications have time series data. In NILM, this model is applied to load identification on the UK Domestic Appliance-Level Electricity (UK-DALE) dataset using active power by Kelly et al. [21]. However, the author pointed out the performance limitation on the appliances have informative events in power signal can be many time steps such as washing machine. Besides, Kim et al. [22] applied LSTM model on the variant power signals are generated from active power in low frequency on several public datasets. This method overcomes the long gaps between event may present a challenge in LSTM. In future NILM works, LSTM model may become more and more promising and effecting learning method for researchers.

Contrary to the *steady state* features, before settling into a steady state value, *transient state* features are short-term fluctuations in current or power. To create transient feature uniquely, the authors [17] defined appliance state transients which are shapes, size, duration, and harmonics by sampling current and voltage waveforms at high frequency. Hence, these transient features can achieve signal uniquely to a high degree. Besides, they capture all operation cycles in high sampling rates in longer monitoring time [23]. For example, Patel et al. [9] proposed a custom hardware built to detect the transient noise in range 0.001 kHz to 100 kHz frequency. Then, the authors used the fact that each appliance in-state operation transmits noise back to the power line. However, the high sampling rate required is the major drawback to obtain transient features such as current spikes, transient response time, repeatable transient power profiles, spectral envelopes, etc. [24]. When using a high sampling rate, the system demands a costly hardware and complicated to be installed in the home to detect these transient features [24]. The cause is smart meters report only low-frequency power. In the transient state analysis, the researchers [25,26] analyzed and captured the load signatures based on wavelet transform and transient energy algorithm. Artificial Neural Network (ANN) and HMM are used to improve the performance of NILM in these researches, respectively. Although the results were very much significant, there are still the little drawbacks. For example, the authors [25] sampled at high frequency for current and voltage waveform data to capture the transient effect. Nevertheless, doing this will increase the cost of energy meters because modern energy meters are not equipped with such functionality. Besides, the authors [26] used the data which have the repeatable transient energy signature for load identification because of the varying transient with these waveforms. Therefore, the sampling of the instantaneous load profile for each turn ON transient is required much diligence. In addition, the authors [27] applied the Wavelet Transform Coefficient (WTC) based on ON/OFF transient signal

identification in data acquisition. Although WTC works better than Fourier Transform, WTC requires much longer computation time. Besides, it needs larger machine resource like memory usage.

In summary, the low-frequency or high-frequency data collected is used in classifying appliance recognition systems based on signature feature extraction. A low-frequency data sampling rate is implemented in without additional installation by using existing meter infrastructure. Contrary to this, high-frequency data sampling rate needs adding more hardware installing in data acquisition. Especially, the limitations of high sampling rate data acquisition are more expensive cost and more complex in signature database management [28,29]. However, a more accurate and precise analysis can be provided with more information and assistance [30]. Hence, using a low-frequency sampling rate is more promising in event detection appliances based on analysis the active power and/or reactive power. A new necessary technique of NILM has valid three most important factors to introduce into services for end-users. First thing is an ability to widely appropriately applied in household. The second thing is the usage of the low-cost device to retrieve the energy consumption. The third thing is able to recognition appliances with the same power signal and appliances with variable power signals in the low-frequency sampling data. Hence, in this work, new transient signal is extracted from the active power signal in the low-frequency sampling rate to overcome the problem in advance studies.

3. The Proposed NILM Framework

The proposed NILM framework is introduced in this Section. Figure 2 shows the proposed NILM system. The proposed NILM system includes several components such as energy stream data, sensing device MCP39F511 (Microchip Technology Inc., Chandler, AZ, USA), an embedded board Jetson TX2 (NVIDIA Corporation, Santa Clara, CA, USA), transient signal extraction and recognition, and energy monitoring and consumption prediction system.

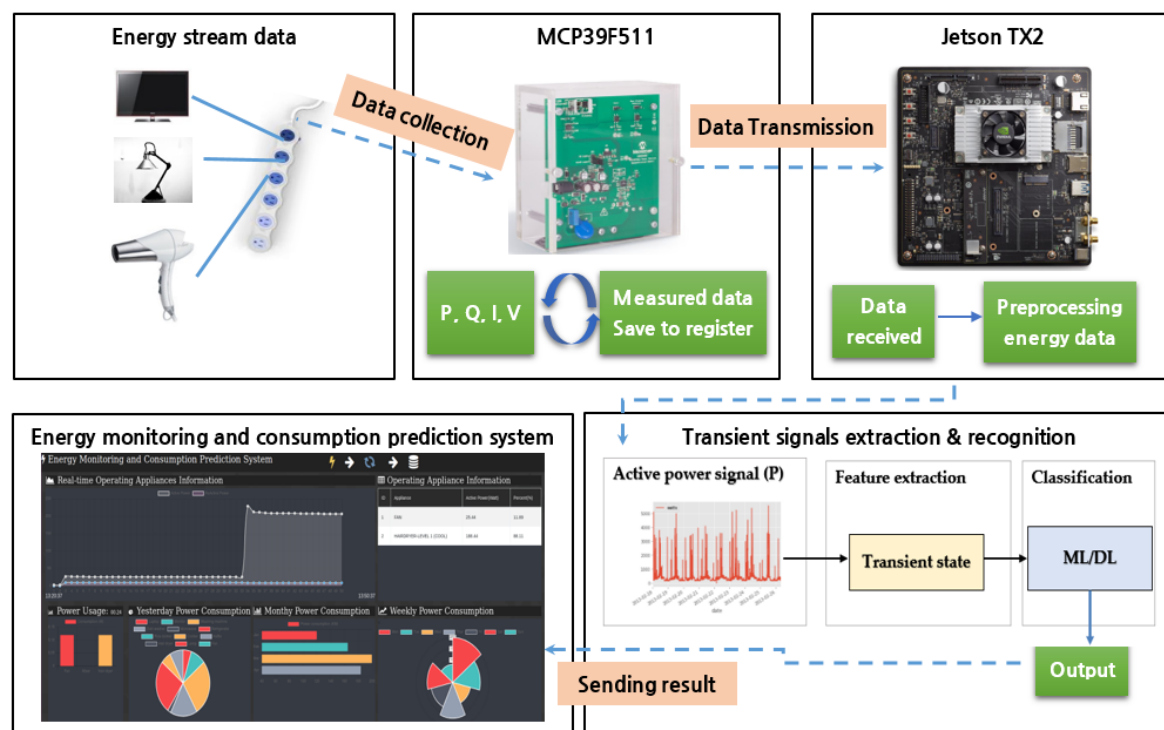


Figure 2. The proposed NILM framework.

3.1. Energy Stream Data

In data acquisition, several household appliances and the multi-tab power (HJ04009-10010C, KC Electric Company, Seoul, Korea) to collect energy data are prepared. Five appliances are collected

including Air-purifier (LG Electronics Inc., Seoul, Korea), Fan (Cixi Xiongsheng Electrical Appliances Co., Ltd., Ningbo, China), Hairdryer (Korea Hanil Electronics, Seoul, Korea), LG monitor (LG Electronics Inc., Seoul, Korea), and Samsung monitor (Samsung Electronics Inc., Suwon, Korea). These appliances are connected to sensing device MCP39F511 via a multi-tab and the sensor's port. Table 1 describes models and power consumption of five appliances which are used to collect energy data.

Table 1. Specifications of five appliances.

Appliance	Model	Power Consumption (W)
Air-purifier	LG AS181DRWT	40
Fan	TESS-S1060	40
Hairdryer	Patech PH-3050	1300
LG monitor	24MP57VQ	23
Samsung monitor	S27D850T	90

Besides, twenty test cases data are also collected. The purpose of testing data collection is to obtain variant transient signals of five appliances. Each test case corresponds to set up that require the number of appliances is different. To each appliance, four combinations between this appliance and 1, 2, 3, or 4 other appliances are create in this work. Therefore, each appliance has four different transient signals in this collecting process. The process to collect testing data includes two steps as follows. The first step is to turn ON simultaneously all appliances required for each test. The second step is to make one appliance transient by turn OFF and then turn ON while other appliances are still ON. Table 2 presents the process to collect twenty test cases in detail.

Table 2. Twenty tests data information.

No	Test	Description
1	Test 1	[Air-purifier, LG monitor, Samsung monitor, Fan, Hairdryer] are ON simultaneously. Turn OFF and then turn ON [Air-purifier] while the other appliances are still ON.
2	Test 2	[Air-purifier, LG monitor, Samsung monitor, and Fan] are ON simultaneously. Turn OFF and then turn ON [Air-purifier] while the other appliances are still ON.
3	Test 3	[Air-purifier, LG monitor, Samsung monitor] are ON simultaneously. Turn OFF and then turn ON [Air-purifier] while the other appliances are still ON.
4	Test 4	[Air-purifier and LG monitor] are ON simultaneously. Turn OFF and then turn ON [Air-purifier] while the other appliance are still ON.
5	Test 5	[Fan and Air-purifier] are ON simultaneously. Turn OFF and then ON [Fan] while the other appliance are still ON.
6	Test 6	[Fan, Samsung monitor, and Air-purifier] are ON simultaneously. Turn OFF and then ON [Fan] while the other appliance are still ON.
7	Test 7	[Fan, Samsung monitor, LG monitor, Air-purifier, and Hairdryer] are ON simultaneously. Turn OFF and then ON [Fan] while the other appliance are still ON.
8	Test 8	[Fan, Samsung monitor, LG monitor, and Air-purifier] are ON simultaneously. Turn OFF and then ON [Fan] while the other appliance are still ON.
9	Test 9	[Hairdryer, Air-purifier, LG monitor, Fan, and Samsung monitor] are ON simultaneously. Turn OFF and turn ON [Hairdryer] while the other appliance are still ON.
10	Test 10	[Hairdryer, Air-purifier, LG monitor, and Fan] are ON simultaneously. Turn OFF and then turn ON [Hairdryer] while the other appliance are still ON.
11	Test 11	[Hairdryer, Air-purifier, and LG monitor] are ON simultaneously. Turn OFF and then tun ON [Hairdryer] while the other appliance are still ON.
12	Test 12	[Hairdryer and Air-purifier] are ON simultaneously. Turn OFF and then turn ON [Hairdryer] while the other appliance are still ON.
13	Test 13	[LG monitor, Samsung monitor, Fan, Air-purifier, and Hairdryer] are ON simultaneously. Turn OFF and then turn ON [LG monitor] while the other appliance are still ON.
14	Test 14	[LG monitor, Samsung monitor, Fan, and Air-purifier] are ON simultaneously. Turn OFF and then turn ON [LG monitor] while the other appliance are still ON.
15	Test 15	[LG monitor, Samsung monitor, and Fan] are ON simultaneously. Turn OFF and then turn ON [LG monitor] while the other appliance are still ON.
16	Test 16	[LG monitor and Samsung monitor] are ON simultaneously. Turn OFF and then turn ON [LG monitor] while the other appliance are still ON.

Table 2. Cont.

No	Test	Description
17	Test 17	[Hairdryer, Fan, Samsung monitor, LG monitor, and Air-purifier] are ON simultaneously. Turn OFF and then turn ON [Samsung monitor] while the other appliance are still ON.
18	Test 18	[Samsung monitor, Fan, LG monitor, and Air-purifier] are ON simultaneously. Turn OFF and then turn ON [Samsung monitor] while the other appliance are still ON.
19	Test 19	[Samsung monitor, LG monitor, and Air-purifier] are ON simultaneously. Turn OFF and then turn ON [Samsung monitor] while the other appliance are still ON.
20	Test 20	[Samsung monitor and Air-purifier] are ON simultaneously. Turn OFF and then turn ON [Samsung monitor] while the other appliance are still ON.

3.2. MCP39F511

The MCP39F511 is a power monitoring device that can measure input power in real time for the consumer, power distribution units, AC/DC power supplies. This sensor supports 2-wire serial protocols and Universal Asynchronous Receiver/Transmitter (UART) with enabling select full speed at up to 115.2 kbps. This sensor has a Power Monitor Demonstration Board which is a fully functional single-phase power. The system calculates and displays active power, reactive power, RMS current, RMS voltage, active energy (both import and export), and four quadrants reactive energy. MCP39F511 changes data acquisition mode compare to a conventional method. In the conventional method, data acquisition mode is getting energy data stored in registers by sending a command from PC. In this sensor device, the mode is getting energy data via connecting to Jetson TX2 (see Figure 3) via some steps as follows.

- Step 1: Jetson TX2 sends a command to switch to single wire mode. This single wire mode includes twenty bytes such as Header Byte (0 × AB), Header Byte 2 (0 × CD), Header Byte 3 (0 × EF), Current RMSs with Byte 0 to Byte 3, Voltage RMSs with Byte 0 to Byte 1, Active Power with Byte 0 to Byte 3, Reactive power with Byte 0 to Byte 3, Line Frequency with Byte 0 to Byte 1, and final is check sum.
- Step 2: Single wire mode is automatically sent whenever the sensing device updates energy data.
- Step 3: The sampling rate is in 15 Hz (see Figure 4).

There is a coherent sampling algorithm to phase lock the sampling rate to the line frequency based on an integer number of sample per line cycle in the computation cycle of MCP39F511. After that, it reports all power output quantities at a 2^N number of line cycles. The power outputs include RMS current, RMS voltage, apparent power, active power. The accumulation interval is defined as an 2^N number of line cycles, where N is the value in the Accumulation Interval Parameter register. Equations (1)–(5) system calculate and display RMS current (I_{RMS}), RMS voltage (V_{RMS}), Apparent power (S), Active power (P), Reactive power (Q), and Power factor as follows.

