

<https://doi.org/10.1109/JIOT.2024.3351838>

Liang, Kangyue, Junxin Chen, Tongyue He, Wei Wang, Amit Kumar Singh, Danda B. Rawat, Houbing Song, and Zhihan Lyu. "Review of the Open Datasets for Contactless Sensing." IEEE Internet of Things Journal, 2024, 1–1. <https://doi.org/10.1109/JIOT.2024.3351838>.

© 2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Review of the open datasets for contactless sensing

Kangyue Liang, Junxin Chen, *Senior Member, IEEE*, Tongyue He, *Graduate Student Member, IEEE*, Wei Wang, *Member, IEEE*, Amit Kumar Singh, *Senior Member, IEEE*, Danda B. Rawat, *Senior Member, IEEE*, Houbing Song, *Fellow, IEEE*, and Zhihan Lyu, *Senior Member, IEEE*

Abstract—Recent years have witnessed the increasing popularity and dramatic progress of contactless sensing technologies, which are able to conduct remote signal acquisition without body contact. Both the physical signs and the physiological parameters can be acquired with contactless sensing. This paper introduces popular contactless sensing technologies, explores their application scenarios, and delves into the underlying theoretical principles. It comprehensively reviews the open datasets released in this field, encompassing collection scenarios, sample counts, data formats, and volunteer information. The performance baseline, typical work, and accessible links are also furnished. In addition, it includes discussions on the primary challenges and potential solutions in the context of contactless sensing with open datasets. Finally, suggestions for establishing a high-quality dataset are also given to the community.

Index Terms—Contactless sensing, Vision sensing, Wireless sensing, Acoustic sensing, Dataset

I. INTRODUCTION

Contactless sensing has drawn increasing attention in recent years [1]–[3]. There is no need to touch the body, so it is able to achieve remote signal acquisition and monitoring conveniently. The applications of contactless sensing can be identified in two folds, i.e., contactless sensing for physical signs and for physiological signs. Physical signs consist of human gestures, movements, postures, etc., which refer to the information revealed through the physical level of the body. Physiological signs are the body's internal physiological signals, such as heart rate, breathing, emotions, etc. With the aging population [4], [5], contactless sensing is increasingly important for healthcare monitoring in hospitals and nursing homes.

This work is funded by the National Natural Science Foundation of China (No. 62171114), the Fundamental Research Funds for the Central Universities (No. DUT22RC(3)099).

Kangyue Liang and Tongyue He are both with the College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110004, China.

Junxin Chen is with the School of Software, Dalian University of Technology, Dalian 116621, China.

Wei Wang is with the Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, China; and also with the School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China.

Amit Kumar Singh is with the Computer Science and Engineering Department, National Institute of Technology Patna, Bihar, India.

Danda B. Rawat is with the Department of Electrical and Computer Science, Howard University, Washington DC 20059, USA.

Houbing Song is with Department of Information Systems, University of Maryland, Baltimore County, MD 21250, USA.

Zhihan Lyu is with Department of Game Design, Faculty of Arts, Uppsala University, Visby 62157, Sweden.

Junxin Chen is the corresponding author. Email: junxinchen@ieee.org.

Numerous contactless sensing technologies are emerging. To the best of our knowledge, it has three primary branches, i.e., radio frequency-based sensing, vision-based sensing, and acoustic-based sensing [6]. Radio frequency signals suitable for contactless sensing consist of WiFi, ultrawideband (UWB), and millimeter-wave (mmWave). The vision-based technologies are based on visible light and infrared (IR). On the other hand, the acoustic method generally refers to using audio devices for signal acquisition or extraction. Previous works demonstrated that contactless sensing has significantly promoted the developments of smart homes, Internet of Things (IoT), human-computer interaction, and healthcare. The use of contactless sensing technology has been deemed practical, applicable, and convenient [7]. For example, hospitals have used contactless sensing technology to monitor patient behaviors and influenza-related indicators [8], while long-term and continuous monitoring methods are suitable for daily healthcare monitoring in a smart home [9]. Finally, acoustic signals and human gestures are considered reliable modalities for human-computer interaction [10], [11]. This paper reviews the open datasets for contactless sensing.

A. Contactless-sensing technology applications

Contactless monitoring of human health is gaining popularity, focusing on two main folds, i.e., physical and physiological signs [6]. Physical signs consist of human body actions and visible information such as facial expressions and human activities. In specific, this review focuses on physical signs in terms of activity recognition, gesture recognition, localization and tracking, face recognition, and speech recognition. Physical signals, such as positions or movements, can reflect a person's physical condition [2]. Effective and accurate information on the activities and behaviors of people helps to ensure a healthy and safe living environment [12], and is also beneficial for earlier disease detection. For example, Parkinson's disease is related to movement characteristics of the human body [13], while speech recognition has been exploited to diagnose Alzheimer's disease [14]. Physiological signs discussed in this paper include heart rate, respiration, emotion recognition, sleeping, and depression detection. It can indicate the health status, lifestyle, emotional state, or the early onset of some diseases. For example, respiration frequency and heart rate are both parameters that assess human body stress levels, and respiration frequency is also used for sleep monitoring [15].

