



# Akte-Liquid: Acoustic-based Liquid Identification with Smartphones

18

XUE SUN, WENWEN DENG, and XUDONG WEI, Northwest University, IoT Research

Center-Northwest University, China

DINGYI FANG, Northwest University, Northwest University-Jingdong Wisdom Cloud Joint Research

Center for AI & IoT, China

BAOCHUN LI, University of Toronto, Canada

XIAOJIANG CHEN, Northwest University, Shaanxi International Joint Research Centre for the

Battery-Free IoT, China

Liquid identification plays an essential role in our daily lives. However, existing RF sensing approaches still require dedicated hardware such as RFID readers and UWB transceivers, which are not readily available to most users. In this article, we propose Akte-Liquid, which leverages the speaker on smartphones to transmit acoustic signals, and the microphone on smartphones to receive reflected signals to identify liquid types and analyze the liquid concentration. Our work arises from the acoustic intrinsic impedance property of liquids, in that different liquids have different intrinsic impedance, causing reflected acoustic signals of liquids to differ. Then, we discover that the amplitude-frequency feature of reflected signals may be utilized to represent the liquid feature. With this insight, we propose new mechanisms to eliminate the interference caused by hardware and multi-path propagation effects to extract the liquid features. In addition, we design a new Siamese network-based structure with a specific training sample selection mechanism to reconstruct the extracted feature to container-irrelevant features. Our experimental evaluations demonstrate that Akte-Liquid is able to distinguish 20 types of liquids at a higher accuracy, and to identify food additives and measure protein concentration in the artificial urine with a 92.3% accuracy under 1 mg/100 mL as well.

CCS Concepts: • Human-centered computing → Ubiquitous and mobile computing systems and tools;

Additional Key Words and Phrases: Liquid identification, acoustic sensing, mobile sensing, neural networks

---

This work is supported by the National Natural Science Foundation of China (61772422, 61972316). This work is also partially supported by NSFC A3 Foresight Program Grant 62061146001, the Shaanxi International Science and Technology Cooperation Program under grant agreement 2019KWZ-05, 2022KW-11, 2018SF-369, 2019GY-012, and the ShaanXi Science and Technology Innovation Team Support Project under grant agreement 2018TD-026.

Authors' addresses: X. Sun, W. Deng, and X. Wei, Northwest University, and IoT Research Center-Northwest University, xi'an, China; emails: sunxue@stumail.nwu.edu.cn, redpanda.brilliant@gmail.com, weixudong@stumail.nwu.edu.cn; D. Fang, Northwest University and Northwest university-Jingdong Wisdom Cloud Joint Research Center for AI & IoT, xi'an, China; email: dyf@nwu.edu.cn; B. Li, University of Toronto, Toronto, Canada; email: bli@ece.toronto.edu; X. Chen, Northwest University and Shaanxi International Joint Research Centre for the Battery-Free IoT, xi'an, China; email: xjchen@nwu.edu.cn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2023 Association for Computing Machinery.

1550-4859/2023/02-ART18 \$15.00

<https://doi.org/10.1145/3551640>

**ACM Reference format:**

Xue Sun, Wenwen Deng, Xudong Wei, Dingyi Fang, Baochun Li, and Xiaojiang Chen. 2023. Akte-Liquid: Acoustic-based Liquid Identification with Smartphones. *ACM Trans. Sensor Netw.* 19, 1, Article 18 (February 2023), 24 pages.

<https://doi.org/10.1145/3551640>

---

## 1 INTRODUCTION

Liquid identification has recently garnered significant research attention, as potential application scenarios include cheap detection of water contamination, adulteration, and food additives in beverages, as well as kidney diseases using urine analysis [4, 19, 26, 28, 29]. Traditionally, a wide range of advanced technologies are adopted in liquid identification, such as Raman spectroscopy, mass-spectrometer-based technologies, and NMR spectroscopy [5]. Those liquid identification systems rely on expensive equipment in specialized labs, which do not meet the demand for identifying liquids in mobile environments.

Many applications, however, would benefit from identifying liquids by mobile and portable devices, especially for lay users. For example, in a convenience store, one would want to ensure that the juice sold is fresh and without food additives. Similarly, for some patients with chronic diseases, such as chronic kidney diseases, patients can perform urine tests at home instead of paying regular visits to the hospital. Naturally, it is most convenient to utilize a smartphone to achieve liquid identification. In this article, we take the initial step toward this vision by developing a reliable liquid identifier of acoustic signals using only smartphones, and consider the feasibility of delivering such services to lay users without dedicated hardware and specialized setup.

Typically, liquid identification needs to measure a particular property of liquids, such as electric permittivity or acoustic absorption, which can be analyzed to identify the type and characteristics of the liquid. Materials' absorption coefficient measurement method in sound measurement inspired us. We explore whether liquid acoustic impedance could be used to identify liquids, since different liquids have different acoustic impedance [13]. Acoustic impedance, an inherent property of liquids, describes the ratio of the sound pressure to the speed of molecule movement caused by the sound pressure of the medium. Since different liquids have different molecular structures, the velocity of molecular motion of different liquids induced by the same acoustic pressure is also different, which leads to different acoustic impedance of different liquids.

The liquid acoustic impedance, however, is very difficult to measure directly by using the speaker and microphone on smartphones. Fortunately, the acoustic impedance of liquids will affect the liquid's reflected signals. By analyzing the reflected signals, we find that the energy distribution in the frequency domain differs between the reflected signals of different liquids. Therefore, we investigate whether it is feasible to extract liquid features from the reflected signal to build our liquid identification system.

In this article, we present Akte-Liquid, a noninvasive and lightweight smartphone application that uses acoustic signals to identify liquids by leveraging the built-in speaker and microphone on off-the-shelf smartphones.

Akte-Liquid, which meets the needs for daily personal use, offers several advantages:

- (1) The convenience of operation. The system could utilize common devices available to lay users, rather than depending on any extra hardware or external infrastructure support, and without requiring a specialized setup. Unlike RF signals (such as Wi-Fi and Bluetooth signals) that need a pre-deployed infrastructure and additional hardware, acoustic signals have the advantage of being easily transmitted by the speaker and received by the microphone available on off-the-shelf smartphones.



Fig. 1. Visualization of Akte-Liquid: an acoustic-based liquid identification system utilizing smartphones.

- (2) Detection efficiency. The system could simplify the identification process and shorten its duration, promptly obtaining results without any noticeable latency.
- (3) High resolution. The system not only can distinguish different liquid types but also has sufficient resolution to detect small differences in solution concentrations.

As shown in Figure 1, Akte-Liquid generates acoustic signals through the front speaker and collects the reflected signals from the liquid using the front microphone. When the sound waves travel from the air into the liquid, the energy distribution of the acoustic reflection signals at frequency is induced by the acoustic signal reflectivity, which is determined by the acoustic impedance in the air and liquid together.

In this article, we first investigate the feasibility of utilizing acoustic signals to obtain the liquid features from the reflected signals for liquid identification. We discover that the different liquids have different energy distributions at specific frequencies. Motivated by this observation, we first extract the amplitude-frequency features from reflected signals, and then we propose a **differential amplitude frequency (DASD)** method and ratio **transfer function (TF)** method to represent liquid features. To address the issue of interference from the liquid volumes, we set a series of weights for each volume to calibrate the difference in amplitude-frequency values caused by the volume. Finally, as extracted features from the reflected signals contain both liquid features and container features, we propose a Siamese network [1] model with a specific training sample selection strategy to reconstruct the extracted features to container-irrelevant features, and further develop a transfer learning method to harness the power of a well-trained model on a large number of measurements, and to achieve high accuracy on new liquid samples without having to train the whole model.

Highlights of original contributions in this article are as follows:

- (1) We present a low-cost solution for a liquid identification system that exploits acoustic signals generated by smartphones and reflected by liquids as a fingerprint of liquids. As a result, our design delivers convenience using only a smartphone discriminating the liquid types, detecting food additives in beverages, tracking protein content in artificial urine, and identifying different water sources without any specialized hardware or any limitation on the container.
- (2) We employ a Siamese-network-based structure to reconstruct the features extracted from reflected signals to container-irrelevant features, which deliver significant performance improvements when our system is used on new containers.
- (3) Our results indicate that on 20 different types of liquids, Akte-Liquid can achieve at least an identification accuracy of 98%. Moreover, we can measure the protein in artificial urine with a low concentration of 1 mg/100 mL and also have the capability of detecting water sources and food additives in beverages at a low concentration of 1%.

The remainder of the article is organized as follows. Section 2 introduces related work. Section 3 presents the background and preliminaries. Section 4 illustrates our system design and overview. In Sections 5–7, we present the detailed design and algorithms used in Akte-Liquid. Our evaluation results are shown in Section 8. Section 9 adds additional insights into the limitations of our system.

## 2 RELATED WORK

In this section, we review the related literature in liquid identification, the acoustic sensing method, and using **Convolutional Neural Networks (CNNs)** for classification.

**Liquid Identification Systems.** Currently, there have been some existing works on liquid identification [4, 6, 8, 9, 11, 26, 28]. For example, LiquID [4] calculates the phase, amplitude, and propagation delay of the UWB signals penetrating the liquid to estimate the liquid's permittivity, which in turn identifies the liquid. RF-EATS [8] utilizes the RF coupling effect between the RFID antenna and the liquid container's tag, which manifests that the frequency response of reflected signals varies with the liquids, so that the liquid can be identified. However, they use special signals (i.e., RFID, UWB), which are unavailable to the general public. Compared with our work, two closely related works are Capcam [29] and Vi-Liquid [11]. They all focus on liquid identification using smartphones. Capcam measures liquid surface tension based on a camera and a Vibro-motor of the smartphone to identify liquids. Vi-Liquid estimates the liquid viscosity by leveraging the Vibro-motor and accelerometer of smartphones. However, they have some restrictions on the containers (specific containers) and smartphones' hardware (cameras' resolution).