RMS current equation (I_{RMS}) with unit is Amps (A):

$$I_{RMS} = \sqrt{\frac{\sum_{n=0}^{2^N-1} (i_n)^2}{2^N}} \quad (1)$$

RMS voltage equation (V_{RMS}) with unit is Volt (V):

$$V_{RMS} = \sqrt{\frac{\sum_{n=0}^{2^N-1} (v_n)^2}{2^N}} \quad (2)$$

Apparent power equation (S) with unit is Volt-Amps (VA):

$$S = I_{RMS} \times V_{RMS} \quad (3)$$

Active power equation (P) with unit is Watts (W):

$$P = \frac{1}{2^N} \sum_{k=0}^{2^N-1} V_k \times I_k \quad (4)$$

In the MCP39F511, Reactive power (Q) with Volt-Amps-Reactive unit (VAR) is measured based on a θ -degree phase shift in the voltage channel. The common degree is 90-degree phase shift. Accumulator Unit (ACCU) acts as the accumulator where has the similar accumulation principles applied to Active power (P). In the Gain Reactive power register, Gain is corrected. In the Reactive power register, an unsigned 32-bit value is located which is the final output. P is measured by the formula below.

$$Q = V_{RMS} \times I_{RMS} \times \sin(\theta) \quad (5)$$

The ratio of P to S or Active power divided by Apparent power is Power factor (PF) measurement.

$$PF = \frac{P}{S} \quad (6)$$

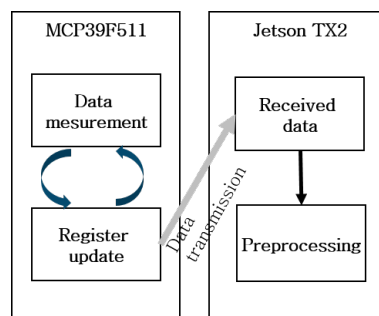


Figure 3. Data transfer process.

A	B	C	D	E	F	G	H	I	J	K	L
Year	Month	Day	Hour	Minute	Second	Millisecond	Current	Voltage	ActivePwr	ReactivPwr	Frequency
2018	3	5	9	21	20	50928	0.0835	222.9	3.82	16.33	60.048
2018	3	5	9	21	20	117481	0.0819	222.9	3.57	16.38	60.074
2018	3	5	9	21	20	184210	0.0827	222.7	3.56	16.43	60.124
2018	3	5	9	21	20	250923	0.0827	222.7	3.78	16.32	60.098
2018	3	5	9	21	20	317605	0.0835	222.8	3.87	16.28	60.098
2018	3	5	9	21	20	384271	0.0827	222.7	3.69	16.32	60.098
2018	3	5	9	21	20	450959	0.0827	222.7	3.69	16.32	60.098
2018	3	5	9	21	20	517666	0.0843	222.8	3.94	16.31	60.098
2018	3	5	9	21	20	584307	0.0835	222.9	3.85	16.31	60.098
2018	3	5	9	21	20	650940	0.0835	222.9	3.81	16.36	60.098
2018	3	5	9	21	20	717508	0.0835	222.8	3.65	16.46	60.098
2018	3	5	9	21	20	784372	0.0827	222.9	3.69	16.4	60.048
2018	3	5	9	21	20	850993	0.0835	222.9	3.73	16.36	60.048
2018	3	5	9	21	20	917845	0.0835	222.8	3.68	16.41	60.048
2018	3	5	9	21	20	984467	0.0835	222.9	3.73	16.4	60.048
2018	3	5	9	21	21	50976	0.0835	222.9	3.77	16.38	60.048
2018	3	5	9	21	21	117763	0.0843	223	3.82	16.35	60.074
2018	3	5	9	21	21	184411	0.0933	223	5.62	16.06	60.098
2018	3	5	9	21	21	251133	0.1014	223	7.36	15.75	60.098
2018	3	5	9	21	21	317812	0.1014	223	7.42	15.8	60.098
2018	3	5	9	21	21	384489	0.099	223	7.1	15.84	60.074
2018	3	5	9	21	21	451071	0.099	222.9	7.1	15.82	60.048
2018	3	5	9	21	21	517752	0.1249	222.6	11.33	15.27	60.074
2018	3	5	9	21	21	584576	0.1371	222.6	13.33	14.98	60.098
2018	3	5	9	21	21	651324	0.1379	222.8	13.53	14.98	60.048
2018	3	5	9	21	21	718006	0.1371	222.9	13.36	14.98	60.048
2018	3	5	9	21	21	784595	0.1371	222.8	13.38	15.01	60.074
2018	3	5	9	21	21	851286	0.1387	222.8	13.66	14.96	60.098
2018	3	5	9	21	21	917898	0.1379	222.8	13.49	14.97	60.098
2018	3	5	9	21	21	984733	0.1395	222.8	13.78	14.95	60.048

Figure 4. The low sampling rate format data collected in 15 Hz of Air-purifier device.

3.3. Jetson TX2

Jetson is the low-power embedded platform in the world's leading. Besides, it enables server-class AI to compute performance everywhere. Jetson's features include an integrated 256-core NVIDIA Pascal GPU, a hex-core ARMv8 64-bit CPU complex, and 8GB of LPDDR4 memory with a 128-bit interface. Figure 5 shows the CPU complex which combines a dual-core NVIDIA Denver 2 alongside a quad-core ARM Cortex-A57.

Table 3 shows Jetson TX2 technical specifications in detail. The installation files of Jetson TX2 are set up including,

- GPU includes Cuda, cudnn
- Machine learning/ deep learning with Tensorflow 1.3
- Python 3rd party lib with Pandas, numpy, jupyter, pyserial, matplotlib etc.
- Power Sensor Interlock is CDC ACM module
- Server includes Node.js 6.11.3, Npm 3.10.10, MongoDB-enterprise



Figure 5. Jetson TX2.

Table 3. Jetson TX2 Technical Specifications.

Feature	Byte
CPU	NVIDIA Pascal TM , 256 CUDA cores
GPU	HMP Dual Denver 2/2 MB L2 + Quad ARM A57/2 MB L2
Monitor	8 GB 128 bit LPDDR4 59.7 GB/s
Data capacity	32 GB eMMC, SDIO, SATA
Connectivity	1 Gigabit Ethernet, 802.11ac WLAN, Bluetooth
Etc.	CAN, UART, SPI, I2C, I2S, GPIOs

It is a necessary to configure the system for collecting energy data from the appliances. The stream energy data is calculated in the power sensing device (MCP39F511). When the embedded board (Jetson TX2) sends a power data request, the sensing device will transfer power data via a single wire mode transmission frame 15 times per second. After that, Jetson TX2's processor operates to sort data received from MCP39F511 and save the data result to database management (MongoDB). For instance, Figure 6 shows the system configuration diagram to collect data from five household appliances (Air-purifier, Fan, Hairdryer, LG monitor, and Samsung monitor).

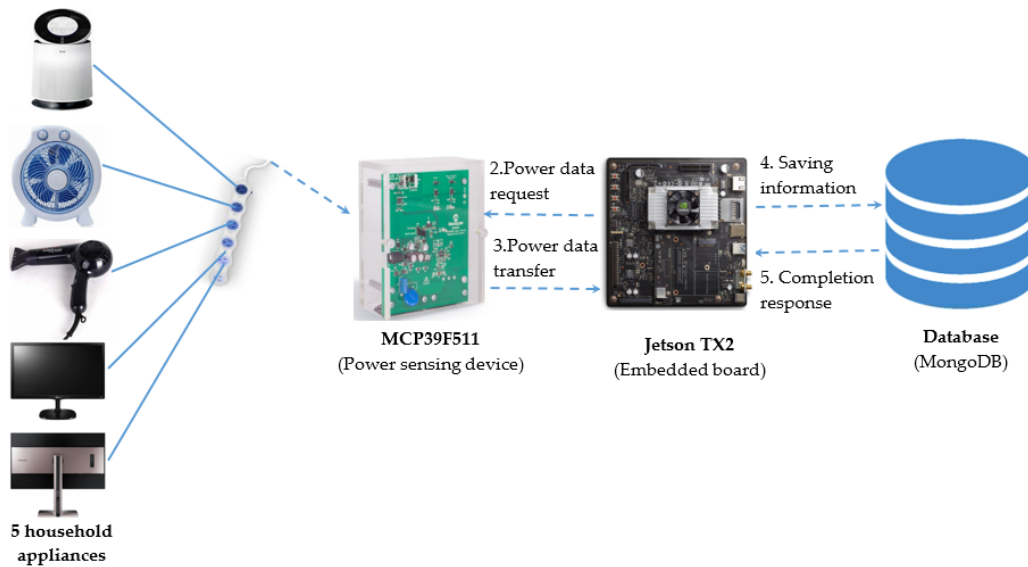


Figure 6. System configuration diagram collection data.

To collecting and storing the energy data, this paper proposes Algorithm 1. The algorithm consists of four steps as follows.

- Step 1. Setting up serial connect from NVIDIA Jetson TX2 to MCP39F511
- Step 2. Checking frame with 20-byte frame
- Step 3. Calculating power data including current, voltage, active power, reactive power, frequency
- Step 4. Storing data into NVIDIA Jetson TX2

Algorithm 1 The requesting and storing energy data of Jetson TX2

Require: *single_wire_frame*

Ensure: current (I), voltage (V), activepwr (P), reactivepwr (Q), frequency (F)

```

1: if Serial.isConnected () then
2:   Serial.write (single_wire_frame)
3: else
4:   print ("serial connection error!")
5:   exit()
6: end if
7: while Serial.isConnected() do
8:   single_wire_mode =Serial.read()
9:   check = check_frame (single_wire_frame)
10:  if check = True then
11:     $I, V, P, Q, F = \text{calculate\_power\_data}(\text{single\_wire\_mode})$ 
12:    save_to_database ( $I, V, P, Q, F$ )
13:  end if
14: end while

```

In this algorithm , it defines two functions. The first function is **check_frame()** to check frame with 20-byte frame in single wire mode. The second one is **calculate_power_data()** to measure energy data information in Root Mean Square (RMS).

3.4. Transient Signal Extraction and Recognition

Based on literature observation above, this paper states three problems needing solve as follows.

Problem 1. Analyzing and extracting new transient signal from original power signal in low sampling rate data in feature extraction.

Problem 2. Labeling ON/OFF data on new transient signals extracted.

Problem 3. Improving the performance for event detection as well as load identification using a new transient signal results and machine learning/deep learning models.

This study selects active power factor which is a unique input in this approach. Figure 7 shows in detail of the proposed solution. The Figure 7a displays the process of this solution. The Figure 7b is to illustrate the process in Figure 7a, respectively. The proposed method includes three processes, such as feature extraction, labeling, and classification. The first, the feature extraction task is to generate the state of the appliance and to extract the transient signal of appliances. The second, labeling task is to label ON/OFF data with state and transient signals of appliance after extracting. The final, classification task is to learn and classify the ON/OFF appliance based on the transient signal and ON/OFF label signal.

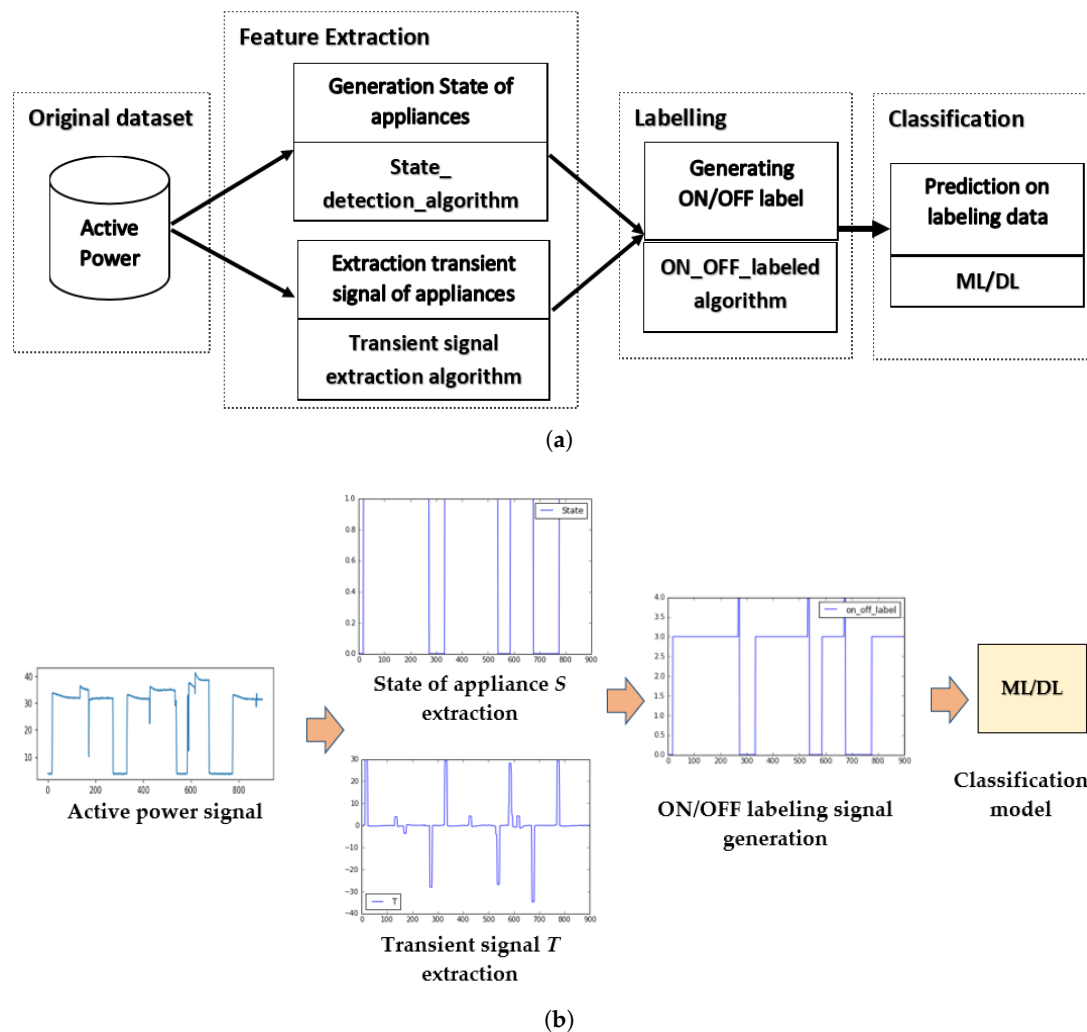


Figure 7. The proposed technique for transient signal extraction and recognition. (a) The proposed concept; (b) The illustration of the proposed process.