B. Comparison between contact-based and contactless-based sensing methods

Contact sensing methods mainly use wearable devices or medical equipment in the hospital. Wearable devices are

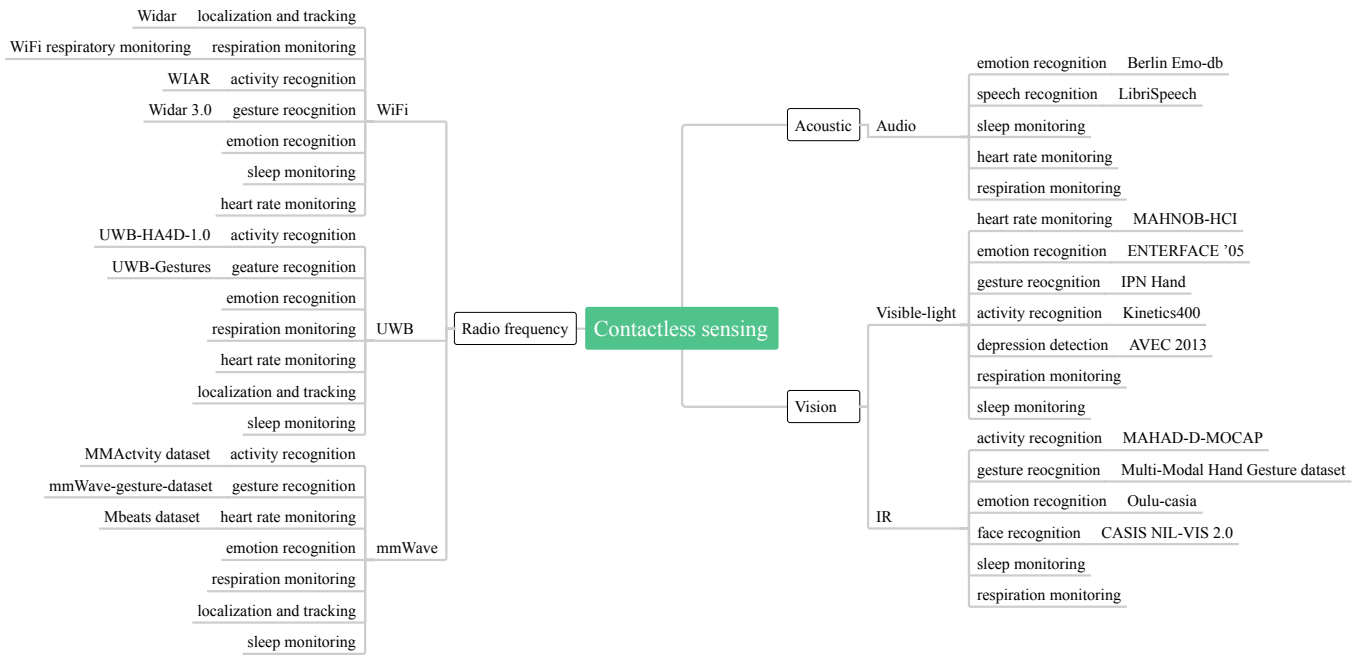


Fig. 1. Overview of contactless sensing technologies

operated by contacting the skin with the sensors or implanting the devices into the human body [16]. However, the user's movements may interfere with the wearable devices, have a certain impact on the results, and further lead to data distortion. The size, form, and weight of the batteries also greatly limit the developments of wearable devices. On the other hand, contactless sensing devices can be supplied with continuous power and do not need batteries [6], [17]. It is noteworthy that the issue of biocompatibility should also be taken into consideration [17].

Utilizing contactless sensing technology, patients can be detected or monitored near the devices without the need to wear or contact them [18]–[20]. It could prevent some problems, such as forgetting to wear the device or to charge it, which are particularly serious for elderly people [21]. Meanwhile, contactless sensing is also able to achieve multi-object monitoring and detection [22], which is impossible for wearable devices.

C. Datasets for contactless sensing

There are three kinds of datasets [23], [24] in the research field of contactless sensing. First, some datasets are freely available without any conditions. These datasets are often hosted on publicly accessible platforms or repositories. Researchers can simply navigate to the website, locate the dataset of interest, and download it without any additional steps. Second, some datasets require researchers to go through an application process before obtaining access. In such cases, interested researchers need to fill out an application form and obtain approval from the dataset owner or administrator. This step ensures that the dataset is being used appropriately and in accordance with any data sharing policies or ethical considerations. Third, certain datasets may only provide a

description of the data collection phase without specifying how to access the actual dataset. This could be due to various reasons, such as privacy concerns or legal restrictions, that led to the dataset not being made public.