**Acoustic-based Sensing Techniques.** Acoustic signals have been widely employed for large applications, such as user identification [7], eye blink detection [17], fine-grained motion tracking [18], and respiration and heartbeat monitoring [31]. For example, EarEcho [7] utilizes the fact that different people's ear canals have different acoustic impedance to achieve ear-canal-based authentication. BlinkListener [17] also utilizes the different acoustic impedance between the eyelid and eyeball to achieve eye blink detection. Akte-Liquid also implements liquid identification depending on the differences in the acoustic impedance of liquids.

**Using CNN for Classification.** CNN was first proposed by LeCun in 1998 [15]. CNN-based deep learning has been used for a wide variety of tasks such as image classification [10, 14, 20] and audio signal classification [2, 16, 32] to extract the audio feature as well as the speaker information to recognize speech and identify speakers. CNN has advantages of modeling spatial context information in the 2-D space and has achieved great success in the field of image classification. In Akte-Liquid, we construct a Siamese-network-based model with a specific training sample selection strategy to reconstruct the extracted features to container-irrelevant features.

## 3 BACKGROUND AND PRINCIPLE

Before delving into the technical details of Akte-Liquid, we first provide the principle of liquid detection exploiting the acoustic impedance.

The liquid acoustic impedance describes the product of liquid density  $\rho_0$  and its acoustic speed  $c_0$  [13]. When an acoustic wave is vertically incident from the air into liquid, the incident acoustic pressure  $P_i$  can be represented as

$$P_i = P_0 \cos(\omega t - kx + \varphi), \quad (1)$$

where  $P_0$  is the initial sound pressure,  $\omega$  is the angular frequency,  $k$  is the wave number associated with wave lengths, and  $\varphi$  is the initial phase. The reflected acoustic pressure  $P_r$  can be represented as

$$P_r = R * P_i, \quad (2)$$

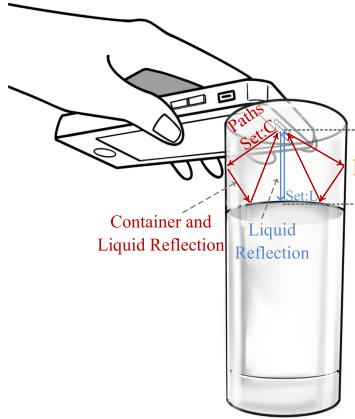


Fig. 2. Multi-path propagation of received signals.

in which  $R$  is the Liquid Reflectance, and  $R$  can be calculated as

$$R = \frac{z_{liquid} - z_{air}}{z_{liquid} + z_{air}}, \quad (3)$$

where  $z_{air}$  is the acoustic impedance of air, and  $z_{liquid}$  is the acoustic impedance of liquid.

In our work, we utilize the acoustic sensing approach to capture the liquid-induced signal variation. Specifically, as shown in Figure 2, for a single-frequency signal  $S(t) = \cos(2\pi ft + \phi_0)$ , with the frequency  $f$  and the initial phase  $\phi_0$ . The acoustic signal is transmitted continuously through the speaker of the smartphone. Then it propagates with different paths in the container, reflected by the container and liquid, and finally received by the smartphone's microphone. The received signals at time  $t$  can be written as

$$R(t) = \sum_{i \in M} A_i \cos(2\pi ft + \phi_i), \quad (4)$$

where  $M$  denotes the set of all paths of acoustic signals, and  $A_i$  is the amplitude reduction of the  $i$ th path. We then utilize the amplitude spectrum to profile the received signals' information. The amplitude spectrum of the received signals can be represented as

$$A(f) = \frac{FFT(R(t))}{N} = \sum_{i \in M} A_i = \sum_{j \in L} A_j + \sum_{j \in L} \sum_{k \in C} A_k + \sum_{q \in B} A_q. \quad (5)$$

It can be seen from Equation (5) that the amplitude  $\sum_{i \in M} A_i$  of signals is mainly composed of three parts,  $\sum_{j \in L} A_j$ ,  $\sum_{k \in C} A_k$ ,  $\sum_{q \in B} A_q$ , where  $L$  is the set of paths reflected by the liquid,  $C$  is the set of paths reflected by the container and liquid, and  $B$  is the set of other possible paths not related to the liquid. The amplitude variation of received signals is not only affected by liquids but also affected by the container and the length of the propagation path, such that changes in the liquid volume changes the smartphone position. As such, we cannot directly utilize the original amplitude-frequency pattern as the liquid features.

Fortunately, we discover that when the container and the signal propagation path are unchanged, the amplitude variation of the reflected signals is mainly affected by liquids. Meanwhile, we discover that the differential amplitude value corresponding with frequency can eliminate the irrelevant factors when the propagation path is identical. Specifically, for two liquids  $a$  and  $b$  with

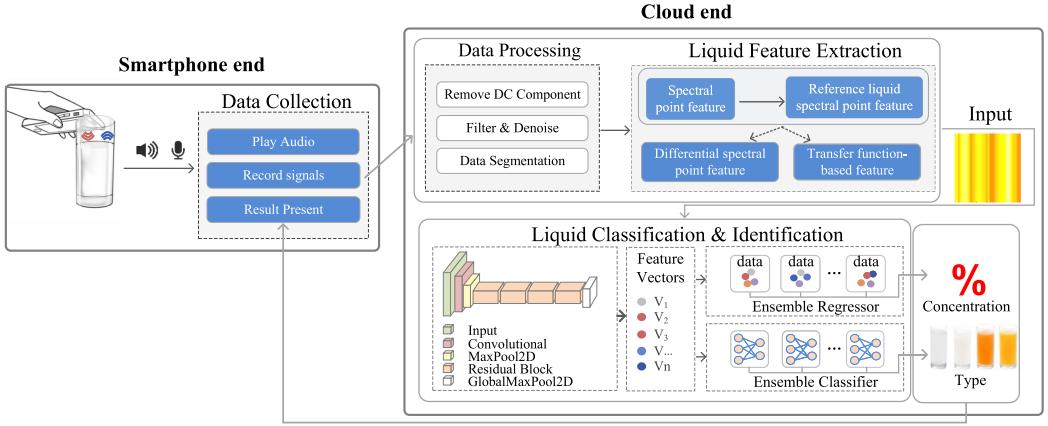


Fig. 3. Architecture of Akte-Liquid.

the same volume, the differential spectral points can be expressed as

$$\begin{aligned} D_A(a, b)(f) &= A(f_a) - A(f_b) \\ &= \sum_{j \in L} (A_{aj} - A_{bj}) + \sum_{j \in L} \sum_{k \in C} (A_{aj} A_{ak} - A_{bj} A_{bk}). \end{aligned} \quad (6)$$

From Equation (6), we find that the first part of the equation mainly reflects the amplitude difference between liquids, and the remaining part reflects the amplitude difference caused by containers and liquids, which illustrates the DASD feature affected by the containers.

We discover that the differential amplitude-frequency features are inadequate in distinguishing similar liquids, such as the same solution with slightly different concentrations. It cannot be employed as the unique feature for liquid identification.

Meanwhile, we are inspired by the absorption coefficient of materials measured in sound measurement methods [3, 12]. In sound measurement, the traditional method is to put the test material on one end of the impedance tube, while the sound source is put on the other end of the impedance tube, and the dual microphones are placed in different positions to receive the reflected signals by the test material in the impedance tube. Then, the ratio of the dual-channel receive signals is the transfer function  $H$ , which can represent the material's absorption properties. It can be represented as

$$H = \frac{P_2}{P_1} = \frac{\exp(jkx_1) + R * \exp(-jkx_1)}{\exp(jkx_2) + R * \exp(-jkx_2)}, \quad (7)$$

where  $x_1$  and  $x_2$  are the propagation path of the two microphones. The transfer function  $H$  can reflect the liquid reflectance  $R$ . Although the propagation path ( $x_1, x_2$ ) is difficult to obtain, the transfer function features can be extracted from the ratio of received signals.

Therefore, according to acoustic propagation characteristics, we design the liquid features based on the differential amplitude-frequency features and the transfer function features to reflect the changes of liquid to achieve the liquid identification.

#### 4 SYSTEM DESIGN

In order to design a mobile, portable, noninvasive liquid identification system, we propose Akte-Liquid, which only utilizes a built-in speaker and microphone of the smartphone to achieve liquid identification. We design and implement Akte-Liquid, which consists of a smartphone end and a cloud end, as illustrated in Figure 3.

**Data Collection.** It is a terminal interface to collect the liquid data, which utilizes the front built-in speaker of the smartphone to transmit the acoustic signals and the front microphone to capture the reflected acoustic signals.

**Data Processing.** In this module, we first synchronize the received signals using cross-correlation between the received signals and the transmitted signals. We then remove the DC component to eliminate the hardware current impact. Due to the transmitted signals used in our system (10 kHz to 16 kHz), we utilize a band-pass filter to remove the low-frequency (<10 kHz) and the high-frequency (>16 kHz) noises. Finally, we segment the received signals based on the period of a chirp signal for the next step, liquid feature extraction.