In feature extraction task, to label ON/OFF for appliance data, this work generates two signals including State of appliance signal and Transient signal. Hence, this paper proposes two algorithms, Algorithms 2 and 3.

The main idea of this task is how to detect whether or not rising and falling signals from the active power signal. The rising signal means appliance is operating (ON state). The falling signal means appliance is changing state from ON to OFF. Sliding window size of time series data is used to trace power signal to detect changing signals. Hence, the performance of the changing detection algorithm

does not depend on the fixed or adaptive threshold, the large measure noise, and similarities among steady state signatures. Figure 8 shows the changing signals from active power signal of the appliance.

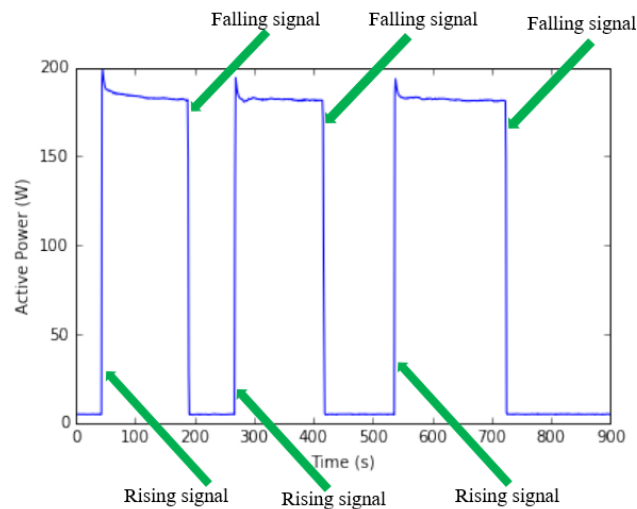


Figure 8. The analysis of active power signal with rising and falling signal.

Algorithm 2 is to generate State of the appliances, denoted by S . S value extracted is used for the input of Algorithm 4.

Algorithm 2 State of the appliances generation algorithm

Require: Activepwr (P)

Ensure: State of appliance (S)

```

1:  $L_p \leftarrow \text{length of } P$ 
2:  $\text{bin} \leftarrow \text{histogram}(P)$ 
3:  $\text{threshold} \leftarrow \text{bin}$ 
4:  $\text{OFF\_state} \leftarrow 0$ 
5:  $\text{ON\_state} \leftarrow 1$ 
6: for  $i$  in range  $(0, L_p)$  do
7:   if  $P[i] = 0$  then
8:      $S \leftarrow S || 0$ 
9:   else if  $P[i] < \text{threshold}$  then
10:     $S \leftarrow S || \text{OFF\_state}$ 
11:   else
12:     $S \leftarrow S || \text{ON\_state}$ 
13:   end if
14: end for
15: return  $S$ 

```

Algorithm 3 is to extract new transient signal from active power signal, denoted by T . T value is determined based on calculating prior P data (pre_data) and post P data (post_data). To assign pre_data and post_data values, the requirement is to use the window size w of P time series data and time shifting time series data of P is +1 for post_data value. The size of w depends on the sample size of P and periodicity of the data. To get smaller smoothing moving the average of time series data, there is an initialization of w value is 5. Hence, confidence intervals for the smoothed values are get. T signal extracted from this algorithm becomes to the input of Algorithm 4. If T value has negative value it means corresponding to P active power has falling signal, and then setting appliance's label is OFF event. In contrast to this, the active power has a rising signal, and then setting appliance's label is ON event.

Algorithm 3 Transient signal extraction algorithm

Require: Activepwr (P)
Ensure: Transient signal extraction (T)

```

1:  $L_p \leftarrow \text{length of } P$ 
2:  $w \leftarrow 5$ 
3:  $T \leftarrow [0]$ 
4: for  $i$  in range ( $w, L_p - w$ ) do
5:    $T\_sum \leftarrow 0$ 
6:   for  $j$  in range (15) do
7:      $pre\_data \leftarrow P[i - w + j]$ 
8:      $post\_data \leftarrow P[i - w + 1 + j]$ 
9:      $T\_list \leftarrow post\_data - pre\_data$ 
10:     $T\_sum \leftarrow \Sigma(T\_list)$ 
11:   end for
12:    $T \leftarrow T || T\_sum$ 
13: end for
14: return  $T$ 

```

In labeling task, this paper generates ON/OFF labeling by the proposed Algorithm 4. T signal and state of the appliances are extracted by Algorithms 2 and 3, respectively. They become to the input of this algorithm. The task of Algorithm 4 is to generate ON_label or OFF_label from T signal. This study defines *threshold* value is -5 which is a maximum threshold value to determine whether event status changing from ON to OFF event. If T value extracted at time t smaller than -5 it means that at that time has occurred event changing from ON to OFF. In case S variable has 0 value, it means that no event operation. If T value is smaller than *threshold*, it is assigned label is OFF_label ; else it is assigned label is ON_label . OFF_label value or ON_label value of each appliance are different values. For example, there are two appliances including Air-purifier and Fan. This paper defines sets of ON/OFF labels of these appliances as follow. Labels of Air-purifier device are $OFF_label=1$ and $ON_label=2$. And labels of Fan device are $OFF_label=3$ and $ON_label=4$. For setting ON/OFF labeling values of each appliance and twenty-test cases data, this paper mentions the labels for testing data in next Section.

Algorithm 4 Labeling ON/OFF algorithm

Require: Transient signal extracted (T), State of appliance (S)
Ensure: ON_OFF_label

```

1:  $L_T \leftarrow \text{length of } (T)$ 
2:  $ON\_OFF\_label \leftarrow []$ 
3:  $threshold \leftarrow -5$ 
4: for  $i$  in range ( $0, L_T$ ) do
5:   if  $S[i] = 0$  then
6:      $ON\_OFF\_label \leftarrow ON\_OFF\_label || 0$ 
7:   else if  $T[i] < threshold$  then
8:      $ON\_OFF\_label \leftarrow ON\_OFF\_label || OFF\_label$ 
9:   else
10:     $ON\_OFF\_label \leftarrow ON\_OFF\_label || ON\_label$ 
11:   end if
12: end for
13: return  $ON\_OFF\_label$ 

```

In the classification task, this paper evaluates the proposed method using the models in both Machine Learning (ML) and Deep Learning (DL) fields. For learning T signal and ON/OFF labeling

signal, Decision Tree and Long Short-Term Memory (LSTM) models are selected in ML and DL, respectively.

- Decision Tree is a supervised learning method. It is used in both classification and regression tasks. The input feature of this model is used to infer the output feature by learning simple decision rules. CART [31] (Classification and Regression Trees) is similar to C4.5, however, two different to C4.5 are it supports the regression task in numerical target variable and does not need to compute rule sets. This algorithm builds binary trees based on the feature and threshold with the largest information gain at each node. In a decision tree algorithm, it needs training vectors $x_i \in R^n, i = 1, \dots, l$ and a vector $y \in R^l$. This algorithm needs to recursively partition such that grouping the same labels in a group together. A feature f and threshold $thre_f$, partition the data into T_{left} and T_{right} subsets are contained in each candidate split $\theta = (j, thre_f)$

$$T_{left}(\theta) = (x, y) | x_f \leq thre_j \quad (7)$$

$$T_{right}(\theta) = T \setminus T_{left}(\theta) \quad (8)$$

The impurity is the choice of the classification or regression task. The impurity at j is calculated based on an impurity function $H(j)$.

$$G(T, \theta) = \frac{n_{left}}{N_j} H(T_{left}(\theta)) + \frac{n_{right}}{N_j} H(T_{right}(\theta)) \quad (9)$$

Select the parameters that minimizes the impurity is formulated by θ^* as follows.

$$\theta^* = \operatorname{argmin}_{\theta} G(T, \theta) \quad (10)$$

Recurse for subsets $T_{left}(\theta^*)$ and $T_{right}(\theta^*)$ reaching the maximum allowable depth, $N_j < \min_{samples}$ or $N_j = 1$. In the classification task, the output classification represents a region R_j with N_j observations taking on value $0, 1, \dots, K - 1$, for node j .

$$pr_{jk} = \frac{1}{N_j} \sum_{x_i \in R_j} I(y_i = k) \quad (11)$$

be the proportion of class k observations in node j . $I(\cdot)$ is a spitting criterion that makes used of the impurity measure.

Common measures of impurity are Gini index which is calculated as follows.

$$H(X_j) = \sum_k p_{jk}(1 - p_{jk}) \quad (12)$$

- LSTM has been designed by Hochreiter and Schmidhuber in 1997 [32]. LSTM is an elegant Recurrent Neural Network (RNN). The LSTM architecture is combined memory cells by replacing the regular units of the neural network. A memory cell consists of three gates: an *input gate*, a *forget gate*, an *output gate* and the state of memory cell is called *cell state* (Figure 9a). In particular, at the *input gate*, it allows incoming signal to alter the *cell state* or block it. On the other hand, the *output gate* can allow the *cell state* to have an effect on other neurons or prevent it. Finally, the *forget gate* can modulate *cell state* of the memory cell. Besides, it allows the cell to remember or forget its previous state as needed.

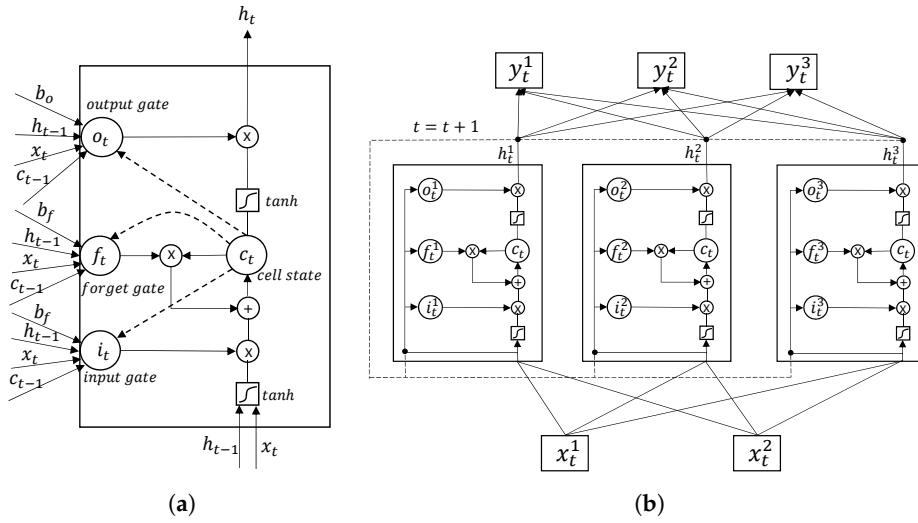


Figure 9. (a) LSTM Cell; (b) An architecture of LSTM model

The difference between LSTM and the original RNN is hidden units. In LSTM, the hidden units are replaced by LSTM cells. Figure 9b shows an architecture of LSTM having two input units, three LSTM cells as hidden units and three output units. The equations below describe how LSTM processes data. Assumption that $X_t = [x_t^1, x_t^2, x_t^3, \dots, x_t^{n_x}]$ is an input vector and $H_t = [h_t^1, h_t^2, h_t^3, \dots, h_t^{n_h}]$, $Y_t = [y_t^1, y_t^2, y_t^3, \dots, y_t^{n_y}]$, $C_t = [c_t^1, c_t^2, c_t^3, \dots, c_t^{n_c}]$ are hidden, output and cell vector, respectively. The elements of each vector are units for layers of LSTM. n_x , n_h , n_c and n_y are a number of each units. σ is the logistic sigmoid function, and i , f and o are the input gate, forget gate and output gate, respectively. The weight matrix superscripts have the obvious meaning. For example, W_t^{hi} is the hidden-input gate weight matrix and W_t^{xo} is the input-output gate weight matrix. b_t^i , b_t^f , b_t^c and b_t^o are bias terms at time t .

First is to compute the value for f_t , the activation of the forget gate. The output range of f_t is from 0 to 1 and the output value will be multiplied by c_{t-1} when calculating c_t . Therefore, f_t means an activation rate of the previous cell state.

$$f_t = \sigma(W_t^{xf} \sum_{i=1}^{n_x} x_t^i + W_t^{hf} \sum_{j=1}^{n_h} h_{t-1}^j + W_t^{cf} \sum_{k=1}^{n_c} c_{t-1}^k + b_f) \quad (13)$$

Second is to compute the value for the input gate i_t . In common with f_t , i_t is the activation ratio of the input value x_t .

$$i_t = \sigma(W_t^{xi} \sum_{i=1}^{n_x} x_t^i + W_t^{hi} \sum_{j=1}^{n_h} h_{t-1}^j + W_t^{ci} \sum_{k=1}^{n_c} c_{t-1}^k + b_i) \quad (14)$$

Third is to compute the value for the cell state c_t . Two factors are combined. The first factor is the previous cell state activated by the forget gate and the second factor is the input value activated by the input gate.

$$c_t = f_t c_{t-1} + i_t \tanh(W_t^{xc} \sum_{i=1}^{n_x} x_t^i + W_t^{hc} \sum_{j=1}^{n_h} h_{t-1}^j + b_c) \quad (15)$$

Final is to compute the value of their output gates and use it for the memory block output.

$$o_t = \sigma(W_t^{xo} \sum_{i=1}^{n_x} x_t^i + W_t^{ho} \sum_{j=1}^{n_h} h_{t-1}^j + W_t^{co} \sum_{k=1}^{n_c} c_{t-1}^k + b_o) \quad (16)$$

$$h_t = o_t \tanh(c_t) \quad (17)$$

Equations (13)–(17) are processed in one LSTM cell. After all process are done in the hidden layer, it can be calculated for the output units with the hidden vector H_t .

$$y_t = \sigma(W_t^{hy} \sum_{i=1}^{n_h} h_t^i) \quad (18)$$

3.5. Energy Monitoring and Consumption Prediction System

This is a web application to display the energy disaggregation results. This paper uses Node.js, Javascript, and HTML to build the user interface for NILM system. The results of classification are passed to energy monitoring system.