For papers related to a certain dataset on contactless sensing, they mainly include the certification or credentials required for the dataset, the equipment used for data collection, the collection scenario, dataset availability, sample counts, data formats, performance baselines, and the volunteers' information.

Currently, there is no review paper focusing on summarizing the open datasets for contactless sensing.

D. Comparison with similar reviews

Dataset reviews are important as they provide researchers with a comprehensive summary of the datasets relevant to a specific domain. It helps to avoid duplicating the collection of existing data, saving research costs and time. With dataset reviews, researchers can comparatively understand the characteristics, scope, and availability of different datasets to guide their own research design. They can also decide whether they need to combine multiple datasets or select the dataset that best matches their study.

Compared with other reviews, which generally focus on the introduction and comparison of different technologies, we focus on the open datasets that have been released for contactless sensing technologies. To the best of our knowledge, there are no such reviews on summarizing open datasets for contactless sensing. Meanwhile, we will provide a comprehensive overview of various contactless sensing technologies, focusing specifically on reviewing open datasets within this domain. There are some contactless tasks that lack public datasets, and we hope that these areas will have public, reliable datasets in the future.

E. Contribution

Our contributions can be summarized as follows.

- 1) We provide a complete picture of contactless sensing technologies and their typical applications.
- 2) We comprehensively summarize open-access datasets for contactless sensing, providing a comparative discussion and including relevant links.
- 3) Discussions about the underlying problems and potential solutions are given, and the expected conditions for establishing an open dataset are also suggested.

F. Structure

This paper will proceed as follows. Section II introduces the literature on human physical and physiological signs as well as contactless sensing technologies. Section III demonstrates open datasets on WiFi-based contactless sensing. Section IV discusses open datasets on UWB-based contactless sensing. Section V presents open datasets on mmWave-based contactless sensing. Section VI shows open datasets on visible-light-based contactless sensing. Section VII surveys open datasets on IR-based contactless sensing. Section VIII concludes open datasets on audio-based contactless sensing. Section IX provides the discussions as well as the challenges and possible solutions in this field.

II. PRELIMINARIES

Collecting real-time data from a person's daily life by contactless ways is an emerging trend [6]. Compared with wearable devices, contactless sensing methods do not need to touch the skin and are able to collect data in a non-invasive way. Fig. 1 demonstrates the popular contactless sensing technologies and their application scenarios, where a typical dataset is followed if it has public datasets. In addition, Table I also comparatively lists the involved applications and their enabling contactless sensing technologies. The entries with checkmarks represent the technologies and their enabled applications. For example, WiFi-based technology can be used for emotion recognition, respiration monitoring, activity recognition, sleep monitoring, gesture recognition, heart rate monitoring, and location and tracking.

A. Human physical signs and physiological signs

The following items list the involved physical and physiological signs reviewed in this paper.

1) *Emotion recognition*: Emotion is defined as the body's reaction to an event, a situation, or a person [25]. Keeping a relaxed mood can promote our happiness and comfort [26]. In [27], authors proposed that facial emotion recognition could be beneficial for many areas, such as education, entertainment, and marketing. EEG has been widely used for emotion recognition, yet this paper will demonstrate that most of the contactless technologies are feasible for contactless emotion recognition.

2) *Respiration monitoring*: Respiration monitoring plays an important role in many applications, such as healthcare, recruitment, the workplace, sport, and exercise [28]. Respiration monitoring is a very common way to diagnose, prognosis, and manage pneumonia in the COVID-19 pandemic and is also used to detect emotional stress and cognitive load [28].

3) *Sleep monitoring*: The task of sleep monitoring can be broadly divided into sleep stage monitoring and sleep posture monitoring. Sleep stage classification focuses on investigating the fundamental sleep cycles, and any disruption or insufficient time spent in a particular cycle can result in disorders [29]. Other studies focus on monitoring sleep postures, examining body movements and positions such as supine, prone, left, and right, as well as the duration of each posture [30].

4) *Activity recognition*: Human activity recognition has been found very useful for facilitating our daily lives [31]. For instance, it can be used to monitor patient situations in hospitals and identify the activities of elderly people in nursing homes to prevent harmful actions such as falling down. In [32], authors proposed that human activity recognition may reduce the need for human resources in the handling processes of manufacturing and logistics.

5) *Gesture recognition*: Generally, gesture recognition has two ways, including sign language recognition and hand movement recognition, similar to activity recognition [33]. Some meaningful actions performed by a person can be recognized by gesture recognition, and it can be used for human-computer interfaces, healthcare, and virtual reality [34]. In addition, gesture recognition could also be used for sign language recognition. There are many deaf people using sign language, and sign language is also used by many hearing people to help other deaf people understand what they say [35].