**Liquid Feature Extraction.** The liquid feature extraction module is mainly to extract both the amplitude frequency feature and transfer function feature from the liquid reflected signals. We then propose a series of methods to deal with the impact from the hardware heterogeneity, the smartphone positions, and the liquid volumes.

**Liquid Classification and Identification.** In this module, we develop a Siamese-network-based approach to reconstruct the extracted features to container-irrelevant features, and further utilize the transfer learning method to adapt the network for new liquids.

## 5 PROCESSING SENSING SIGNALS

In this section, we first introduce the preliminary of choosing the chirp signals as the transmitted signals in our system, and we then propose the detailed parameters of chirp signals.

### 5.1 Background of Transmitted Signal Design

First, we take the multi-frequency continuous wave of  $\sum_i^n A \cos(2\pi f_i t)$  as the transmitted signals, where  $A$  is the amplitude,  $f$  is the frequency of the acoustic signals, and we set  $n$  as seven. We conduct experiments on two liquids (i.e., Coca-Cola, Pepsi) to evaluate the discrimination of utilizing the multi-frequency continuous signals. We extract the amplitude information from the reflected signals of the two liquids. Figure 4 plots the amplitude values of the two liquids; we discover that the amplitude of the two liquids has a slight difference only at the fourth frequency point. We utilize the **dynamic time warping (DTW)** [25] algorithm to calculate the difference between the amplitude frequency profiles of the two liquids' reflected signals. The DTW distance of the above two curves is  $D_{dtw} = 13.5$ . We wonder if it is feasible to obtain a collection of amplitude information and use such a profile as the liquid feature, and whether the liquid feature will be more distinguishable as the number of spectral points increases. We increase the number of frequency points in transmitted signals from 7 to 300. The DTW distance of the below two curves is  $D_{dtw} = 337.26$ . It indicates that the difference between feature points of the two liquids is more obvious.

### 5.2 Transmitted Signal Design

In order to obtain richer frequency-domain characteristics for liquid identification, we generated a chirp signal with frequency sweeping from 10 kHz to 16 kHz. The duration of the chirp is 0.05 s, corresponding to 2,400 samples. A **pseudo-noise (PN)** preamble is added at the beginning of the transmitted signals for synchronization. Figure 5 illustrates the transmitted signal in the time domain and frequency domain.

The reason for such a design is twofold: On one hand, the sampling frequency of smartphones usually is 48 kHz, so the transmitted frequency can reach 24 kHz. However, our preliminary experiments indicated that the frequency response in the high-frequency range is very weak to sustain liquid distinction. As shown in Figure 6, the amplitude value of the two liquids is almost identical above 16 kHz, as the speaker and microphone of the commercial devices suppress acoustic signals above 16 kHz. The acoustic signals below 10 kHz are easily polluted by environmental noise. So

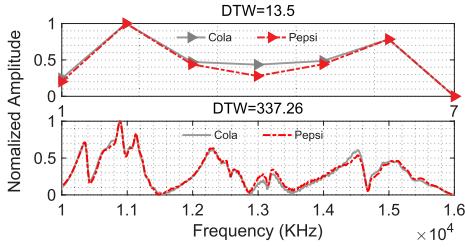


Fig. 4. DTW distance of the amplitude frequency profile between two liquids.

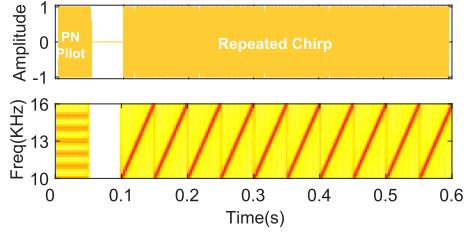


Fig. 5. Example of the transmitted signals between 10 kHz and 16 kHz.

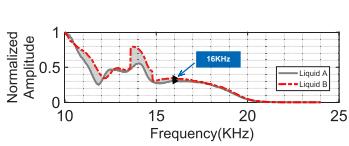


Fig. 6. Transmitted signals frequency and bandwidth selection.

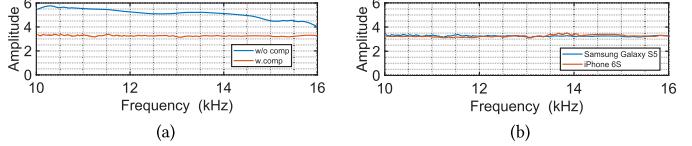


Fig. 7. Frequency response compensation.

we choose the frequency from 10 kHz to 16 kHz. This frequency range not only is much higher than the frequency range of the ambient noise but also has a relatively flat frequency response for the speaker on the smartphone. On the other hand, although a longer signal has a higher **signal-to-noise ratio (SNR)**, it also leads to more severe multi-path propagation effects. To balance this, we set the chirp period as 50 ms.

## 6 LIQUID FEATURE EXTRACTION

In this section, our goal is to extract a unique and stable liquid feature for liquid identification. We then extract the differential amplitude frequency and transfer function features to represent the liquids.

### 6.1 Differential Amplitude Frequency Feature Extraction

According to Section 3, we can calculate the differential amplitude-frequency feature to represent the liquid. However, it is still challenging to directly utilize this feature due to some key issues for liquid identification. Concretely, there are three types of interference: (1) hardware heterogeneity, (2) smartphone position, and (3) liquid volumes.

**6.1.1 Hardware Heterogeneity.** Hardware heterogeneity is mainly caused by the speaker of smartphones. Smartphones' speakers will strengthen or attenuate the signals at certain frequencies, which leads to a non-flat frequency response. Meanwhile, the speakers of different smartphones have different frequency responses. This will result in different smartphones acquiring inconsistent liquid features.

To solve the problem, we utilize the frequency response compensation method to eliminate the difference on different smartphones. Specifically, we first measured the frequency response of the speaker when the smartphone was placed in a multi-path-free environment, and we designed a digital filter whose frequency response is the reciprocal of the frequency response of the speaker. Then the transmitted signals that we designed passed through this filter before being emitted by the speaker. The result is shown in Figure 7(a); we discover that the smartphone (iPhone 6s) has a relatively flat frequency response after the frequency response compensation method. Then we

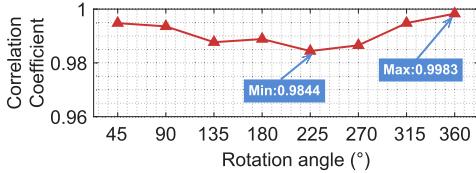


Fig. 8. Impact of the smartphone rotation angle on the feature profile.

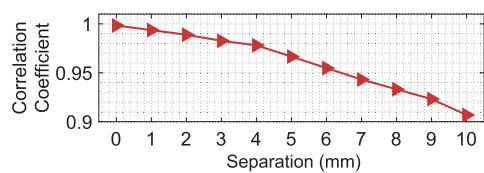


Fig. 9. Impact of smartphone position on feature profile.

measure the frequency response of the different smartphones; we discover that the different smartphones have different frequency responses, and two smartphones can have a similar frequency response through the frequency response compensation.

**6.1.2 Smartphone Position Selection.** Random changes in the smartphone position will cause variations in the amplitude-frequency feature. Fortunately, we discover that there is also a stable relationship between smartphone position and liquid feature. Specifically, we set the initial position of the smartphone with the top speaker and microphone facing the center of the container. We rotated the smartphone's angle with respect to the central point of the container from  $0^\circ$  to  $360^\circ$  in steps of  $45^\circ$ . We then calculated the correlation coefficients of the liquid feature from the different angles of the smartphone with  $0^\circ$ . The result is shown in Figure 8; the correlation coefficients were all close to 99%. It illustrates that the liquid feature exhibited highly consistent patterns when we kept the speaker in this position and rotated the smartphone randomly. Then, we shifted the smartphone forward from the initial position to 10 mm (in a step of 1 mm) to detect the feature changes caused by the smartphone position. Figure 9 plots the correlation coefficients of the different position features with the initial position feature. The result illustrates that when the position is offset from the initial position, it leads to a gradual decrease in the correlation coefficient of features. But we discover that when the physical offset is within 5 mm, the correlation coefficient is within 0.95. Therefore, the position of the smartphone does not need to be fixed or tightly controlled; it is tolerable to deviate from the center position by 5 mm forward and backward.

In summary, we set the measure position as the front speaker and microphone of the smartphone facing the center of the container. Meanwhile, when the speaker and microphone of the smartphone are vertical to the center of the container and relative to the shift within 5 mm of the center position, the amplitude-frequency profile changes caused by the position can be ignored.

**6.1.3 Different Liquid Volumes.** We discover that the changes in liquid volume lead to the variation in the reflected signal. Specifically, different volumes of liquid in the container cause changes in the signal propagation path, resulting in the same liquid having different amplitude-frequency features.

In order to tackle the problem, we discover that when the propagation path does not change, the amplitude attenuation of signals caused by the path is fixed. The difference in amplitude of the reflected signals is mainly caused by the difference in the liquids. Therefore, in order to obtain the amplitude change caused by the liquid, we first utilize the DASD method to eliminate the path-induced attenuation. We conduct a benchmark experiment to verify the feasibility of this method. We experiment with different volumes of sugar solutions (i.e., 100 mL, 150 mL, ..., and 500 mL). We take the 400 mL as the reference volume. We first calculate the correlation coefficient of the ASD and DASD profiles of different volumes with the 400 mL. The results are shown in Figure 10. The similarity of the ASD feature between different volumes is very low. The similarity of the DASD feature is improved relative to the ASD feature, but the similarity of the different volumes

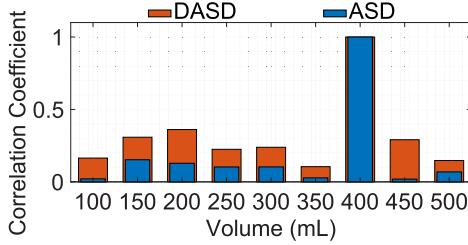


Fig. 10. Correlation coefficient of ASD and DASD between different volumes.