4. Experiments

This section points data preparation for experiment and experiment results in this approach. The results including T signals extracted and classification ON/OFF event. This study sets up the environment for implementation as following Intel® Core™ i7-4790 CPU @3.60GHz; GPU: NVIDIA GeForce GTX 750 (NVIDIA Corporation, Santa Clara, CA, USA); RAM: 16GB and Operating System (OS): Windows 10; the language programming in Python.

Besides, this paper uses confusion matrix (CM) to evaluate the approach classification model. CM includes four categories such as True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). This paper assumes that the positive event means an appliance is turned ON and when the appliance is turned OFF that it is the negative event. Hence, Recall, Precision, Accuracy, F1-score are calculated for evaluating the NILM model based on CM.

Recall is a ratio of the number correctly classified to the total number of actual positive samples.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (19)$$

Precision is a ratio of the number of correctly classify to the total number of predicted positive samples.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (20)$$

Accuracy is a ratio of correctly classify to the total test data.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{FP}) + (\text{FN} + \text{TN})} \quad (21)$$

F1-score is the harmonic average of Recall and Precision.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (22)$$

Furthermore, this paper measures loss of NILM model using Loss function such as Mean Squared Error (MSE).

$$\text{Loss} = \frac{1}{n} \sum_{i=1}^n (Y_t - Y_p)^2 \quad (23)$$

where Y_t is the expected output of sample data, Y_p is the predicted output of sample data by NILM approach model.

4.1. Data Preparation for Experiment

Five appliances data individual and twenty test cases are collected and saved in *.csv files. As mentioned in the previous Section, this paper selected active power data for input feature in the experiments. Figure 10 illustrates the active power signals from five active power data of five appliances.

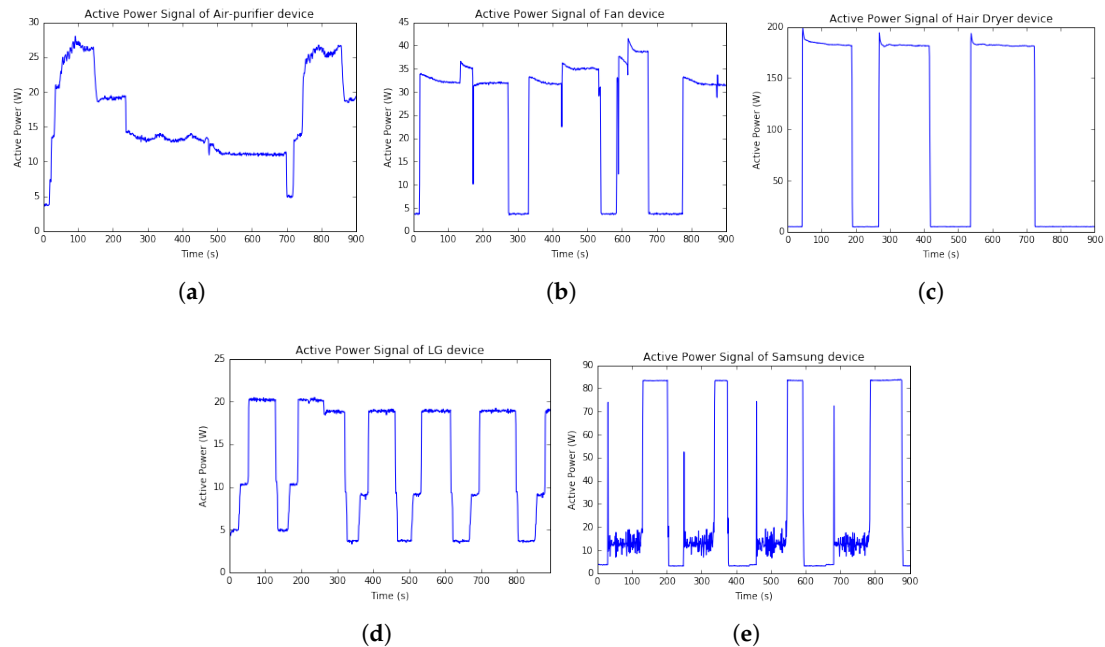


Figure 10. Active power signals: (a) Airpurifier; (b) Fan appliance; (c) Hairdryer appliance; (d) LG monitor appliance; (e) Samsung monitor appliance.

In testing data, this paper named twenty-test cases data files following in Table 4.

Table 4. Twenty-test cases data file name.

No	Test	File Name	Description
1	Test 1	Airpurifier_Transient_LG_Samsung_Fan_Hairdryer_Steady	Air-purifier transient
2	Test 2	Airpurifier_Transient_LG_Samsung_Fan_Steady	
3	Test 3	Airpurifier_Transient_LG_Samsung_Steady	
4	Test 4	Airpurifier_Transient_LG_Steady	
5	Test 5	Fan_Transient_Airpurifier_Steady	Fan transient
6	Test 6	Fan_Transient_Samsung_Airpurifier_Steady	
7	Test 7	Fan_Transient_Samsung_LG_Airpurifier_Hairdryer_Steady	
8	Test 8	Fan_Transient_Samsung_LG_Airpurifier_Steady	
9	Test 9	Hairdryer_Transient_Airpurifier_LG_Fan_Samsung_Steady	Hairdryer transient
10	Test 10	Hairdryer_Transient_Airpurifier_LG_Fan_Steady	
11	Test 11	Hairdryer_Transient_Airpurifier_LG_Steady	
12	Test 12	Hairdryer_Transient_Airpurifier_Steady	
13	Test 13	LG_Transient_Samsung_Fan_Airpurifier_Hairdryer_Steady	LG monitor transient
14	Test 14	LG_Transient_Samsung_Fan_Airpurifier_Steady	
15	Test 15	LG_Transient_Samsung_Fan_Steady	
16	Test 16	LG_Transient_Samsung_Steady	
17	Test 17	Samsung_Transient_Airpurifier_LG_Hairdryer_Fan_Steady	Samsung monitor transient
18	Test 18	Samsung_Transient_Airpurifier_LG_Fan_Steady	
19	Test 19	Samsung_Transient_Airpurifier_LG_Steady	
20	Test 20	Samsung_Transient_Airpurifier_Steady	

Besides, this paper plotted twenty active power signals from twenty test cases in Figure 11.

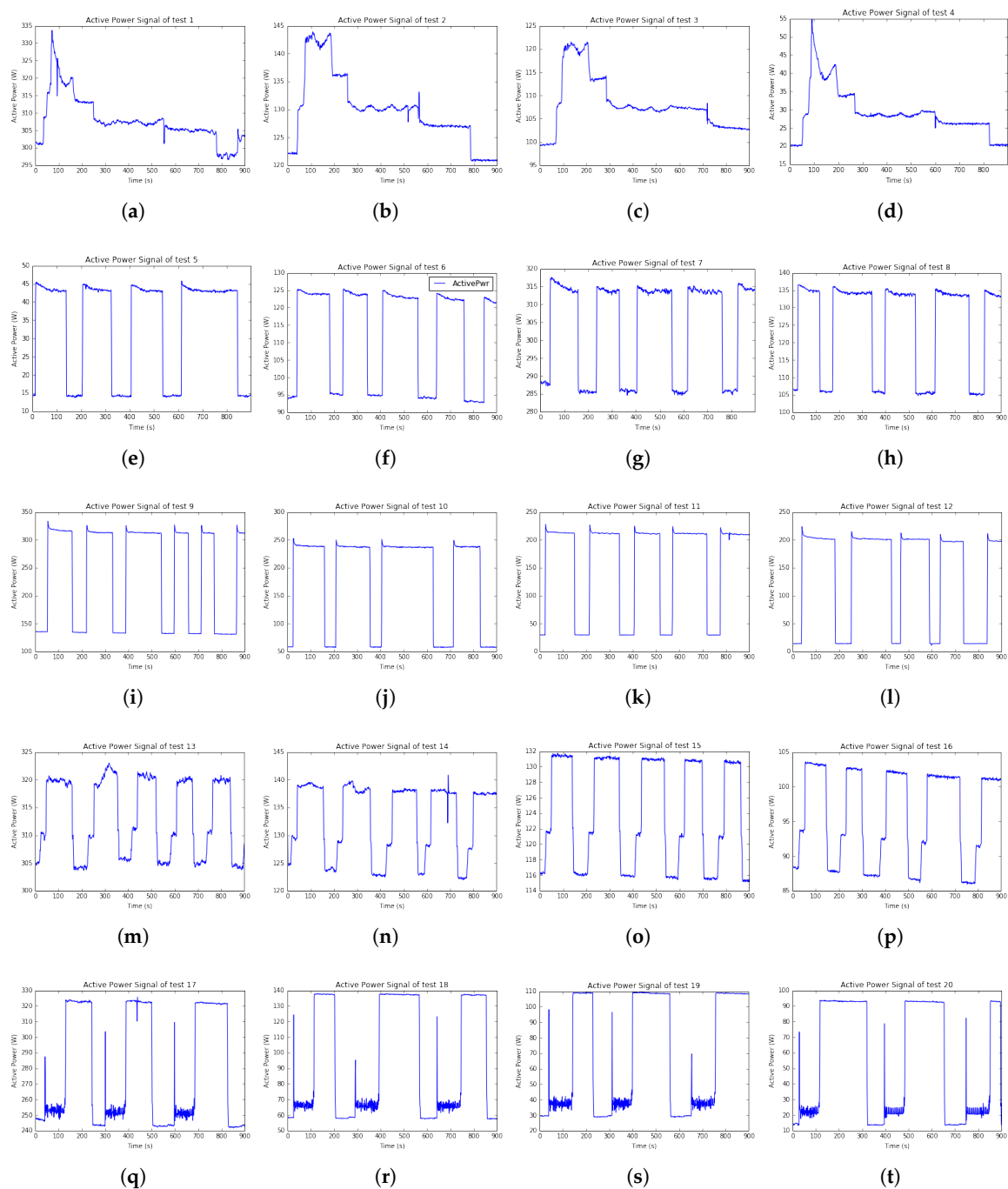


Figure 11. Active Power Signals of twenty test cases data: (a–d): test 1 to test 4; (e–h): test 5 to test 8; (i–l): test 9 to test 12; (m–p): test 13 to test 16; (q–t): test 17 to test 20.

4.2. Experiment Results

4.2.1. State of Appliance, Transient Signal and ON/OFF Label Results

Firstly, this paper determined and declared ON/OFF label value to use in Algorithm 4. Tables 5 and 6 show ON/OFF labels for five appliances and twenty test cases, respectively.

Table 5. Labeling ON/OFF for each appliance.

Appliance Name	ON/OFF Labeling	
	ON Label	OFF Label
Airpurifier	1	2
Fan	3	4
Hairdryer	5	6
LG monitor	7	8
Samsung monitor	9	10

Table 6. Labeling ON/OFF for twenty test cases.

Test Name	ON/OFF Labeling	
	ON Label	OFF Label
Test 1, Test 2, Test 3, Test 4	1	2
Test 5, Test 6, Test 7, Test 8	3	4
Test 9, Test 10, Test 11, Test 12	5	6
Test 13, Test 14, Test 15, Test 16	7	8
Test 17, Test 18, Test 19, Test 20	9	10

Secondly, this paper implemented Algorithms 2 and 3 to extract *State* of the appliance and *T* signals, respectively. After that, the Algorithm 4 are implemented to label ON/OFF based on each *T* signals from five appliances and twenty test cases data. The left, middle, and right of Figure 12 are *State S*, *T* and *ON/OFF_label* signals obtained from five appliances data, respectively. From observing these *T* signal results in this figure, this paper can determine the number of changing status of event ON/OFF in each appliance. There are two, three, three, five, and four times changing status from ON to OFF corresponding to air-purifier, fan, hairdryer, LG monitor, and Samsung monitor.

In summary, this paper recognizes that when *T* signal changes the value from positive to negative value and smaller than -5 , the event ON/OFF on device will occur. In particular, when the *T* signal obtains positive values, it means that the appliance is operating (ON). Contrast to this, the appliance switches to OFF.

For visualizing the results of twenty test cases, this paper pointed the *S*, *T*, and ON_OFF label results of test 1 to test 4 in Figure 13. Besides, the *S*, *T*, and ON_OFF label results of test 5 to test 20 are plotted in Figures A1–A3 at Appendix A, respectively.