6) *Heart rate monitoring*: Heart rate refers to the number of heartbeats per minute, which is very important for disease diagnosis and healthcare monitoring [36]–[38]. Heart rate measurement has three folds: ballistocardiography (BCG), photoplethysmography (PPG), and remote photoplethysmography (rPPG). BCG is a heart rate monitoring method using involuntary head movement during blood circulation [37], while PPG and rPPG are both based on the blood volume change in microvascular tissue during a heartbeat [36].

7) *Localization and tracking*: Indoor localization is an application in an indoor environment to obtain a person's or a device's location [39]. It is applicable in many scenarios. For example, customers can use localization systems to find the location of the product they need in the supermarket; in natural disasters, the location system is helpful for searching for victims; and in military applications, it can detect location and intrusion events [40].

8) *Depression detection*: Traditional depression assessment involves clinical interviews or self-assessments. The evaluation forms always consist of various questions that are used to elicit responses from patients [41]. Based on the received responses, the patients are assigned a score that correlates with their level of depression. One of the latest advancements in the field of depression detection is the use of automatic detection.

9) *Face recognition*: Face recognition is developed to identify individuals based on their facial features [42]. It is used to

Table I. Comparison of various contactless sensing technologies

Sensing Category	Emotion Recognition	Respiration Monitoring	Sleep Monitoring	Activity recognition	Gesture Recognition	Heart Rate Monitoring	Localization and tracking	Depression Detection	Face Recognition	Speech Recognition
WiFi	✓	✓	✓	✓	✓	✓	✓			
UWB	✓	✓	✓	✓	✓	✓	✓			
mmWave	✓	✓	✓	✓	✓	✓	✓			
Visible-light	✓	✓	✓	✓	✓	✓		✓		
IR	✓	✓	✓	✓	✓				✓	
Audio	✓	✓	✓			✓				✓

determine whether two faces in different images belong to the same person or search for a face in a large dataset. Biometric security systems use facial recognition to uniquely identify individuals during user logins and enhance the security of authentication activities.

10) *Speech recognition*: Speech recognition is the process of transforming a sound signal into text, and the output of the system is text that corresponds to the speech signals [43]. It is able to make human-computer interaction more fast and convenient without external equipment. In addition, speech recognition can also be used in the healthcare industry, which can help elderly and disabled people. Of course, audio signals are used in this task.

B. Contactless sensing technologies

We investigate different kinds of contactless sensing technologies, which can be identified into three primary branches, i.e., radio frequency-based sensing, vision-based sensing, and acoustic-based sensing, as demonstrated in Fig. 1. The underlying theories explaining why specific techniques are suitable for certain applications are important and will be described in this subsection [44].

- 1) Radio frequency-based techniques, including WiFi, UWB, and mmWave, utilize the propagation and interaction of electromagnetic waves for information extraction. The operating frequency of WiFi devices is 2.4~5 GHz, while that of UWB is 3.1~10.6 GHz with a bandwidth of at least 500 MHz. In addition, the mmWave radar operates at 30~300 GHz. Considering that WiFi, UWB, and mmWave have been widely deployed, they will be listed in independent sections in the following parts.
- 2) Vision-based techniques include visible light and IR. Visible light presents light with wavelengths of 400 to 700 nm, which is the same spectrum perceived by the human eye. IR thermal imaging uses IR to generate thermal images, so that the temperature distribution and thermal characteristics of the target can be detected [45]. The visible light and IR are listed as independent sections in the following parts.
- 3) Acoustic-based technology mainly contains audio signal processing [46]. Audio signals are produced by acoustic-based equipment that converts sound waves into electronic signals. The audio possesses a frequency range spanning approximately 20 to 20,000 Hz, closely aligning with the audible range of the human ear. Acoustic-based technology, renowned for its non-contact, non-

visual data acquisition, and heightened sensitivity, finds extensive applications for emotion recognition and speech recognition.

The characteristics of the contactless sensing technologies are listed in Table II, and are further described as follows.

1) *WiFi-based contactless sensing*: Recent years have witnessed the exploration of WiFi signals for emotion recognition, respiration monitoring, activity recognition, sleep monitoring, gesture recognition, heart rate monitoring, and localization and tracking [9], [52]. Qian *et al.* [53] proposed Widar2.0, a system that can achieve passive human positioning and tracking on a single WiFi commercial device. Ma *et al.* [35] used WiFi-CSI data for recognition of 276 American sign languages. Wsleep was proposed by Liu *et al.* [54], which used WiFi-CSI signal to extract breathing-related rhythm patterns and body movements during sleep. This paper introduces open datasets for localization, activity recognition, respiration monitoring, and gesture recognition.

2) *UWB-based contactless sensing*: UWB sensing is advantageous in its fast speed and wide range, and it does not need light [47]. Various UWB-based solutions have been proposed for emotion recognition, respiration monitoring, activity recognition, sleep monitoring, gesture recognition, heart rate monitoring, and localization and tracking. For example, a UWB-based human indoor localization and recognition system was proposed by Cheng *et al.* [55], and a UWB-based radar was developed by Lee *et al.* [56] for heart rhythm and heart rate measurement. A gesture recognition algorithm using UWB radar was presented by Ahmed *et al.* [57]. This paper focuses on open datasets for UWB-based activity recognition, heart rate monitoring, and gesture recognition.