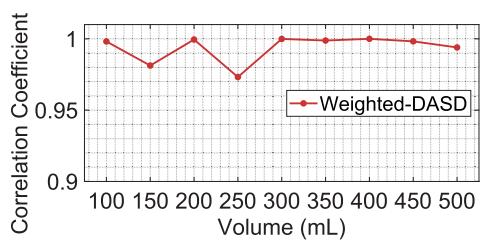


Fig. 11. Correlation coefficient of weighted DASD between different volumes.

is still less than 0.5. This illustrates that the liquid volumes can cause inconsistencies in the DASD feature.

To overcome the above issue, we introduce a weight-based method to help unify the amplitude difference between liquid volumes. The volume weight  $W_{dasd}$  is computed by

$$W_{dasd} = \left[ \frac{da_1^{vol}}{da_1^{ref}}, \dots, \frac{da_i^{vol}}{da_i^{ref}}, \dots, \frac{da_n^{vol}}{da_n^{ref}} \right], \quad (8)$$

where  $da_i^{vol}$  is the amplitude at frequency bin  $i$  in the magnitude spectrum of a certain volume of liquids, and  $n$  is the number of frequency points.

Before applying the corresponding volume weight vector  $W$ , we first need to acquire the volume of liquid. To solve the problem, we discover that calculating the liquid volume can be converted into calculating the height of the liquid. This value is relatively easy to calculate; it can be simplified to FMCW-based acoustic localization [18] such that the reflected signals are mixed with the transmitted signals and then passed through a low-pass filter. The mixed signals can be derived as

$$m(t) = \alpha \cos \left( 2\pi \left( \frac{B}{T} \tau_i t + f_0 \tau_i - \frac{B}{2T} \tau_i^2 \right) \right), \quad (9)$$

where  $f_p = \frac{B}{T} \tau_i$ , and  $f_p$  is determined by the first peak in the spectrum of the mixed signals. Thus, the height  $h_s$  of the smartphone to the liquid's surface can be derived as

$$h = \frac{f_p * c * T}{B}, \quad (10)$$

where  $c$  is the propagation speed of sound,  $T$  is the sweep time, and  $B$  is the signal's bandwidth. Therefore, based on the method, we also can get the height of the container, which is denoted as  $h_c$ . So the liquid's height  $h_l = h_c - h_s$ . As the liquid volume  $V_l$  increases, the  $h_l$  increases. Therefore, we can get the liquid volume  $V_l$  by mapping the  $h_l$ .

Therefore, by combining the method mentioned above ( $W_{dasd}$ ,  $h_l$ ), we eliminate the interference caused by different volumes. As shown in Figure 11, the weighted DASD profiles of different volumes are very similar, and the residuals of the curve fit are only 0.019.

## 6.2 Transfer Function-based Feature Extraction

Although the DASD feature can reflect the difference between different types of liquids, we discover that the feature is not enough in distinguishing the similar liquids, such as the same solution with slightly different concentrations (low-concentration salt solution). We take the three low concentration solutions (1%, 2%, and 3%) as an example; we then plot the DASD feature extracted from the three liquids in Figure 12. We can discover that it is difficult to distinguish the three liquids only utilizing the DASD feature.

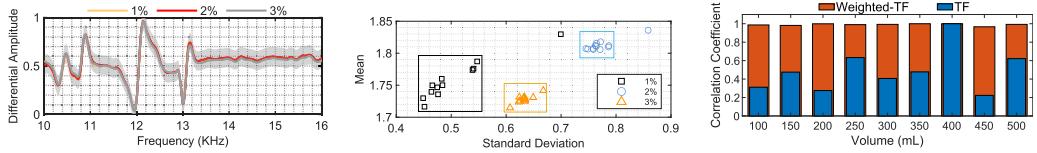


Fig. 12. DASD feature for different concentrations of the same solution. Fig. 13. TF feature for different concentrations of the same solution. Fig. 14. Weighted TF feature across different volumes of the same solution.

In order to improve the usability of Akte-Liquid, we discover that the TF method to measure the material's absorption coefficient has high accuracy in sound measurement [3]. According to Equation (7), the transfer function is related to the reflection coefficient of the material. Specifically, there is a single-input (one speaker) and single-output (one microphone) system in Akte-Liquid. It is different from the sound measurement, where the two microphones are placed on different reflection positions. In Akte-Liquid, we calculate the transfer function ratio between the two liquids.

The transfer function can be calculated as

$$H(f) = \frac{P_{yx}(f)}{P_{xx}(f)}, \quad (11)$$

where  $P_{yx}$  is the cross-power spectral density between the recorded and transmitted signals.  $P_{xx}$  is the self-power spectral density of the transmitted signal. We observe that when the information from the same volumes of two liquids is exploited, a high identification accuracy can be obtained compared to only one liquid's information utilized. We therefore define the ratio-liquid transfer function as

$$H_{ratio}(f) = \frac{H_{test}(f)}{H_{ref}(f)}, \quad (12)$$

with  $H_{test}(f)$  and  $H_{ref}(f)$  representing the transfer function of the test liquid and reference liquid, respectively.

Then, we obtain the feature set  $H = [H(f_1), H(f_2), \dots, H(f_k)]$ , where  $H(f_k)$  represents the transfer function of the  $k$ th frequency bin. In order to illustrate the discriminative power of the TF feature, we calculate the mean value and standard deviation of the extracted TF feature for each liquid. The result is illustrated in Figure 13. This shows that the points from the same concentration exhibit a cluster and are obviously different from each other.

Then we evaluate the robustness of the TF feature of the different volumes. We take the different volumes of salt solutions (i.e., 100 mL, 150 mL, ..., and 500 mL) as an example and calculate the correlation coefficients of the TF between different volumes and the reference volume (400 mL). The result is shown in Figure 14; we discover that the similarity of the different volumes is still poor, all below 0.6. This illustrates that the liquid volume changes lead to feature inconsistency. In order to solve the problem, we weight the TF feature from different volumes and unify the amplitude to eliminate the volume difference. We utilize 400 mL as the reference liquid. The volume weight vector  $V_{tf}$  is defined as

$$V_{tf} = \left[ \frac{tf_1^{vol}}{tf_1^{ref}}, \dots, \frac{tf_i^{vol}}{tf_i^{ref}}, \dots, \frac{tf_n^{vol}}{tf_n^{ref}} \right], \quad (13)$$

where  $tf_i^{vol}$  is the amplitude at frequency bin  $i$  in the magnitude spectrum of a certain volume liquid, and  $n$  is the number of frequency points. As shown in Figure 14, the weighted TF profiles of different volumes are very similar.

Finally, due to the temporal stability of the TF being not as good as the DASD feature, the variance of the transfer function value of the same liquid over the different time periods varies greatly. In order to keep the discrimination among different liquids as well as maintain feature consistency in each liquid, Akte-Liquid utilizes the DASD-based feature and TF-based feature to represent liquids.

### 6.3 The Speaker Volume Control on Smartphones

The volume of transmitted acoustic signals plays a critical role in extracting valid liquid features for liquid identification. However, the signals are more easily distorted when the volume of the transmitted acoustic signals is set to the maximum (i.e., 100%), due to hardware imperfections in the speaker. Moreover, the transmitted acoustic at the maximum volume will make the sensing process more uncomfortable to users, since our system utilizes audible frequency ranges. In contrast, if the acoustic is transmitted at 1% of the maximum volume, the reflections are too weak to be picked up by the smartphone microphones. Based on our preliminary experiments, setting the phone volume at 60% is optimal for liquid sensing. Even though this setting varies from one smartphone model to another, only a one-time calibration is required to determine the optimal setting.

## 7 BUILDING THE LIQUID TESTING MODEL

In this section, we first introduce how to reconstruct extracted features to container-irrelevant features, and we then build a liquid identification model utilizing the reconstructed features.

### 7.1 Liquid Feature Reconstruction

Because the received signals are mainly from the liquid and container reflections, due to different materials having different reflection coefficients, the containers of different materials will also lead to the reflected signals' variation. So the extracted liquid features from reflected signals contain both liquid feature and container feature.

**7.1.1 Container-irrelevant Feature Reconstruction.** To build a convenient liquid identification system, we need to eliminate the influence of containers, i.e., reconstruct features to container-irrelevant features. In order to solve the problem, we utilize the CNN [30] to extract the features. CNN has the advantage of modeling spatial context information in the 2-D space, and has achieved great success in the field of image classification. Thus, we utilize the CNN to extract the frequency domain features from the inputs. In Akte-Liquid, we employ a Siamese-network-based model with a specific training sample selection strategy to reconstruct the extracted features to container-irrelevant features. The Siamese network is a novel approach to recognize activities or object tracking [24]. The basic principle of this network is using a pair of neural networks with the same architecture and weights to compute a distance metric for two input samples. The particular structure enables customized feature extraction by the selection of sample pairs during training.