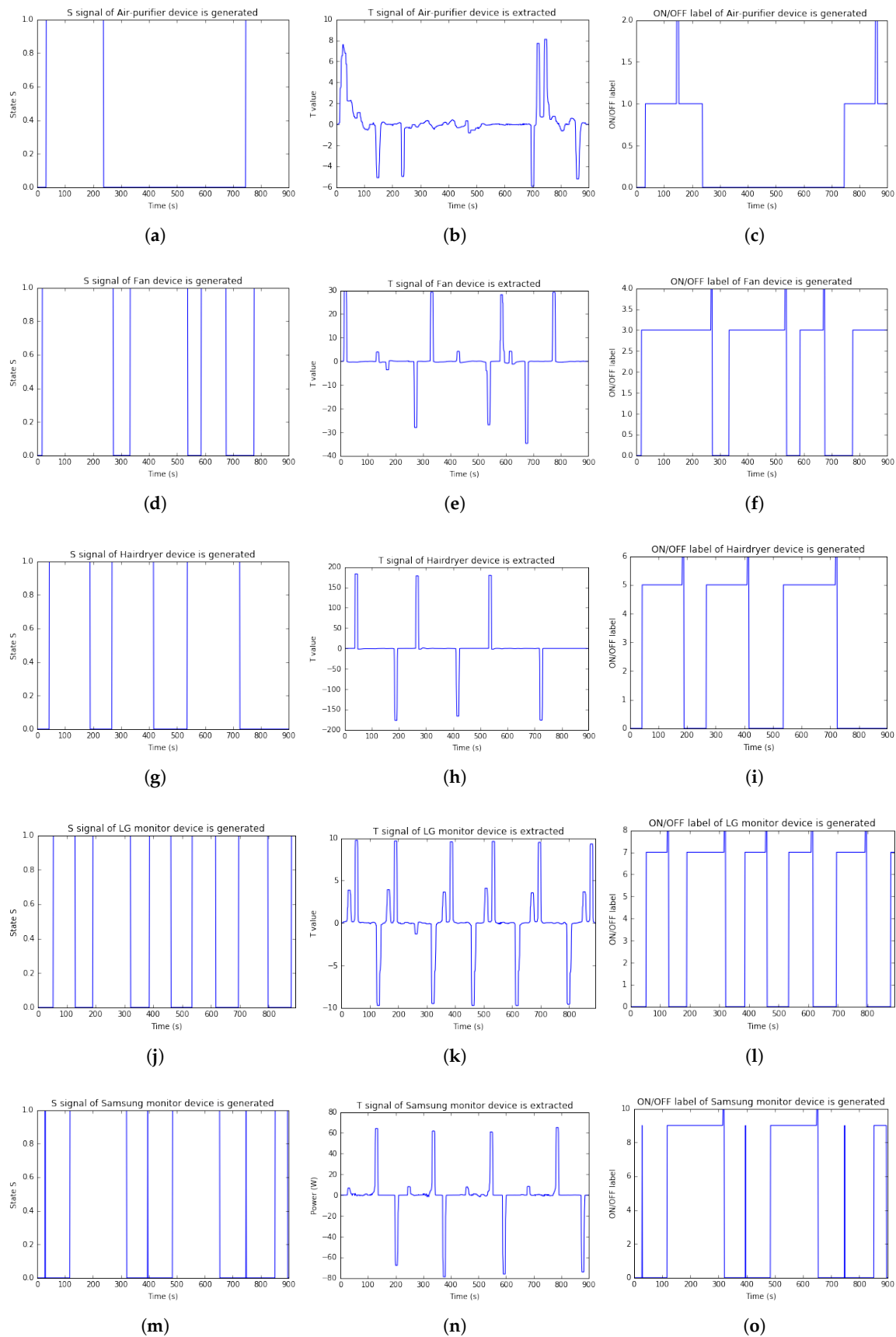


Figure 12. State S , T & ON/OFF label signals of 5 appliances: (a–c): Air-purifier device; (d–f): Fan device; (g–i): Hair Dryer device; (j–l): LG monitor device; (m–o): Samsung monitor device.

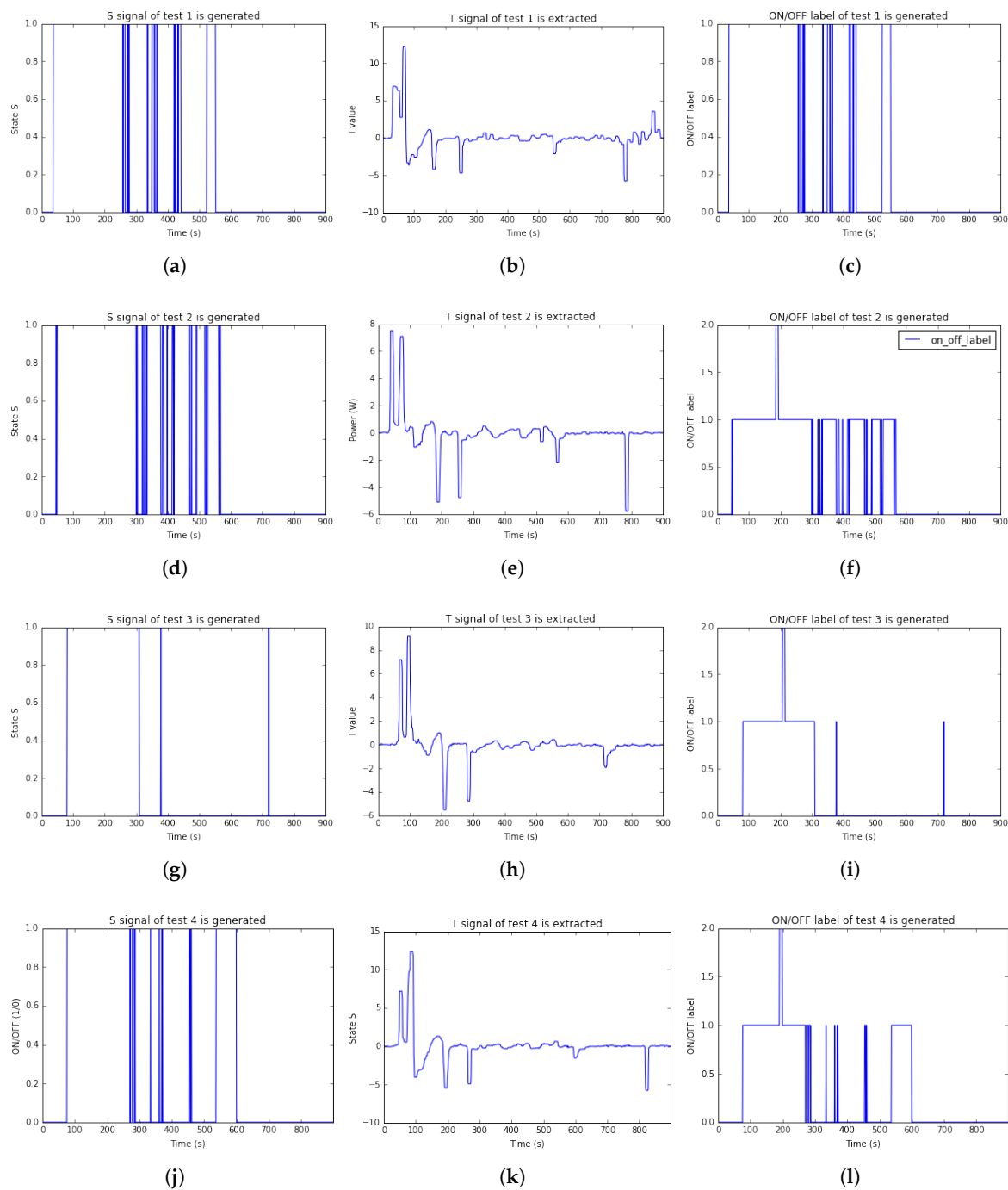


Figure 13. State S , T & ON/OFF label signals of test 1 to test 4 data: (a–c) test 1; (d–f) test 2; (g–i) test 3; (j–l) test 4.

4.2.2. Classification Results

- In ML, this paper applied Decision Tree model for classification ON/OFF event of the household appliance in this study. The reason for selecting Decision Tree model in classification task related to classification accuracy. This study tried to apply other models such as SVM, Random Forest, Multilayer Perceptrons (MLP), etc., however, the result of them obtained not well. This paper used T signals for input feature and output feature is ON/OFF labeling for the approached models. Five T signals extracted from five appliances are training data. The testing data is twenty T signals extracted from twenty tests data. This paper used CM to display the classification result.

The performance classification results for detecting ON/OFF of Decision Tree model are shown in Figure 14. The main diagonal of confusion matrix represents the number of correctly samples are predicted by the applied model.

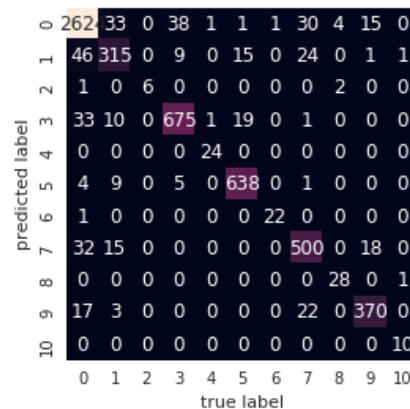


Figure 14. Confusion matrix of Decision Tree model for classification ON/OFF label.

Table 7 is classification report for each event detection of five appliances using other performance metrics such as precision, recall, F1-score. The average accuracy of event detection model on these appliances data obtained 93%.

Table 7. Classification result for event detection ON/OFF of each appliance.

Label	Event Device	Precision	Recall	F1-Score	Support
0	No Event	0.95	0.96	0.95	2747
1	Air-purifier_ON	0.82	0.77	0.79	411
2	Air-purifier_OFF	1.00	0.67	0.80	9
3	Fan_ON	0.93	0.91	0.92	739
4	Fan_OFF	0.92	1.00	0.96	24
5	Hairdryer_ON	0.95	0.97	0.96	657
6	Hairdryer_OFF	0.96	0.96	0.96	23
7	LG-monitor_ON	0.87	0.88	0.87	565
8	LG-monitor_OFF	0.82	0.97	0.89	29
9	Samsung-monitor_ON	0.92	0.90	0.91	412
10	Samsung-monitor_OFF	0.83	1.00	0.91	10
avg/total		0.93	0.93	0.93	5629

In summary, Table 8 pointed the average performance of the approached model. The performance classification of our approach model achieved 92.64% and loss rate was 1.67.

Table 8. Classification result of Decision Tree model on energy data.

Metric	Performance
Accuracy	0.926413
Precision	0.927533
Recall	0.926413
F1	0.926778
Loss	1.66637

- In DL, LSTM model is used for classification ON/OFF event of appliances. Similar to Decision Tree model, this paper used T signals and ON/OFF labeling for input and output features, respectively. For testing data, this research also used twenty T signals of twenty tests data. Setting up hyper-parameters of LSTM model are described in Table 9 as follows.

Table 9. Hyper-parameter of LSTM model in the approach.

Hyper-Parameter Name	Value
Input	T signals
Hidden size	100
Batch size	893
time step	1
Number of LSTM layers	3
Epochs	5000
Output	On_Off_labeling

To evaluate LSTM model, this paper uses accuracy metric and Root Mean Square Error (RMSE) for measuring performance and loss of the model, respectively. RMSE is the standard deviation of the residuals (prediction errors). The formula to calculate RMSE is as follows.

$$\text{RMSE} = \sqrt{(pr - ex)^2} \quad (24)$$

where pr is predicted output and ex is expected output.

Figure 15 presents the results of LSTM model for classification ON/OFF of five appliances data. Besides, LSTM model's results on test 1 to test 4 data are shown in Figure 16 and the others tests' results in Figure A4 at Appendix A. The average accuracy of LSTM model obtained 96.85% for detecting ON/OFF appliances. The loss of LSTM model (RMSE) obtained 0.60632.

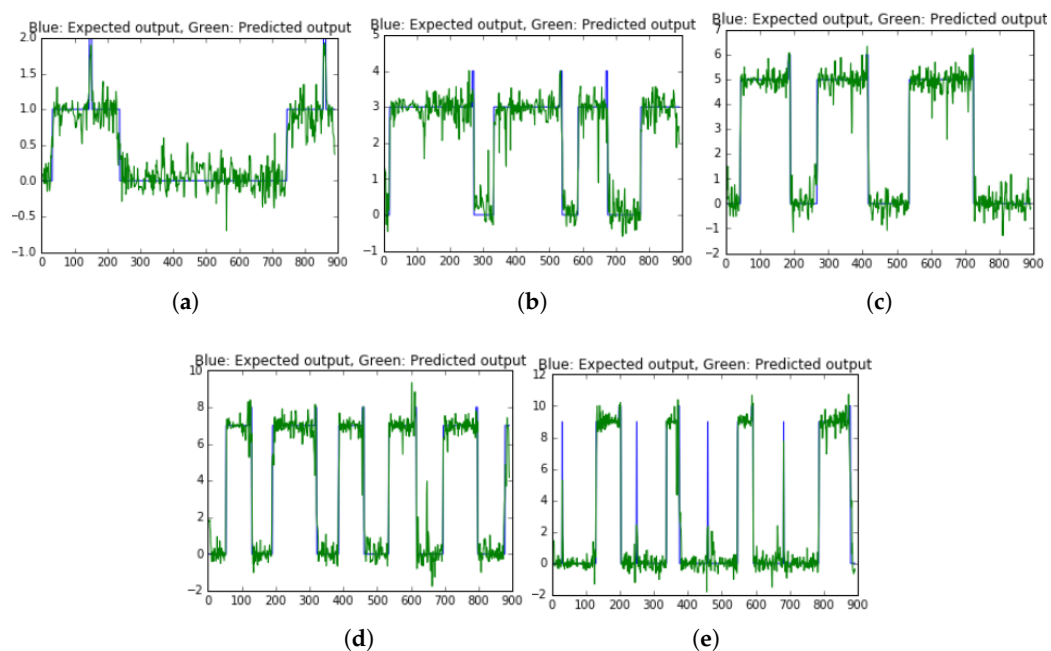


Figure 15. Classification results of LSTM model on 5 appliances data: (a) Airpurifier device; (b) Fan device; (c) Hairdryer device; (d) LG monitor device; (e) Samsung monitor device.