3) *mmWave-based contactless sensing*: Benefiting from the advantages in terms of real-time monitoring, large coverage, and high resolution, mmWave radar has been exploited for emotion recognition, respiration monitoring, activity recognition, sleep monitoring, gesture recognition, heart rate monitoring, and localization and tracking [58]–[60]. Singh *et al.* [58] utilized mmWave radar for activity classification, while mmWave radar is exploited for gesture recognition in [59]. This paper concentrates on the mmWave-based public datasets that are used in activity recognition, gesture recognition, and heart rate monitoring.

4) *Visible-light-based contactless sensing*: Visible-light-based contactless sensing has found broad applications in the fields of emotion recognition, respiration monitoring, activity recognition, sleep monitoring, gesture recognition, heart rate

Table II. Characteristics of different contactless sensing technologies

Technology	Contactless	Sensitivity	Accuracy	Cost	Advantages	Disadvantages	Ref
WiFi-based	Yes	High	Mid	Low	it is widely deployed and can run in the dark.	narrow bandwidth, difficult for sensing multi-subject	[9]
UWB-based	Yes	High	high	high	fast data transmission speed, wide sensing range, it can run in dark	cost high, easily disturbed by other frequency	[47]
mmWave-based	Yes	High	high	high	small size, large bandwidth, high sensitivity and precision, can work in dark	high energy consumption, short measurement range	[48]
Visible-light-based	Yes	High	mid	low	widely used, easy to deploy	susceptible to light interference	[49]
IR-based	Yes	High	high	high	long distance sensing, do not need light	affected by temperature	[50]
Audio-based	Yes	Low	mid	low	do not depend on light, simple and convenient for measurement	affected by noise	[51]

monitoring, and depression detection. For example, Kwon *et al.* [61] used video captured by a smartphone to monitor heart rate, while a video-based algorithm for emotion recognition was developed by Kaya *et al.* [62]. This paper focuses on open accessible datasets for vision-based emotion recognition, heart rate monitoring, and depression detection.

5) *IR-based contactless sensing*: In addition to its vision-based advantages, IR-based sensing further has high sensitivity, long distance, and does not rely on a light source. It has been widely employed for emotion recognition, respiration monitoring, activity recognition, sleep monitoring, gesture recognition, and face recognition [50]. Magdalena *et al.* [63] used an IR camera to monitor heart rate when the users were driving. The method of using IR facial images to identify emotion on the images was proposed by Cruz *et al.* [64].

6) *Audio-based contactless sensing*: Acoustic-based sensing is advantageous because it is cheap and ubiquitous, and it does not rely on light sources. The use of Audio-based contactless sensing technology has become prevalent in emotion recognition, respiration monitoring, sleep monitoring, heart rate monitoring, and speech recognition. A framework for emotion recognition from speech signals was suggested by Wang *et al.* [65], and they verified their model on two different languages (Chinese and German). Hannun *et al.* [66] presented an end-to-end speech recognition system. In this paper, we introduced open datasets for audio-based emotion recognition and speech recognition.

In addition, these sensing technologies can be used in a hybrid, which can be called hybrid contactless sensing technologies. It combines multiple contactless techniques for a more accurate and comprehensive perception of different objects or environments. For example, audio-visual joint algorithms were proposed to recognize and analyze a person's emotional state [67], [68]. On the other hand, a combination of WiFi and video has been explored for detecting and tracking human poses and movements. Zou *et al.* [69] proposed WiVi, a device-free human activity recognition scheme that utilizes WiFi-based and vision-based devices.

The following Sections will review the open datasets of contactless sensing technologies for their enabled applications. They are divided into several sections by the contactless sensing technologies, and each section will be further divided by its enabled applications which have open accessible dataset. In other words, the following part will not cover all the

ticked entries in Table I, because only the combination of a contactless sensing technology and an enabled application with open datasets will be reviewed. For example, WiFi has been exploited for sleep monitoring, as marked in Table I. However, as of now, there are no open-access datasets available in this field. Consequently, this application will not be reviewed in Section III, but open datasets for the other three WiFi-based sensing applications are comprehensively investigated.

III. WiFi-BASED CONTACTLESS SENSING

Nowadays, WiFi infrastructure has been widely deployed in schools, homes, and office buildings. The broad deployments of WiFi devices facilitate the investigation and application of human physical and physiological signs' monitoring. Two kinds of WiFi information can be used for contactless sensing, i.e., RSSI, which stands for radio signal strength, and CSI, which refers to channel state information [80]. The availability of data for different tasks is as follows.

- Tasks have open datasets: indoor localization, activity recognition, gesture recognition, and respiration monitoring.
- Tasks don't have open datasets: sleep monitoring and emotion recognition.