Concretely, we apply a training sample selection scheme, as shown in Figure 15. If the liquid labels are the same, we select the sample pairs from different containers as the training samples, so the model can learn to ignore the container differences for samples from the same liquid. If the liquid labels are different, we choose the sample pairs from the same container, and thus the model can learn to distinguish samples from different liquids not based on the container difference. Therefore, by selecting the proper training samples as the input to the Siamese network, we can reconstruct container-irrelevant features.

As shown in Figure 16, we illustrate the architecture of the Siamese network. Considering a pair of features as the input, we reconstruct the container-irrelevant feature representation from the extracted reflected signal features through two identical sub-networks. Then, we calculate the

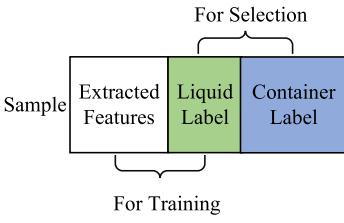


Fig. 15. Sample selection.

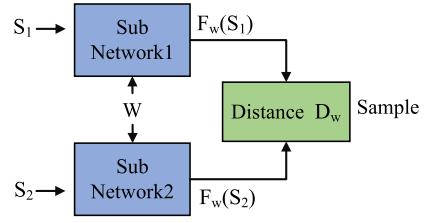


Fig. 16. Illustration of Siamese network.

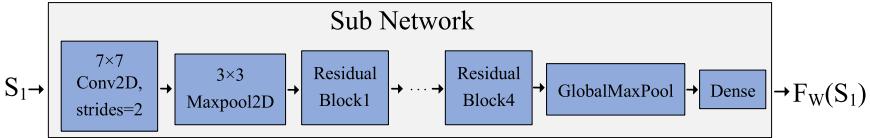


Fig. 17. Sub-network design in Akte-Liquid.

distance  $D_w$  of the reconstructed features  $F_w(S_1)$  and  $F_w(S_2)$ . We denote it as  $D_w$

$$D_w(s_1, S_2) = ||F_w(S_1) - F_w(S_2)||. \quad (14)$$

So if the two input samples are the same liquid but from different containers, we make the extracted features have the minimal  $D_w$ . Otherwise, we maximize the distance  $D_w$  if extracted features are from the different liquids.

The structure of the sub-network is designed with four residual blocks, and each residual block contains two convolutional layers, and the shortcut connection and pooling are adopted between the residual blocks, as shown in Figure 17. We denote the parameters of the sub-network as  $W$ ; the loss function of sub-network we denoted as  $L(W)$ :

$$L(W) = \sum_{i=1}^N Y(D_W^i)^2 + (1 - Y)\max(M - D_W^i, 0)^2, \quad (15)$$

where  $Y$  indicates whether two input samples are related to the same liquid; i.e., for the same liquid  $Y = 1$ , and otherwise,  $Y = 0$ .  $D_W^i$  is the Euclidean distance of the  $i^{th}$  input samples,  $M$  is the margin that represents the decreased interval. From Equation (15), we minimize the loss function  $L(w)$  of the designed Siamese network by minimizing the distance between samples of the same liquid and maximizing the distance between samples of different liquids.

**7.1.2 The Input of Model.** The model input is the jointing of the DASD and the transfer function features for each sub-network. The extracted two-dimensional features matrixes are transformed into an image as the input of the model, because CNN is known to work well in image identification. The size of the DASD feature  $W_{dasd}$  is  $1*300$ ; the size of the transfer function feature  $W_{tf}$  is  $1*300$ . We perform feature fusion by concatenating them as  $W = W_{dasd} \oplus W_{tf}$ .

**7.1.3 The Extracted Liquid Feature by Model.** Finally, in order to verify the effectiveness of the extracted feature, we output the intermediate process of the training model.

The results are shown in Figure 18. We can find that the features extracted by CNN can maximize the distance between classes and minimize the distance within classes. Moreover, the extracted liquid features can eliminate the influence of the container and only focus on the features of the liquids.

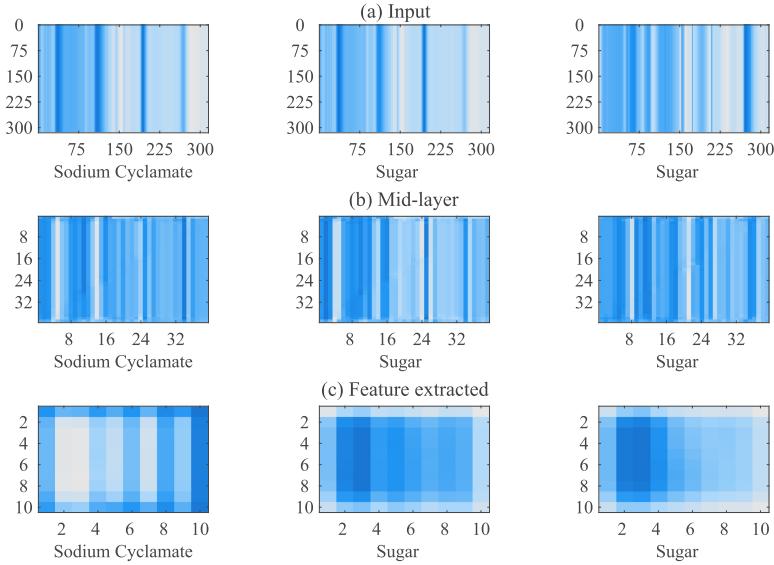


Fig. 18. Visualization of the mid-process output features of the model.

We take two solutions (sodium cyclamate, sugar solution) as an example. Specifically, Figure 18(a) presents the extracted DASD feature and transfer function feature of liquids, which are the input of the model. The first and second diagrams are the liquid features of two different liquids obtained in the same container; the third diagram is the liquid features of the same sugar solution in another container. From Figure 18(a), we can find that due to the influence of the container, the different liquids in the same container have higher similar liquid features than the same liquids in different containers. It easily leads to recognition errors. Figure 18(b) is the output liquid features by the model's middle layer; its size is compressed to 39\*39 by the model. We can find that the difference between the second and third diagrams is narrowing. Figure 18(c) is the output liquid features by the last feature extraction layer; its size is compressed to 10\*10 by the model. The first and second pictures show that the features of two liquids in the same container are completely different. The second and third pictures show that the features of the same liquid are highly similar even in the different containers.

In order to verify the validity of the proposed model, we visualize the reconstructed features from four containers and nine types of liquids with t-SNE in Figure 19. We can discover that the same liquid from different containers has a coherent feature distribution, while different liquids have more distinct feature distributions. The result demonstrates that the designed Siamese network model can effectively reconstruct container-irrelevant features.

## 7.2 Transfer Learning to New Liquids

To reduce the cost of re-training for new liquid samples while keeping the generation capability, we refer to the idea of transfer learning. Some common layers of a well-trained container can be directly frozen and transferred to new liquid samples. This design significantly reduces the number of parameters that need to be learned for new tasks and reduces the size of the dataset and training time required to achieve high accuracy. Figure 20 shows the structure of transfer learning.

Specifically, the training for new liquids is twofold process. In the first training stage, the outputted features of the source domain, denoted as  $F_{source}$ , are utilized as the training goal of the

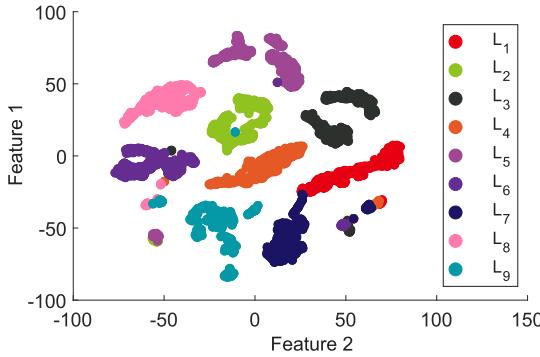


Fig. 19. t-SNE visualization of container-irrelevant features for nine types of liquids.

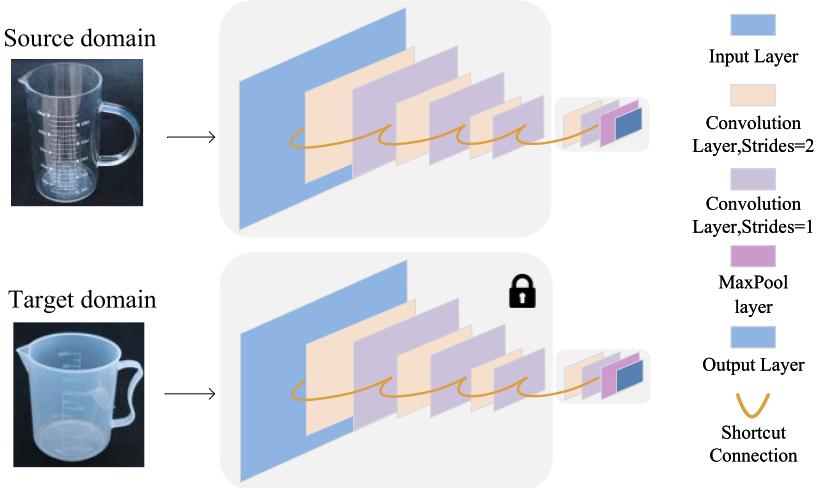


Fig. 20. Transfer learning model structure of our CNN model.

target domain's output, denoted as  $F_{target}$ . The loss function of the source domain and target domain  $L_1$  is calculated as the cross-entropy of  $F_{source}$  and  $F_{target}$ :

$$L_1 = H_{F_{source}}(F_{target}) = - \sum_i F_{source} \times \log(F_{target}). \quad (16)$$

Then, the target domain refines the parameters of the network with its own training data from new liquid samples, and the loss function is calculated as Equation (15). Based on the above steps, we are able to extract the container-irrelevant features of new samples and also reduce the training effect for new liquids and keep a good performance.