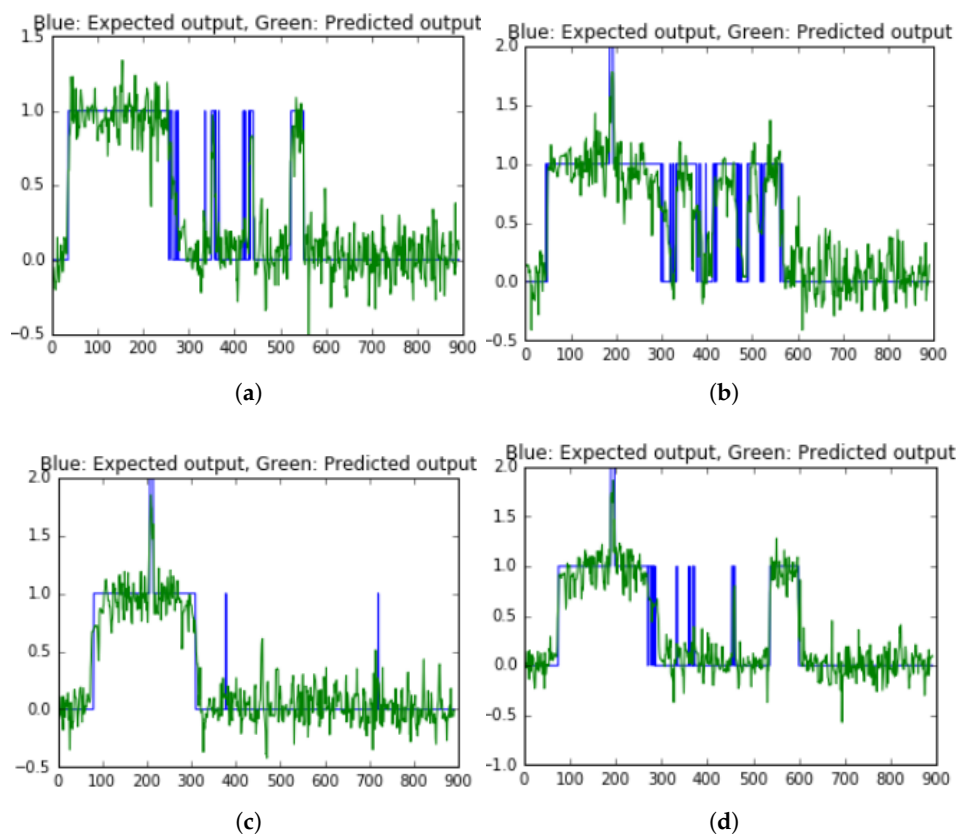


Figure 16. Classification results of LSTM model on test 1 to test 4 data: (a) Test 1; (b) Test 2; (c) Test 3; (d) Test 4.

4.2.3. Applying the Proposed Method on a Publicly Available Dataset

There are several NILM available datasets such as Building-Level fully-labeled dataset for Electricity Disaggregation (BLUED), UK-DALE, Residential Energy Disaggregation Dataset (REDD), The Almanac of Minutely Power dataset (Version 2) (AMPds2) etc. This paper applied the proposed method on AMPds2 [33] dataset which is low sampling rate data. Although several data file formats are published such as Original file format, Tab-delimited, Rdata format, Variable metadata, etc. The original file format in CSV is used in this paper. In this dataset, there are 19 appliances data in isolation have already collected and tested in electricity data. There are no publicly aggregated data between the loads in this dataset. Therefore, this paper only performed and tested on individual load. Among available loads, 5 appliances data are randomly selected to use in this experiment. These appliances are Master and South Bedroom, Basement Plugs and Lights, Instant Hot Water Unit, Entertainment: TV, PVR, AMP, Kitchen Fridge. Besides, ON/OFF label value are determined and declared from 1 to 10 corresponding to each appliance. The information of these appliances and their ON/OFF labels are described in Table 10 as follows.

Table 10. Five appliances information in AMPds2 dataset selected [34].

CSV File Name	Appliance Name	Watts (W)	ON Label	OFF Label
Electricity_B2E	Master and South Bedroom	175	1	2
Electricity_BME	Basement Plugs and Lights	387	3	4
Electricity_HTE	Instant Hot Water Unit	70	5	6
Electricity_TVE	Entertainment: TV, PVR, AMP	415	7	8
Electricity_FGE	Kitchen Fridge	525	9	10

About processing dataset, in this dataset, there are 11 electric data features in each appliance data. Timestamp (TS) feature is Unix timestamp value in this dataset. Electricity measurement is at one minute intervals. This dataset was collected to total of 1,051,200 readings per meter for 2 years of monitoring from 2012 to 2014. Therefore, these loads data are big dataset. In testing performance of the presently proposed methodology, the range of TS in the first 500 readings/samples value data from 1333263600 to 1333323540 is selected. This range of TS value is converted to Universal Time Coordinated (UTC) corresponding to the date time range from 2012-4-01 7:00 a.m. to 2012-4-01 15:19 p.m. The total of time series data is 500 min. Because real power P is active power data, hence P is selected for the input data in the experiment. Figure 17 visualizes five real power P signals time series data were selected.

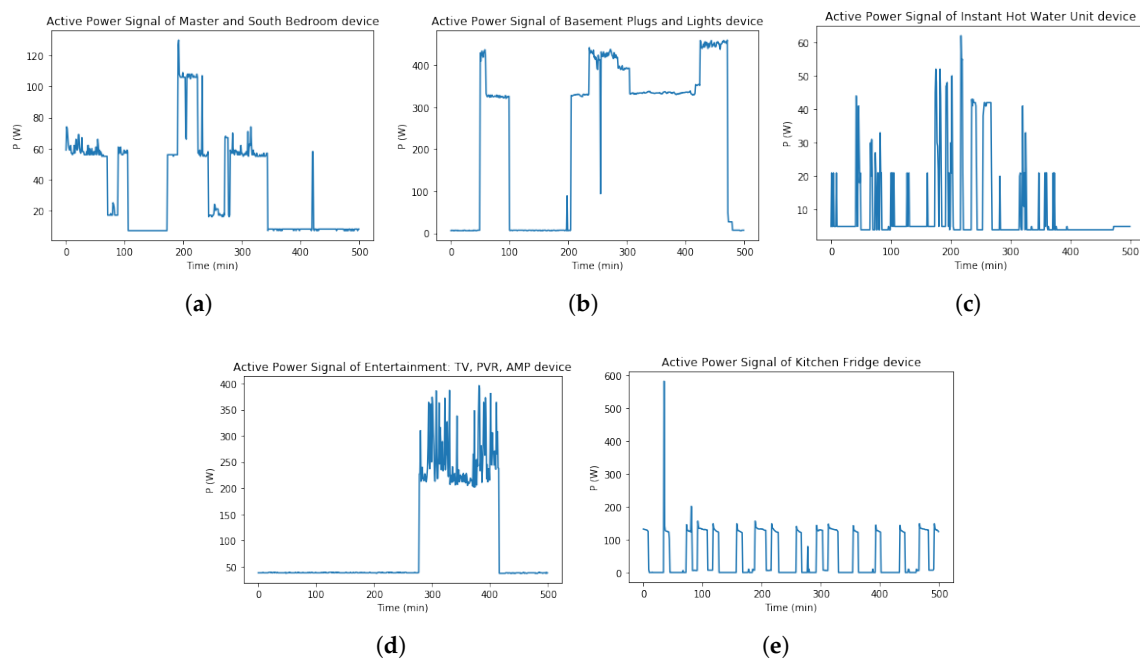


Figure 17. Active power signal of appliance data: (a) Master and South Bedroom; (b) Basement Plugs and Lights appliance; (c) Instant Hot Water Unit appliance; (d) Entertainment: TV, PVR, AMP appliance; (e) Kitchen Fridge appliance.

About feature extraction result, this paper applied the proposed solution to generate and extract S , T , ON/OFF label signals of each appliance. Figure 18 illustrates these results of each appliance.

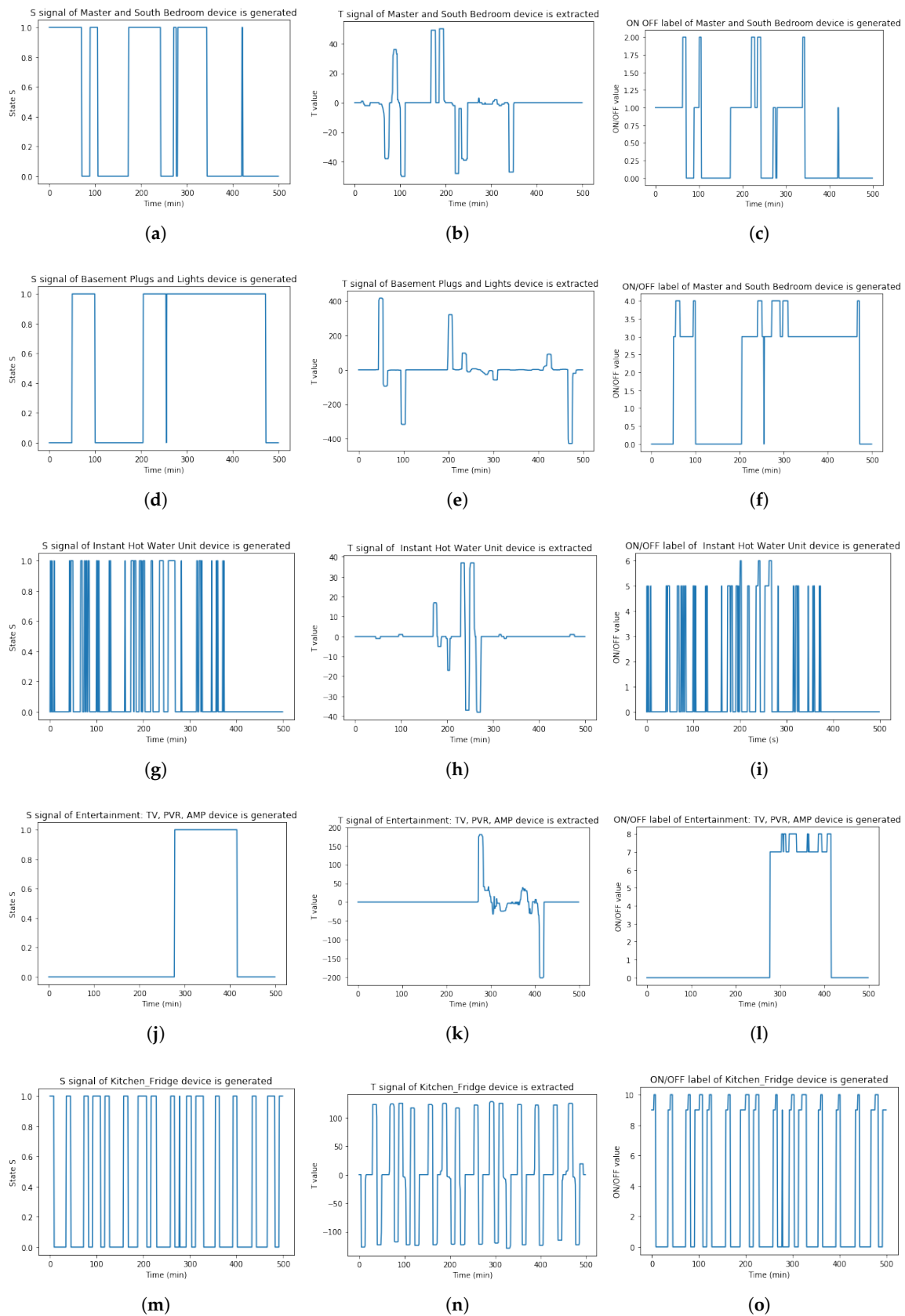


Figure 18. State S, T & ON/OFF label signals of 5 appliances: (a–c) Master and South Bedroom device; (d–f) Basement Plugs and Lights device; (g–i) Instant Hot Water Unit device; (j–l) Entertainment: TV, PVR, AMP device; (m–o) Kitchen Fridge device.

About classification result, this paper applied two approached learning models on this selected dataset.

- In ML, the classification result of Decision Tree model is presented in Figure 19. Furthermore, Table 11 presents the event detection results in detail of this confusion matrix result. The average classification performance of Decision Tree model on this dataset achieved 98.6% accuracy and loss rate was 0.488.

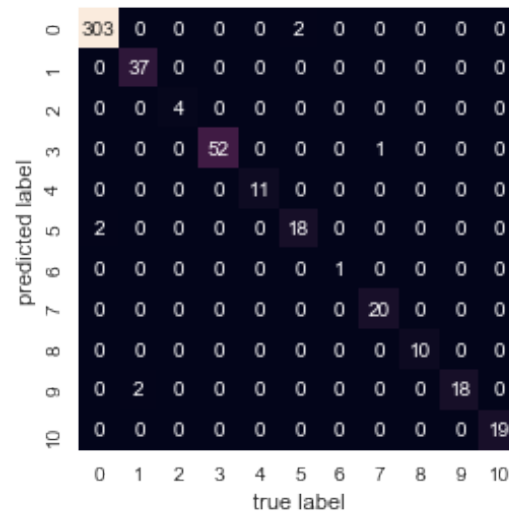


Figure 19. Confusion Matrix of Decision Tree model for classification ON/OFF label on five appliances data in AMPds2.

Table 11. Classification result for event detection ON/OFF of each appliance in AMPds2.

Label	Event Device	Precision	Recall	F1-Score	Support
0	No Event	0.99	0.99	0.99	305
1	Master and South Bedroom_ON	0.95	1.00	0.97	37
2	Master and South Bedroom_OFF	1.00	1.00	1.00	4
3	Basement Plugs and Lights_ON	1.00	0.98	0.99	53
4	Basement Plugs and Lights_OFF	1.00	1.00	1.00	11
5	Instant Hot Water_ON	0.90	0.90	0.90	20
6	Instant Hot Water_OFF	1.00	1.00	1.00	1
7	Entertainment: TV, PVR, AMP_ON	0.95	1.00	0.98	20
8	Entertainment: TV, PVR, AMP_OFF	1.00	1.00	1.00	10
9	Kitchen Fridge_ON	1.00	0.90	0.95	20
10	Kitchen Fridge_OFF	1.00	1.00	0.91	19
	avg/total	0.99	0.99	0.99	500

- In DL, the classification result of LSTM model obtained 96.78% accuracy and RMSE was 0.5124. Figure 20 illustrates the predicted results ON/OFF label of 5 appliances compare to the expected outputs.