The following subsections will review the open datasets in WiFi-based contactless sensing works.

A. WiFi-based localization and tracking

This task involves utilizing WiFi signals for indoor position identification, and there are various solutions. In [70], [77], the velocity and location of the users are estimated by analyzing the signal propagation path, and for the WiFi fingerprint method, it is divided into two steps: adjusting and handling [81]. In the first step, a graph of the place in radio ways where we want to detect is established. While in the operation step, the signal strength of the user's location is estimated and sent to the server to compare the signal strength with the strength on the radio map. The user position is located at the most similar place [73]. In [77], authors solved indoor localization problems using the pattern classification method, and the position was predicted as being in a work area, kitchen, etc. Open datasets for indoor localization using WiFi signals are listed in Table III, and described in detail as follows.

Table III. Wi-Fi-based localization and tracking

Dataset	Author	Introduction	Typical works	Access
Widar	Qian <i>et al.</i> [53], [70]	This dataset contains six person walking in different shapes (line, circle, etc.) Wi-Fi-CSI data, and uses video tracking as ground truth.	[53], [70], [71]	http://tms.thss.tsinghua.edu.cn/wifiradar
WiFi fingerprints dataset	Laurila <i>et al.</i> [72]	This dataset contains 4,038 WiFi fingerprints.	[72]	https://zenodo.org/record/889798#.Y10ZurZBxPY
UjiIndoorLoc	Torres <i>et al.</i> [73]	This dataset contains 21,048 Wi-Fi-RSSI signal data in three buildings of their university.	[73]–[76],	https://archive.ics.uci.edu/ml/datasets/UJIIndoorLoc
Wireless Indoor Localization Data Set	Rohra <i>et al.</i> [77]	This dataset contains 2,000 data and each data contains 7 Wi-Fi signal strengths.	[77], [78]	https://archive.ics.uci.edu/ml/datasets/Wireless+Indoor+Localization
2.4 GHZ INDOOR CHANNEL MEASUREMENTS	Chen <i>et al.</i> [79]	This dataset was collected in four different rooms and each room contains 1,177,960 signal data which contains 196 positions.	[79]	https://ieee-dataport.org/documents/24-ghz-indoor-channel-measurements

Widar. The Widar dataset was collected in an indoor environment by Tsinghua University [70]. A total of 5 volunteers participated in data acquisition, and all of them walked along different shapes of trajectories that were designed by the authors. This dataset consisted of eighty Wi-Fi signals. The researchers used digital cameras to record the volunteers' walking path as ground truth. The Widar 2.0 dataset was an extension. It was collected in three different indoor environments, which contained a cramped hallway, a compact office space containing multiple furnishings, and a spacious classroom with no furniture. Six people participated in this dataset [53]. Widar has been widely used for indoor localization. Tian *et al.* [71] reported a Wi-Fi-based indoor target angle estimation algorithm with Widar2.0, and the results performed best in corridors, where they achieved a 3.62° average angle error, 7.18° in classrooms, and 12.16° in offices. Both of the datasets are openly accessible.

WiFi fingerprints dataset. A research team at the University of Tampere in Finland collected a total of 4,648 Wi-Fi fingerprints, and this dataset was split into 697 Wi-Fi fingerprints for training, and 3,951 Wi-Fi fingerprints for testing [72]. In total, eight volunteers participated in data acquisition. This dataset contained the user location information and the RSSI levels from the access points. It was recorded in its five-floors university buildings and contained date and device model information. In the data collection stage, the users held their mobile phones to move in different areas of the building, and entered the location information into the mobile phone. The location information and RSSI levels from the access points of the user were recorded. This dataset can be directly downloaded from the website.

UjiIndoorLoc. UjiIndoorLoc was an indoor localization dataset collected by a research team from the University of Jaime I in Spain [73]. There were twenty volunteers participating, and a total of 21,049 Wi-Fi-RSSI samples were collected in three different indoor environments. The researchers who collected this dataset used a basic localization system for verification, and they achieved 7.9 meters in average position error. This dataset has been widely employed for indoor localization. Kim *et al.* [74] used a deep neural network (DNN) architecture to classify the human locations in different buildings and on different floors. Bozkurt *et al.* [75] found K-

Nearest Neighbors (KNN) performed best for indoor localization tasks by comparing different machine learning algorithms in this dataset. Anyone who wants this dataset can download it from the website accessed in the table.

Wireless Indoor Localization Data Set. A research team from the Vellore Institute of Technology University established this dataset [77]. The dataset was collected in three indoor environments, which consisted of an office, a kitchen, and an indoor sports room. It contained seven parameters of the Wi-Fi signal strength and had two thousand rows of data. Gomes *et al.* [78] compared different algorithms, and they found that random forest performed the best. This dataset can be downloaded directly from the website provided in the table.