## 8 EVALUATION

In this section, we evaluate the overall performance of Akte-Liquid in liquid identification. We first introduce the implementation of Akte-Liquid. We then evaluate the ability of our system from four aspect applications. Finally, we verify the effectiveness and robustness of Akte-Liquid from each sub-component's contribution to the system, and the impact of different parameters (e.g., environmental changes, container changes, the different devices).

## 8.1 Implementation

**Experiment Implementation.** We implement our system on an iPhone 6s based on the open-source acoustic sensing framework LibAS [21]. In the test environment, the user just holds onto the smartphone in their hands, as shown in Figure 21. The smartphone is placed on top of a glass container with the speaker and microphone facing the liquid surface vertically.

**Software.** The algorithms in Akte-Liquid are implemented in MATLAB and Tensorflow. We implement MATLAB to control the generated acoustic signals and receive the reflected signals and then design a series of algorithms to extract the liquid features.

**Data Collection.** We recruit 15 volunteers (7 females, 8 males) to conduct experiments for evaluating the performance of Akte-Liquid. We collect the liquid samples from 8 containers and 10 different environments. In the training phase, we first collect different volumes of reference liquid data and different kinds of test liquids for each new container. For each measurement, Akte-Liquid collects for 2 minutes and is repeated five times in one container and environment. The data is stored in the database in advance. Meanwhile, the liquid dataset can be expanded as users need. In the real-time testing phase, we only send 1-second signals for the test liquid.

In dataset split, we randomly split our dataset into 10 parts. We select five of them as training, three of them as testing, and two as verifying.

**Evaluation Metrics.** We evaluate Akte-Liquid along with five metrics:

- (1) Correlation Coefficient: Describing the liquid feature similarity between different liquids, it can be represented as  $\text{corr}_{xy} = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ , where  $\text{Cov}(\cdot)$  and  $\sigma(\cdot)$  calculate the covariance and standard deviation, respectively. The scope is  $[0, 1]$ ; the closer the correlation coefficient is to 1, the greater the correction between the two liquids.
- (2) Accuracy:  $\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$ , which represents the fraction of samples that are correctly predicted. TP represents the number of true positives (i.e., the number of items correctly labeled as belonging to the positive class), TN is the true negative, FP represents the number of items incorrectly labeled as belonging to the positive class, and FN is the false negative.
- (3) Precision:  $\text{Precision} = \frac{TP}{TP+FP}$ , which is the fraction of relevant instances among the retrieved instances.
- (4) Recall:  $\text{Recall} = \frac{TP}{TP+FN}$ , which is the fraction of relevant instances that were retrieved.
- (5) F1 Score:  $F1 = 2 * \frac{P * R}{P + R}$ , which is the harmonic average of the precision and recall.

## 8.2 Different Applications

In this section, we describe the possible applications of Akte-Liquid in liquid identification from daily life, water sources, health, and food safety.

**8.2.1 Liquid Identification Performance.** First, we verify the effectiveness of Akte-Liquid to identify different types of liquids. We prepare 20 different kinds of liquids, which are Coca-Cola, Pepsi, Diet Pepsi, two kinds of juice, whole milk, skim milk, three concentrations (1%, 5%, 10% of alcohol, salt, sucrose solution), and three kinds of coffee. For data collection, we use the standard glass graduate whose volume is 500 mL as a testing container, placing the phone on top of the container.

Figure 22 shows the result as a confusion matrix. The different rows represent the actual liquids, and the different columns represent predicted results. From Figure 22, we observe that Akte-Liquid can classify across the different liquids to reach an accuracy of more than 97.9%. The detailed performances are listed in Table 1. We can find the precision is more than 98.1%, and the recall and the F1 score are more than 97%.



Fig. 21. Experimental implementation.

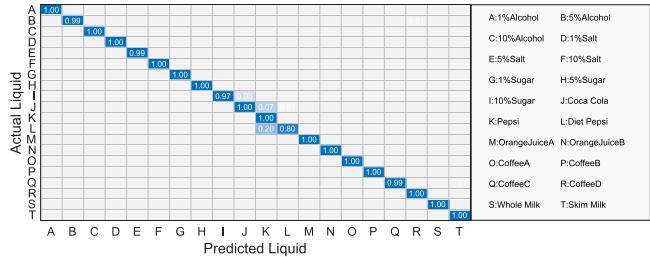


Fig. 22. Liquid classification results.

Table 1. Overall Performance of Identifying Different Types of Liquids

	Mean	Median	Standard Deviation
Accuracy	0.9796	0.9801	0.0024
Recall	0.9796	0.9801	0.0024
Precision	0.9815	0.9811	0.0011
F1 Score	0.9805	0.9801	0.0014

**8.2.2 Detection of Different Water Sources.** In this application, we are interested in evaluating the ability of our system in detecting different water sources. We compare four different water sources: (1) Mineral water (which is water from a mineral spring that contains various minerals, such as salts and sulfur compounds); (2) Spring water (which is self-purified natural drinking water from unpolluted mountainous areas); (3) Purified water (which has been mechanically filtered or processed to remove impurities and make it suitable for use); and (4) Tap water (distinguishes it from other main types of freshwater that may be available). In this experiment, we prepare three containers and each liquid with five volumes to evaluate the performance of our system. Figure 23 plots the result of detecting different water. We discover that our system can successfully distinguish between water from different containers and with different volumes.

**8.2.3 Detection Protein Concentration in Artificial Urine.** We hope Akte-Liquid can provide an easy and convenient way to help patients detect urine protein at home. Because it is hard to control urine substances, we utilize artificial urine, which simulates the components and pH of urine to replace real urine. Artificial urine is also the mixture solution, which is mainly composed of a series of substances such as calcium chloride, sodium chloride, magnesium chloride, sodium sulfate, potassium chloride, and so forth.

In this experiment, we utilize human serum albumin instead of urine protein; they both belong to albumin and share similar physical properties. In order to evaluate the predictability of solution protein concentrations, we run 10 independent experiments with each protein solution. In Figure 24, the results show that our system can detect different concentrations of protein even if there is a small difference between them. From Figure 24, it can be seen that as the protein concentration increases, the identification accuracy increases gradually. Although we only achieve about 70.76% accuracy under 0.5 mg/100 mL, we can achieve more than 92.3% accuracy under 1 mg/100 mL. Note that when the protein concentration in the uric is higher than 3 mg/100 mL, it indicates the patient has the complication of microalbuminuria [27]. Therefore, these results indicate the probability of utilizing a smartphone to detect kidney disease.

	A	B	C	D
A	0.94	0.02	0.04	
B	0.02	0.92	0.06	
C	0.02	0.02	0.96	
D		0.02	0.04	0.94
Predicted Liquid	A	B	C	D

A: Mineral Water  
B: Spring Water  
C: Purified Water  
D: Tap Water

Fig. 23. Detection of different water sources.

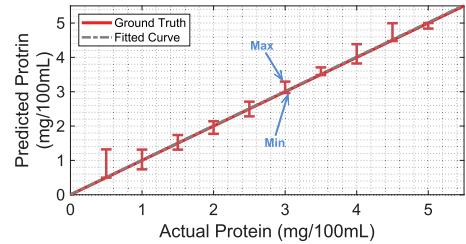


Fig. 24. Detection protein concentration in artificial urine.

Table 2. Recipes of Food Additive Experimentation

Food Additive	Recipe
Water	400 mL Water (no additive)
Citric Acid	0.2 g, 0.5 g, 1 g with 400 mL water, respectively
Citric Acid and Sugar	0.2 g, 0.5 g, 1 g with 400 mL water and 3 g sugar, respectively
Potassium Sorbate	0.2 g, 0.5 g, 1 g with 400 mL water, respectively
Potassium Sorbate and Sugar	0.2 g, 0.5 g, 1 g with 400 mL water and 3 g sugar, respectively
Aspartame	0.2 g, 0.5 g, 1 g with 400 mL water, respectively
Aspartame and Sugar	0.2 g, 0.5 g, 1 g with 400 mL water and 3 g sugar, respectively
Sodium Cyclamate	0.2 g, 0.5 g, 1 g with 400 mL water, respectively
Sodium Cyclamate and Sugar	0.2 g, 0.5 g, 1 g with 400 mL water and 3 g sugar, respectively

**8.2.4 Characterization of the Food Additives in Beverages.** Food additives are also a food safety problem that people often face [22]. Therefore, we would like to employ our system to detect the content of food additives in beverages. In reality, food additives are used in mixed liquids, or with other interfering agents in liquids. Therefore, in order to simulate the real scene, we use sucrose as an interfering agent in liquid. We measure the above four food additives separately in the different solutions. Specifically, as shown in Table 2, we prepared two kinds of experiments; one trial is 400 mL water with, respectively, 0.2, 0.5, 1, 2, and 3 grams of citric acid or potassium sorbate without interfering agents; we take distilled water as a baseline class. Another trial is the above recipe but with 3 g sugar in each solution. Now we test the capability of identifying low-concentration food additives in the beverage.