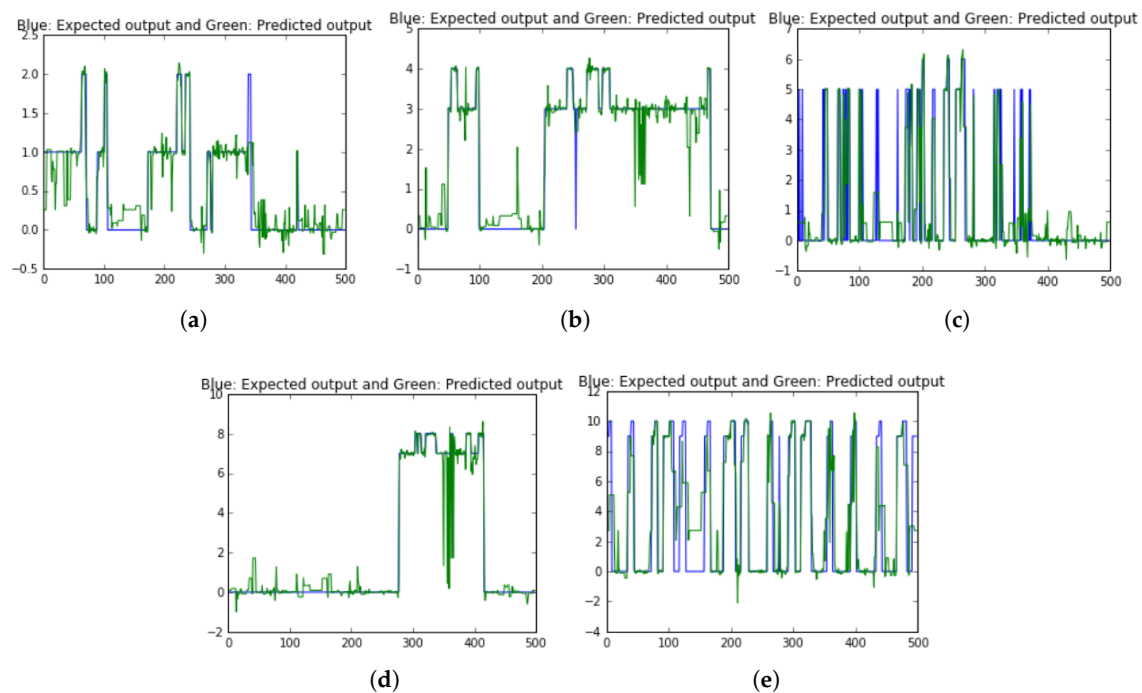


Figure 20. Classification results of LSTM model on 5 appliances data selected in AMPds2: (a) Master and South Bedroom; (b) Basement Plugs and Lights appliance; (c) Instant Hot Water Unit appliance; (d) Entertainment: TV, PVR, AMP appliance; (e) Kitchen Fridge appliance.

In summary, the presently proposed method achieved high-performance accuracy for load identification on the AMPds2 dataset with over 96% accuracy on both approached learning models.

5. Conclusions and Further Works

This study built a complete NILM framework including data acquisition, appliance feature extraction, classification data and monitoring energy data. First, this paper collected the personal NILM data in low-sampling rate including five household appliances energy data and twenty tests case data in the data acquisition task. Secondly, state of the appliance and transient signals are extracted to generate ON/OFF label on personal data in feature extraction task. The proposed transient signal is a transient signal to detect ON/OFF event of appliances. Thirdly, two models in ML and DL fields are stated to detect event on loads. In particular, the Decision Tree model and LSTM model are applied to perform two classification models with average accuracies achieved 92.64% and 96.85%, respectively. Finally, in monitoring energy data, a website platform system are developed to display the results of load identification. Besides, the proposed method is applied on a publicly available dataset, AMPds2. Both approach models obtained over 96% on this sub-data. In future work, this solution can be applied to extract transient signal on other publicly available datasets to widely apply the proposed methodology.

6. Data Availability

The household appliances energy consumption data set was downloaded from the figshare repository with DOI https://figshare.com/articles/NILM_EnergyData/7269692, and other data generated or analyzed during this study are available from the corresponding author on reasonable request.

Author Contributions: T.-T.-H.L. stated the problems and their solution. H.K. managed this project and reviewed the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (2014-1-00743) supervised by the IITP (Institute for Information & communications Technology Promotion). This work was supported by the Human Resources Development program (No. 20184030202060) of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government Ministry of Trade, Industry and Energy.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

NILM	Non-Intrusive Load Monitoring
ILM	Intrusive Load Monitoring
UART	Universal Asynchronous Receiver/Transmitter
EEPROM	Electrically Erasable Programmable Read-Only Memory
RMS	Root Mean Square
AI	Artificial Intelligent
ARM	Acorn RISC Machines
GPU	Graphic Processing Unit
SVM	Support Vector Machine
DNN	Deep Neural Network
MLP	Multilayer Perceptrons
KNN	K-Nearest Neighbour
HMM	Hidden Markov Model
MSE	Mean Squared Error
ML	Machine Learning
DL	Deep Learning
CART	Classification and Regression Trees
LSTM	Long Short Term Memory
RMSE	Root Mean Square Error
ANN	Artificial Neural Network
WTC	Wavelet Transform Coefficient
CM	Confusion Matrix
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
UTC	Universal Time Coordinated
BLUED	Building-Level fully-labeled dataset for Electricity Disaggregation
REDD	Residential Energy Disaggregation Dataset
UK-DALE	UK Domestic Appliance-Level Electricity
AMPds2	The Almanac of Minutely Power dataset (Version 2)

Appendix A

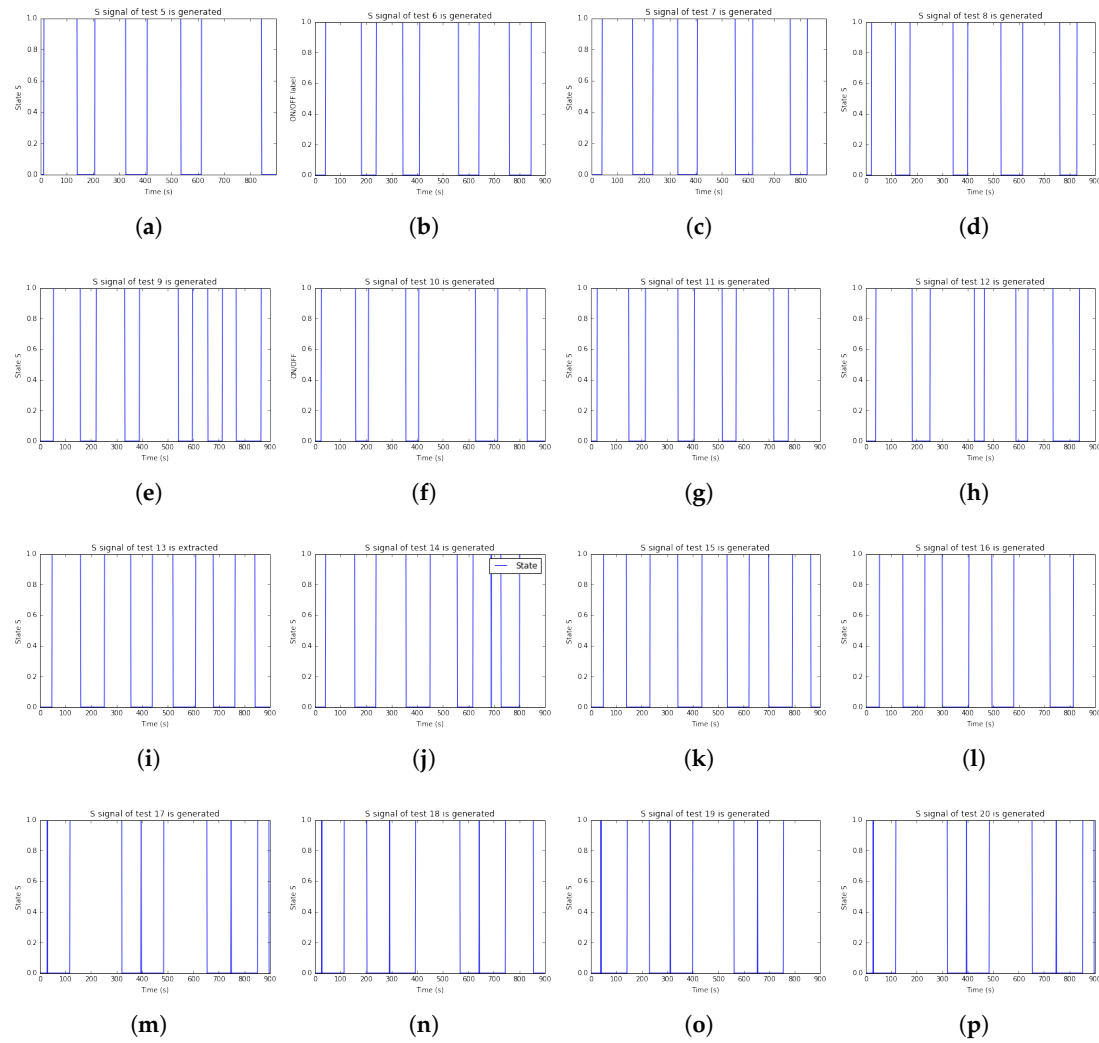
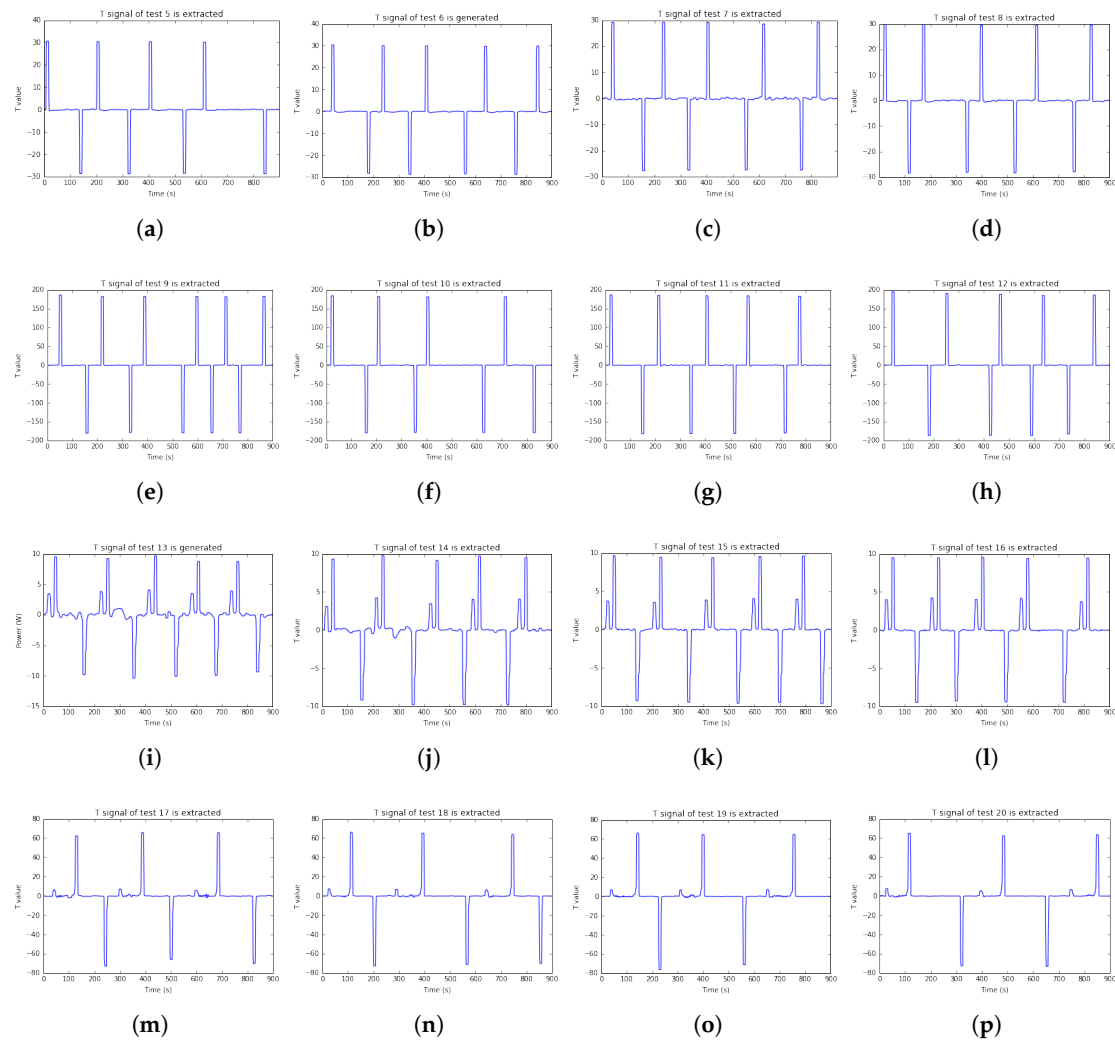


Figure A1. State S signals of test 5 to test 20 data.

Figure A2. Transient T signals of test 5 to test 20 data.