2.4 GHZ INDOOR CHANNEL MEASUREMENTS. This dataset was archived by the Institute for Infocomm Research in Singapore [79]. It was collected in four different rooms (the laboratory, narrow corridor, lobby, and sports hall) on the Khalifa University campus in Sharjah, UAE. Each room has 196 square cells as anchor points, and data is collected 10 times at each point. In total, each room collected 1,177,960 pieces of the signal, which contained 196 locations. The training set consisted of 75% of the collected data, while the remaining 25% was reserved for performance testing. The authors also proposed a deep learning-based localization method to verify this dataset, and it achieved 97.8% accuracy. This dataset can be directly downloaded from the website.

B. Wi-Fi-based respiration monitoring

Human respiration induces fluctuations in Wi-Fi signals, and these rhythmic patterns can be discerned [90]. Table IV lists an open dataset for Wi-Fi-based respiration monitoring.

Wi-Fi respiratory monitoring. A research team at Pakistan's National University of Science and Technology used Wi-Fi to collect data on three activities, including breathing movements, static environments, and random movements [82]. Each record collected a ten-second signal, and the total number of records was 480. Volunteers sat in front of the Wi-Fi device and performed breathing, static environments, and random movements. CNN has been employed to verify their dataset, and it achieved 94.85% in static scenes. Interested researchers need to submit an application on the website.

Table IV. WiFi-based respiration monitoring

Dataset	Author	Introduction	Typical works	Access
WiFi respiratory monitoring	Khan [82]	This dataset contains 1,440 signals which divided into three activities including breathing movements, static environments, and random movements.	[82]	http://ipt.seecs.nust.edu.pk/

Table V. WiFi-based activity recognition

Dataset	Author	Introduction	Activities	Typical works	Access
UT-HAR	Yousefi [80], [83]	This dataset contains totally 720 WiFi-CSI data (six person, six activities, and twenty times).	Six (lying, falling, walking, running, sitting, and standing)	[80], [84]	https://github.com/ermongroup/Wifi_Activity_Recognition
WIAR	Guo [85]	This dataset contains 8,000 WiFi signals performed by 10 people in three indoor environments.	Sixteen (clap hand, take a seat, move around, drink water, make a call, different leg movements including kick forward, kick side, bend, and deep bend, wave hands in different directions and heights including wave in horizontal, wave two hand in same time, wave hand high, throw movement, throw paper, draw tick, and draw x)	[85]	https://github.com/lintere sa/WiAR
Human-to-Human activities dataset	Alazrai [86]	This dataset contains 4,800 WiFi-CSI data (forty person, twelve activities, and ten times).	Twelve (body movements including come and leave, movements on leg such as kick in left leg and in right leg, hand activities including shake hand, wave hand, hug, push, point in left hand and right hand, punch in left hand and right hand)	[86], [86], [87]	https://data.mendeley.com/datasets/3dhn4xnjxw/draft?a=90c726d4-5493-4efc-9ee6-973bcd922b31
NTU-Fi HAR	Yang [83], [88]	400 samples data were collected in a lab.	Six activities (jog, punch, stroll, fall down, revolve hand, sweep the floor)	[88]	https://drive.google.com/drive/folders/1R0R8SIVbLI1iUFQCzh_mH90H_4CW2iwt
NTU-Fi Human-ID	Wang [83], [89]	This dataset was collected in two places(a lab and an office) with four environment settings, and each activity 2,000 times.	Six activities (jogging, punching, strolling, falling down, revolving hand, and sweeping the floor)	[89]	https://drive.google.com/drive/folders/1R0R8SIVbLI1iUFQCzh_mH90H_4CW2iwt

C. WiFi-based activity recognition

Datasets for activity detection using WiFi signals are listed in Table V, and described in detail as follows.

UT-HAR. This dataset, containing 4GB WiFi-CSI amplitude data, was collected at the University of Toronto and Stanford [80], [83]. It recorded CSI amplitude information for six users with six activities (lying, falling, walking, running, sitting, and standing). Every activity contained 20 cases of information. Chen *et al.* [84] suggested using long short-term memory with an attention mechanism to carry out recognition activity with this dataset.

The overall accuracy exceeds 95%, with the recognition accuracy for fall activities reaching approximately 99%. Researchers who want to use this dataset can download it directly from the website in the table.

WIAR. The WIAR dataset was collected by the Dalian University of Technology [85]. It contained WiFi signals reflected by 16 activities. There were ten volunteers, including five men and five women, and each volunteer performed all 16 actions in three different rooms (a room without furniture, a workspace, and a meeting room), and each action was repeated 50 times. The authors have verified their dataset, and the achieved accuracy was 94.7% in a meeting room, 93% in

an empty room, and 87% in an office. This dataset is easily available from the website provided in the table.