Figures 25 and 26 demonstrate the scatter graphs of liquid features for food additive classification. Figures 27 and 28 show the results of the other two food additives.

From the figure, we can discover that with or without sugar, the extracted liquid feature shows obvious differences. Furthermore, we notice that all the classes are well separated, and the same food additive concentration with or without interfering agents also has distinct classes.

Finally, we evaluate the accuracy of our system to detect the four food additives in the water. The result is shown in Table 3; we can see that the average precision is close to 95%. The results illustrate the capacity of Akte-Liquid to identify the food additive, and it is possible to employ Akte-Liquid to learn the food additive content of a mixed solution.

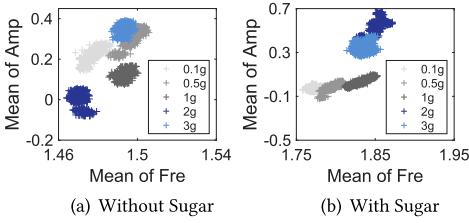


Fig. 25. Food additive in sorbate solution.

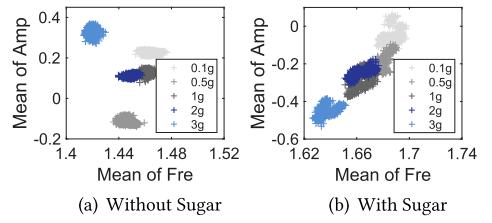


Fig. 26. Food additive in citric solution.

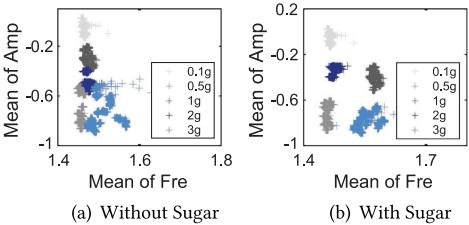


Fig. 27. Food additive in aspartame solution.

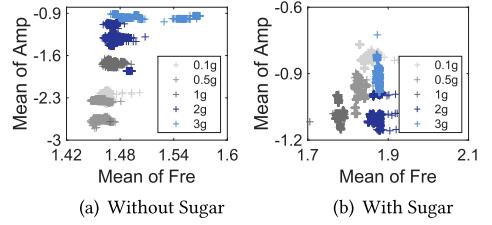


Fig. 28. Food additive in sodium cyclamate solution.

Table 3. Precision and Recall of Detecting a Particular Food Additive

	Sorbate	Citric	Aspartame	Cyclamate	Average
Precision	85.37%	98.92%	94.60%	97.70%	94.14%
Recall	92.32%	95.00%	96.82%	88.46%	93.28%

### 8.3 Robustness and Generalization Evaluation

We provide several micro-benchmarks of Akte-Liquid, including the method validation and the running time of each component in Akte-Liquid. Then, we evaluate several factors that impact the performance of our system, such as the containers, the training set size, environment noise, the smartphones, and so on.

**8.3.1 Method Validation.** We would like to quantify the accuracy gains arising from each of Akte-Liquid's sub-components. We utilize the data samples from the different environments and containers, which include the different types of liquids, different concentrations, and volumes of the same liquid. Then, we compared Akte-Liquid with the other ensemble learning methods (Random Forest and Adaboost) in four different inputs: (1) original received signals, (2) transfer-function-based feature, (3) DASD-based feature, and (4) DASD combined with transfer function feature. We set the depth of the tree as 7 for the RF method, and we set the number of the estimator as 100 for the Adaboost method.

Figure 29 plots the accuracy and F1 score of the above approaches. The result shows that when we apply a common ensemble learning to classify liquids based on raw data, the accuracy is below 70%. Additionally, only using the TF feature or DASD feature, the identification accuracies based on the RF method and Adaboost are both lower than our method. Meanwhile, Akte-Liquid has the highest performance over 90%. Finally, the results also indicate that the DASD-based feature, transfer-function-based feature, and liquid identification model contributes to all system performance.

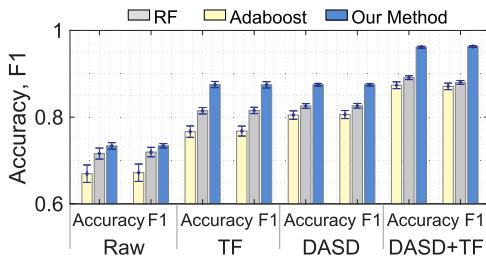


Fig. 29. Accuracy and F1 measure under different features and methods.

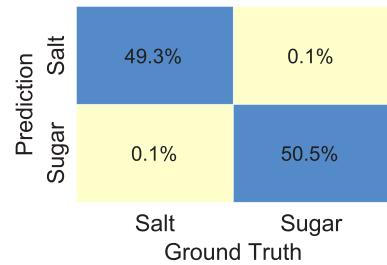


Fig. 30. Performance across different volumes.

Table 4. Running Time of Akte-Liquid

Device	ASD	DASD	Transfer Function	CNN	Total
Galaxy S5	0.005 s	0.009 s	0.418 s	1.781 s	2.215 s
Galaxy S6	0.005 s	0.01 s	0.5 s	1.813 s	2.328 s

**8.3.2 Running Time Evaluation of Each Component.** For evaluating the running time performance of Akte-Liquid, we test the running time for each component in Akte-Liquid on two types of smartphones, i.e., Samsung Galaxy S5 and Samsung Galaxy S6. Table 4 shows the results of detection efficiency evaluation. The total running time is less than 3 s, which shows that Akte-Liquid can promptly obtain results without any noticeable latency.

**8.3.3 Impact of Liquid Volume.** In this experiment, we discuss the impact of liquid volume on the accuracy of liquid identification. We measured the different volumes of salt solutions and sugar solutions (i.e., 100 mL, 150 mL, . . . , 500 mL). Figure 30 shows the results after eliminating the volume variation. We can find that the different volumes of the same liquid are grouped together.

**8.3.4 Impact of the Smartphone Position Relative to Container.** In this experiment, we study the impact of smartphone position relative to the container for Akte-Liquid by placing the smartphone on eight different positions (start 1 cm from the edge of the container and move to 8 cm). We experimented with two different liquids.

Figure 31 shows the confusion matrix of identification results. (a) is the without the DASD method, directly utilizing the ASD feature to classify the two different liquids with different smartphone positions. We can find that the results of clustering are very poor, with only 50% accuracy. (b) is with the DASD method, directly utilizing the DASD feature to classify the two different liquids with different smartphone positions. We can see that the clustering result increased to 75%, which illustrated that the DASD method can reduce the impact of the smartphone positions. This may be due to the fact that when the phone is close to the container wall, the received signals are more reflected by the container wall. So we set the measure position at the front speaker and microphone of the smartphone in the center of the container.

**8.3.5 Impact of Different New Containers.** We evaluate the accuracy of our system for liquid identification on new containers. We train the model on six different materials and size containers, which have different shapes, heights, and opening diameters, as shown in Table 5. Then we test the accuracy on the two new different containers, where one is a glass container and the other is a steel container. We experiment with sugar solution, salt solution, and sugar-like solution on each container. We prepare three different concentrations for each solution. So there are nine liquid samples

Table 5. Different Containers

<b>Container Picture</b>				
<b>Dimension</b>	Top:80mm Bottom:80mm Height:150mm Material:glass	Top:95mm Bottom:95mm Height:115mm Material:plastic	Top:85mm Bottom:65mm Height:120mm Material:glass	Top:85mm Bottom:85mm Height:101mm Material:china
<b>Container Picture</b>				
<b>Dimension</b>	Top:75mm Bottom:53mm Height:87mm Material:paper	Top:78mm Bottom:64mm Height:105mm Material:steel	Top:81mm Bottom:54mm Height:119mm Material:rubber	Top:92mm Bottom:63mm Height:109mm Material:plastic

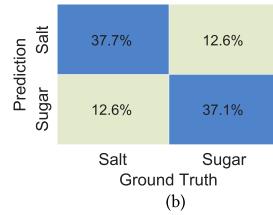
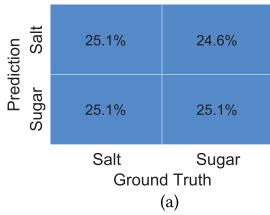
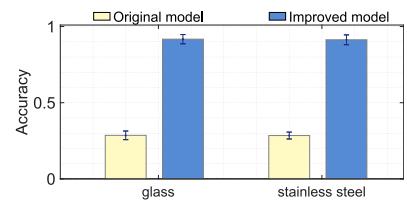


Fig. 31. Performance across different positions.



in total. Figure 32 shows the identification accuracy of each container. Compared to our previous model, the network trained all the samples, and the reconstructed features contain both container features and liquid features. The accuracy is low when testing on a new container. It can be seen that there is no obvious difference in identification accuracy on different containers utilizing the improved model. The result indicates that by extracting container-irrelevant features, our system is not sensitive to the changes of containers and can reach a good performance on new containers.

**8.3.6 Impact of Different Environments.** We trained our liquid identification model in four environments and tested the performance in six new environments. The different environments had different sound pressure levels. In each new environment, we collect three similar liquids, sugar solution, salt solution, and sugar-like solution. For each solution, we prepare three different concentrations. So there are nine types of liquids in total. We collect each liquid sample in three different containers. The result is shown in Figure 33. We can discover that Akte-Liquid is robust to different environments. We can obtain a good identification performance in unseen environments.