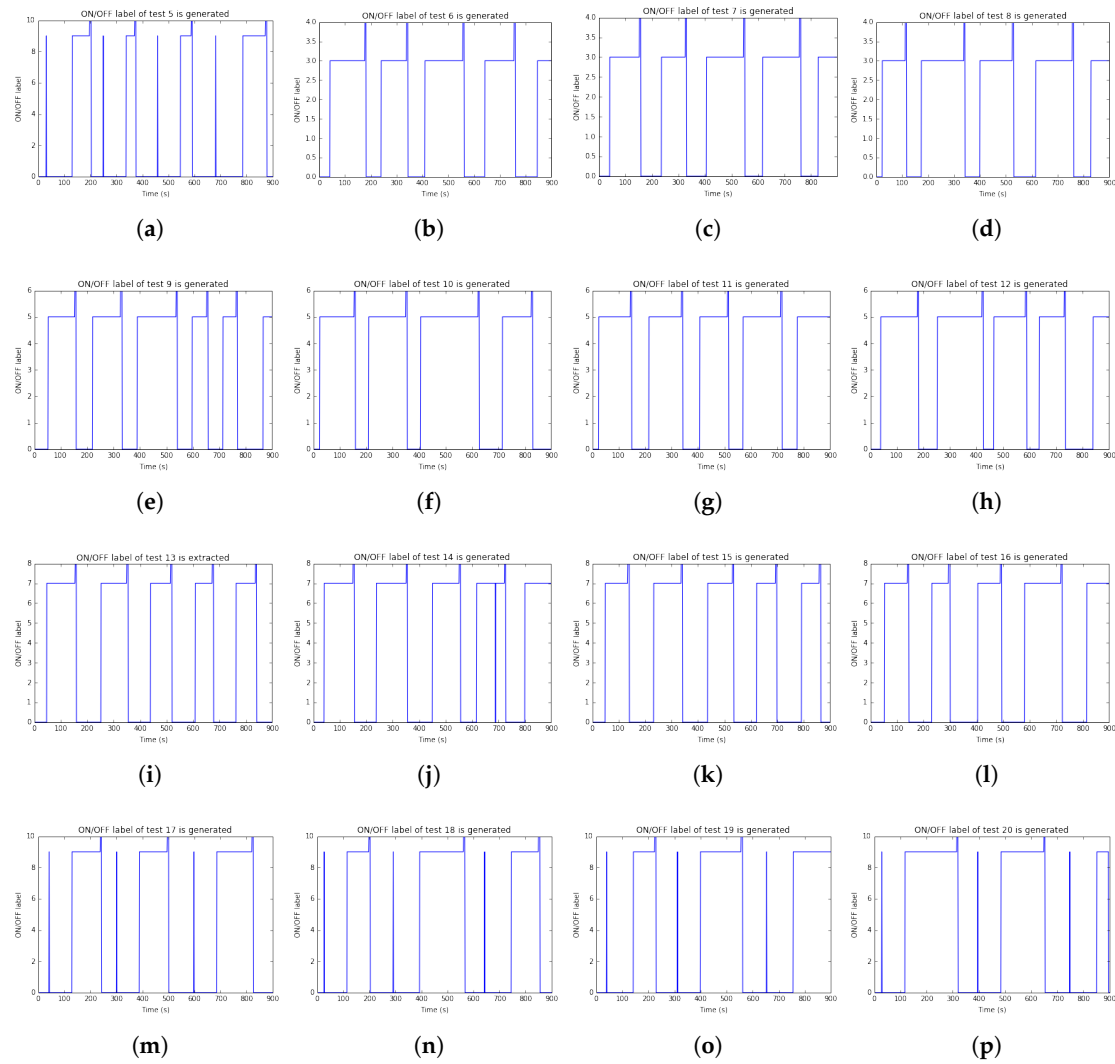


Figure A3. ON/OFF labels of test 5 to test 20 data.

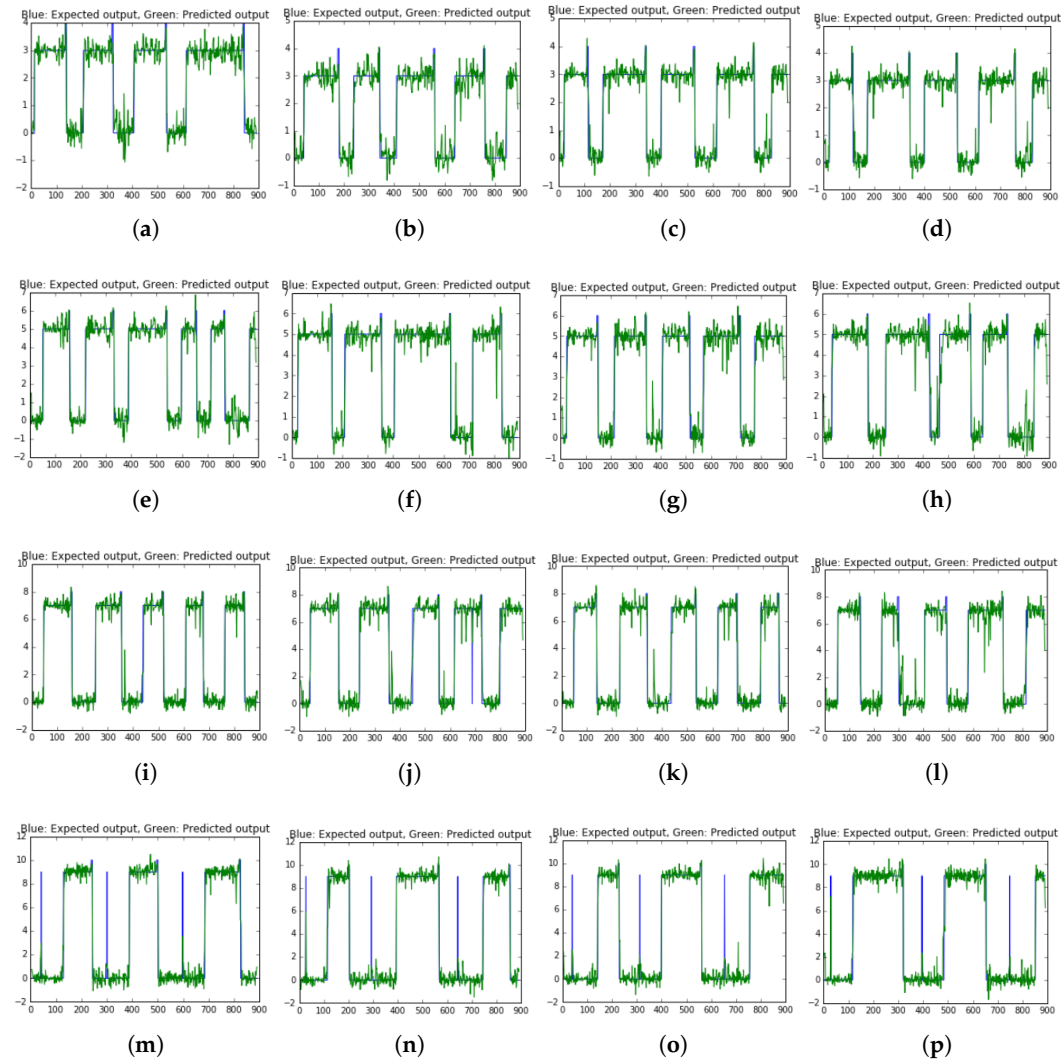


Figure A4. Classification results of LSTM model on test 5 to test 20 data: ON/OFF label is expected output with Blue color; ON/OFF label is predicted output with Green color.

References

1. Lee, J.; Bagheri, B.; Kao, H.A. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf. Lett.* **2015**, *11*, 18–23. [[CrossRef](#)]
2. Darby, S. *The Effectiveness of Feedback on Energy Consumption a Review for DEFRA of the Literature on Metering, Billing and Direct Displays*; Environmental Change Institute, University of Oxford: Oxford, UK, 2006.
3. Kelly, J.; Knottenbelt, W. Does disaggregated electricity feedback reduce domestic electricity consumption? *arXiv* **2016**, arXiv:1605.00962.
4. Abubakar, I.; Khalid, S.N.; Mustafa, M.W.; Shareef, H.; Mustapha, M. Application of load monitoring in appliances energy management: A review. *Renew. Sustain. Energy Rev.* **2017**, *67*, 235–245. [[CrossRef](#)]
5. Hart, G.W. Nonintrusive appliance load monitoring. *IEEE Proc.* **1992**, *80*, 1870–1891. [[CrossRef](#)]
6. Roos, J.G.; Lane, I.E.; Botha, E.C.; Hancke, G.P. Using neural networks for non-intrusive monitoring of industrial electrical loads. In Proceedings of the Instrumentation and Measurement Technology Conference, Hamamatsu, Japan 10–12 May 1994; pp. 1115–1118.
7. Sultanem, F. Using appliance signatures for monitoring residential loads at meter panel level. *IEEE Trans. Power Deliv.* **1991**, *6*, 1380–1385. [[CrossRef](#)]
8. Laughman, C.; Lee, K.; Cox, R.; Shaw, S.; Leeb, S.; Norford, L.; Armstrong, P. Power signature analysis. *IEEE Power Energy Mag.* **2003**, *1*, 56–63. [[CrossRef](#)]
9. Patel, S.N.; Robertson, T.; Kientz, J.A.; Reynolds, M.S.; Abowd, G.D. At the flick of a switch: detecting and classifying unique electrical events on the residential power line. In *International Conference on Ubiquitous Computing*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4717, pp. 271–288, ISBN 9783540748526.
10. Barsim, K.S.; Streubel, R.; Yang, B. An Approach for Unsupervised NonIntrusive Load Monitoring of Residential Appliances. In Proceedings of the 2nd International Workshop on Non-Intrusive Load Monitoring, Austin, TX, USA, 3 June 2014.
11. Yung, F.W.; Ahmet, Y.S.; Tom, D.; Voon, S.W. Recent approaches to non-intrusive load monitoring techniques in residential settings. In Proceedings of the IEEE Symposium on Computational Intelligence Applications in Smart Grid, Singapore, 16–19 April 2013; pp. 73–79.
12. Hana, A.; Jing, L.; Lina, S.; Vladimir, S. A low-complexity energy disaggregation method: Performance and robustness. In Proceedings of the IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG), Orlando, FL, USA, 9–12 December 2014.
13. Hyungsul, K.; Manish, M.; Martin, A.; Geoff, L.; Jiawei, H. Unsupervised Disaggregation of Low Frequency Power Measurements. In Proceedings of the 11th SIAM International Conference on Data Mining, Phoenix, AZ, USA, 28–30 April 2011; pp. 747–758.
14. Zico, K.; Tommi, J.; Kolter, J.Z. Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Canary Islands, Spain, 21–23 April 2012; pp. 1472–1482.
15. Oliver, P.; Ghosh, S.; Weal, M.; Rogers, A. Non-intrusive Load Monitoring using Prior Models of General Appliance Types. In Proceedings of the twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI-12), Toronto, ON, Canada, 22–26 July 2012.
16. Stephen, M.; Fred, P.; Ivan, V.B.; Bob, G.; Lyn, B. Exploiting HMM Sparsity to Perform Online Real-Time Nonintrusive Load Monitoring. *IEEE Trans. Smart Grid* **2016**, *7*, 2575–2585.
17. Ahmed, Z.; Alexander, G.; Muhammad, A.I.; Sutharshan, R. Non-intrusive Load Monitoring approaches for disaggregated energy sensing: A survey. *Sensors* **2012**, *12*, 16838–16866.
18. Lukas, M.; Karim, S.B.; Bin, Y. How well can HMM model load signals. In Proceedings of the 3rd International Workshop on NonIntrusive Load Monitoring (NILM 2016), Vancouver, Canada, 14–15 May 2016.
19. Liang, J.; Ng, S.K.; Kendall, G.; Cheng, J.W. Load Signature Study Part I: Basic Concept, Structure, and Methodology. *IEEE Trans. Power Deliv.* **2010**, *25*, 551–560. [[CrossRef](#)]
20. Faustine, A.; Mvungi, H.N.; Kaijage, S.; Michael, K. A Survey on Non-Intrusive Load Monitoring Methodies and Techniques for Energy Disaggregation Problem. *arXiv* **2017**, arXiv:1703.00785.
21. Kelly, J.; Knottenbelt, W. Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments (BuildSys'15), Seoul, Korea, 4–5 November 2015; pp. 55–64.

22. Kim, J.; Le, T.T.; Kim, H. Nonintrusive load monitoring based on advanced deep learning and novel signature. *Comput. Intell. Neurosci.* **2017**, 2017, 4216281. [[CrossRef](#)] [[PubMed](#)]
23. Kanghang, H.; Lina, S.; Jing, L.; Vladimir, S. Non-Intrusive Load Disaggregation using Graph Signal Processing. *IEEE Trans. Smart Grid* **2016**, 9, 1739–1747.
24. Figueiredo, M.B.; De Almeida, A.; Ribeiro, B. *An Experimental Study on Electrical Signature Identification of Non-Intrusive Load Monitoring (NILM) Systems*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2011; Volume 6594, pp. 31–40, ISBN 9783642202667.
25. Chang, H.H.; Chen, K.L.; Tsai, Y.P.; Lee, W.J. A New Measurement Method for Power Signature of Non-intrusive Demand Monitoring and Load Identification. In Proceedings of the Industry Applications Society Annual Meeting (IAS), Orlando, FL, USA, 9–13 October 2011.
26. Kofi, A.A.; Sekyung, H.; Soohye H. Real-Time Recognition Non-Intrusive Electrical Appliance Monitoring Algorithm for a Residential for a Residential Building Energy Management System. *Energies* **2015**, 8, 9029–9048. [[CrossRef](#)]
27. Chang, H.H.; Lian, K.L.; Su, Y.C.; Lee, W.J. Power-Spectrum-Based Wavelet Transform for Nonintrusive Demand Monitoring and Load Identification. *IEEE Trans. Ind. Appl.* **2014**, 50, 2081–2089. [[CrossRef](#)]
28. Zeifman, M.; Akers, C.; Roth, K. Nonintrusive appliance load monitoring: Review and outlook. *IEEE Trans. Consum. Electron.* **2011**, 57, 76–84. [[CrossRef](#)]
29. Hassan, T.; Javed, F.; Arshad, N. An Empirical Investigation of V-I Trajectory Based Load Signatures for Non-Intrusive Load Monitoring. *IEEE Trans. Smart Grid* **2013**, 5, 870–878. [[CrossRef](#)]
30. Nilson, H.; Kodjo, A.; Suosso, K.; Sayed, S.H.; Michael, F. Power Estimation of Multiple Two-State Loads Using A Probabilistic Non-Intrusive Approach. *Energies* **2018**, 11, 88. [[CrossRef](#)]
31. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: Wadsworth, OH, USA; Belmont, CA, USA, 1984.
32. Sepp, H.; Jurgen, S. Long short-term memory. *Neural Comput.* **1997**, 9, 1735–1780.
33. Makonin, S.; Ellert, B.; Bajic, I.V.; Popowic, F. Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Sci. Data* **2016**, 3, 160037. [[CrossRef](#)] [[PubMed](#)]
34. Makonin, S. AMPds2: The Almanac of Minutely Power Dataset (Version 2). 2016. Available online: <https://doi.org/10.7910/DVN/FIE0S4> (accessed on 30 October 2018).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).