**8.3.7 Impact of Number of Containers Used for Training.** Due to the training sample selection scheme, the number of containers used for training is important for our system to extract container-independent features. A larger number of samples of different containers will improve the performance of our system, but also will lead to tedious work. Hence, we evaluate our system performance under different numbers of containers.

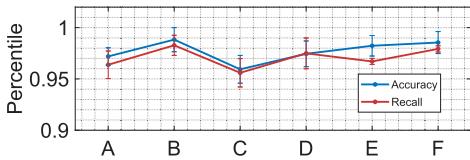


Fig. 33. Performance with different environments (A: supermarket, B: cafe, C: bar, D: restaurant, E: office, F: library).

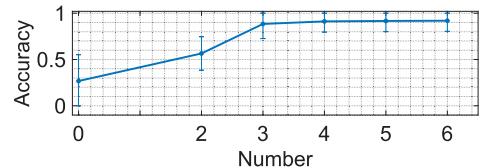


Fig. 34. Impact of number of containers used for training.

Table 6. Accuracy of Different Smartphones

Device	iPhone 8	Samsung Galaxy S6
Accuracy	92.28%	93.35%
Recall	92.44%	93.37%

In this experiment, we utilize the 15 kinds of liquids (nine types of salt, sugar, and sugar-like solutions; four types of water samples; and two similar beverages, Coke and Pepsi). The result is shown in Figure 34; we can discover that as the number of containers increases from 0 to 6, the accuracy first increases and then remains stable when the number reached about 4.

**8.3.8 Impact of Different Smartphones.** We utilize the trained model from three different smartphones (iPhone 6s, iPhone 11, and Samsung Galaxy S5) to test the performance on two untrained smartphones (iPhone 8 and Samsung Galaxy S6). We utilize the liquid dataset from the above experiment. The result is shown in Table 6; we can discover that our system also has high performance on different devices.

## 9 DISCUSSIONS

**Limitations on the Containers.** Akte-Liquid has certain requirements on the container type. First of all, the opening of the container needs to be relatively large. If the opening of the container is too small (e.g., less than 30 mm), it will be difficult for the smartphone to receive the liquid-surface-reflected signals. This is due to the observation that a smaller container will cause the multi-path propagation effects in the container to be more significant, and liquid-surface-reflected signals will be flooded by the multi-path signals.

**Limitations on the Position of the Smartphone.** The position of the smartphone will cause identification errors. We should place the speaker and microphone of the smartphone parallel to the liquid surface for each measurement. Moreover, we should keep the smartphone in the center of the container as much as possible. Placing the smartphone's speaker and microphone too close to the container wall can have an impact on identification accuracy.

**Requirements of Smartphones.** In our experiments, we implement our system using the front speaker and microphone of the smartphone. Hence, it is important to choose a smartphone that has its speaker and microphone on the same side or facing the same direction.

**Unseen Liquid Identification.** Akte-Liquid evaluates some partial applications in liquid identification. Moreover, Akte-Liquid is based on the deep learning architecture that enables high precision in materials that have already been learned. It is challenging to extend our system to identify unseen liquid without training. It is an interesting research direction in our future work.

## 10 CONCLUSION

In this article, we propose Akte-Liquid, which utilizes the speaker and microphone of the smartphone to transmit the acoustic signals and receive the liquid's reflected signals to identify liquids. Our experimental results have demonstrated that Akte-Liquid is able to distinguish different kinds of liquids, as well as robustness across different containers and different environments.

## REFERENCES

- [1] L. Bertinetto, J. Valmadre, Joo F. Henriques, A. Vedaldi, and Phs Torr. 2016. *Fully-convolutional Siamese Networks for Object Tracking*. Springer, Cham.
- [2] Anurag Chowdhury and Arun Ross. 2019. Fusing MFCC and LPC features using 1D triplet CNN for speaker recognition in severely degraded audio signals. *IEEE Transactions on Information Forensics and Security* 15 (2019), 1616–1629.
- [3] J. Y. Chung and D. A. Blaser. 1980. Transfer function method of measuring in-duct acoustic properties. II. Experiment. *The Journal of the Acoustical Society of America* 68, 3 (1980), 914–921.
- [4] Ashutosh Dhekne, Mahanth Gowda, Yixuan Zhao, Haitham Hassanieh, and Romit Roy Choudhury. 2018. Liquid: A wireless liquid identifier. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 442–454.
- [5] David I. Ellis, Victoria L. Brewster, Warwick B. Dunn, J. William Allwood, Alexander P. Golovanov, and Royston Goodacre. 2012. Fingerprinting food: Current technologies for the detection of food adulteration and contamination. *Chemical Society Reviews* 41, 17 (2012), 5706–5727.
- [6] Chao Feng, Jie Xiong, Liqiong Chang, Fuwei Wang, Ju Wang, and Dingyi Fang. 2021. RF-Identity: Non-intrusive person identification based on commodity RFID devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–23.
- [7] Yang Gao, Wei Wang, Vir V. Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using ear canal echo for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
- [8] Unsoo Ha, Junshan Leng, Alaa Khaddaj, and Fadel Adib. 2020. Food and liquid sensing in practical environments using RFIDs. In *Proceedings of the 17th USENIX Symposium on Networked Systems Design and Implementation (NSDI'20)*.
- [9] Unsoo Ha, Yunfei Ma, Zexuan Zhong, Tzu-Ming Hsu, and Fadel Adib. 2018. Learning food quality and safety from wireless stickers. In *Proceedings of the 17th ACM Workshop on Hot Topics in Networks*. 106–112.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.
- [11] Yongzhi Huang, Kaixin Chen, Yandao Huang, Lu Wang, and Kaishun Wu. 2021. Vi-liquid: Unknown liquid identification with your smartphone vibration. In *MobiCom*. 174–187.
- [12] ISO. 1998. Acoustics-determination of sound absorption coefficient and impedance in impedance tubes-Part 2: Transfer-function method. (1998).
- [13] Gordon S. Kino. 1987. *Acoustic Waves: Devices, Imaging, and Analog Signal Processing*. Vol. 107. Prentice-Hall, Englewood Cliffs, NJ.
- [14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. 1097–1105.
- [15] Y. Lecun and L. Bottou. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 11 (1998), 2278–2324.
- [16] Yantao Li, Hailong Hu, Zhangqian Zhu, and Gang Zhou. 2020. SCANet: Sensor-based continuous authentication with two-stream convolutional neural networks. *ACM Transactions on Sensor Networks* 16, 3 (2020), 1–27.
- [17] J. Liu, D. Li, L. Wang, and J. Xiong. 2021. BlinkListener: “Listen” to your eye blink using your smartphone. *Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–27.
- [18] Wenguang Mao, Jian He, and Lili Qiu. 2016. CAT: High-precision acoustic motion tracking. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 69–81.
- [19] Tauhidur Rahman, Alexander T. Adams, Perry Schein, Aadhar Jain, David Erickson, and Tanzeem Choudhury. 2016. Nutrilyzer: A mobile system for characterizing liquid food with photoacoustic effect. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. 123–136.
- [20] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [21] Yu-Chih Tung, Duc Bui, and Kang G. Shin. 2018. Cross-platform support for rapid development of mobile acoustic sensing applications. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 455–467.

- [22] U.S. FOOD and DRUG ADMINISTRATION. 2004. Overview of Food Ingredients, Additives, Colors. <https://www.fda.gov/food/food-ingredients-packaging/overview-food-ingredients-additives-colors/>.
- [23] Hideo Utsuno, Toshimitsu Tanaka, Takeshi Fujikawa, and A. F. Seybert. 1989. Transfer function method for measuring characteristic impedance and propagation constant of porous materials. *The Journal of the Acoustical Society of America* 86, 2 (1989), 637–643.
- [24] R. R. Varior, M. Haloi, and G. Wang. 2016. Gated siamese convolutional neural network architecture for human re-identification. *European Conference on Computer Vision*. (2016).
- [25] Jue Wang and Dina Katabi. 2013. Dude, where's my card?: RFID positioning that works with multipath and non-line of sight. *Computer Communication Review* 43, 4 (2013), 51–62.
- [26] Ju Wang, Jie Xiong, Xiaojiang Chen, Hongbo Jiang, Rajesh Krishna Balan, and Dingyi Fang. 2017. TagScan: Simultaneous target imaging and material identification with commodity RFID devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. 288–300.
- [27] M. R. Weir. 2007. Microalbuminuria and cardiovascular disease. *Clinical Journal of the American Society of Nephrology* 2, 3 (2007), 581–590.
- [28] Binbin Xie, Jie Xiong, Xiaojiang Chen, Eugene Chai, Liyao Li, Zhanyong Tang, and Dingyi Fang. 2019. Tagtag: Material sensing with commodity RFID. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*. 338–350.
- [29] Shichao Yue and Dina Katabi. 2019. Liquid testing with your smartphone. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 275–286.
- [30] Bin Zhang, Changqin Quan, and Fuji Ren. 2016. Study on CNN in the recognition of emotion in audio and images. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS'16)*. IEEE, 1–5.
- [31] Fusang Zhang, Zhi Wang, Beihong Jin, Jie Xiong, and Daqing Zhang. 2020. Your smart speaker can “hear” your heartbeat! *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–24.
- [32] Hanbin Zhang, Chen Song, Aosen Wang, Chenhan Xu, Dongmei Li, and Wenyao Xu. 2019. Pdvocal: Towards privacy-preserving Parkinson’s disease detection using non-speech body sounds. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

Received 15 October 2021; revised 29 June 2022; accepted 30 June 2022