NTU-Fi HAR. This dataset was collected by a research team at Nanyang Technological University [83], [88]. It was used for two tasks, human activity recognition and human identification. For activity recognition, they collected 400 samples in a lab, which contained six human activities (jog, punch, stroll, fall down, revolve the hand, and sweep the floor) performed by twenty volunteers. Each volunteer performed each activity twenty times. The human identification part contains 15 volunteers walking through a WiFi device, and each volunteer repeated it 60 times. The authors also verified their dataset, and achieved an accuracy of 98.6% for activity recognition. Researchers who want to use it can directly get it from the website.

NTU-Fi Human-ID. This dataset was also collected at Nanyang Technological University, and the activity categories were similar to NTU-Fi HAR [83], [88]. There were two places for the data collection (a lab and an office), and each place was designed to contain four environmental settings. Each human activity was performed 2,000 times by volunteers. Researchers who want to use this dataset can get it from the website in the table.

Table VI. WiFi-based gesture recognition

Dataset	Author	Introduction	Gestures	Typical works	Access
SignFi	MA [35]	This dataset contains 8,280 gesture instances, with 5,520 captured in a laboratory setting and the remaining 2,760 recorded in a home environment.	276 categories American Sign Language	[35]	https://yongsen.github.io/SignFi/
Widar3.0	Zhang [91]	This dataset contains two parts, which are 12,000 WiFi-CSI data (16 people, 5 locations, 5 directions, six gestures, and 5 times) and 5,000 WiFi-CSI data (two people, 5 locations, 5 directions, 10 gestures, and 10 times).	16 gestures (wave hand, clap hand, slide hand, push and pull, draw circle and zigzag in air, and draw 0-9 in horizontal plane)	[91], [92]	http://tns.thss.tsinghua.edu.cn/widar3.0
ARIL	Wang [93]	This dataset contains 1,394 samples of WiFi-CSI signals of 6 gestures repeat 15 times in 16 locations by one volunteer.	six gestures (palm raise, palm down, palm to left, palm to right, palm circle, and palm cross)	[93]	https://github.com/geekfei/w/ARIL

Human-to-Human activities dataset. This dataset was collected by the University of Jordan in Germany, and it contained the RSSI and CSI data of 12 kinds of human-to-human activities performed by 40 volunteers [86]. It was collected in an indoor environment (an office), and each activity was performed ten times. The total number of records is 4,800. With this dataset, Alazrai *et al.* [87] proposed a deep learning algorithm that achieved an accuracy value of 86.3% for activity recognition. The dataset can be directly downloaded.

D. WiFi-based gesture recognition

This task goal is to use the WiFi signals affected by in-air hand movement to recognize gestures [94]. Datasets for gesture recognition using WiFi signals are listed in Table VI and described in detail as follows.

SignFi. The SignFi dataset was collected by a research team at the College of William & Mary [35]. It consisted of 8,280 pieces of American Sign Language WiFi-CSI data performed by six users, and the gestures' categories were 276. The authors collected 5,520 samples and 2,760 samples in the lab and at home, respectively. They used a deep learning algorithm to verify this dataset. The result of the average classification accuracy was 98.01% in the lab, 98.91% in the home, and 94.81% in both the lab and home environments. Anyone who wants to use it can get it directly.

Widar3.0. The Widar3.0 gesture dataset was collected by Tsinghua University [91]. It contained 17,000 pieces of WiFi-CSI records collected from different scenarios (including different locations, orientations, and environments). Two kinds of gesture data were collected in this dataset. The first category contained six kinds of gestures (wave hand, clap hand, slide hand, push and pull, draw circle, and zigzag in the air) in different directions, and sixteen people did all the gestures five times. The second part has ten simple gestures (drawing 0-9 in a horizontal plane) done by two volunteers in five locations, five directions, and ten times. The authors also verified their dataset, and the result of accuracy was 92.7%. The dataset has been uploaded and can be freely downloaded.

ARIL. The ARIL dataset was collected by Xian Jiaotong University [93]. A total of 1,394 WiFi-CSI signals were collected. It contained six gestures, including palm raise, palm down, palm to left, palm to right, palm circle, and palm cross, performed by one volunteer. Each activity was

repeated 15 times in 16 locations, and the locations were evenly distributed in a room. As it contains different locations, this dataset can also be used for localization tasks. The author used ResNet1D to verify this dataset and achieved 88.13% accuracy for gesture recognition and 95.68% accuracy for localization. The provided website allows for direct download of this dataset.

IV. UWB-BASED CONTACTLESS SENSING

Due to its high accuracy and resolution, UWB has attracted the interest of researchers and is widely used in various fields. The availability of data for different tasks is as follows:

- Tasks have open datasets: activity recognition, gesture recognition, and heart rate monitoring.
- Tasks don't have open datasets: respiration monitoring and fall detection.

The following subsections will review the open datasets in UWB-based contactless sensing works.

A. UWB-based activity recognition

UWB-based activity recognition is similar to that of WiFi-based methods. It can recognize or monitor human activity and does not depend on wearable devices [95]. Table VII lists the available datasets which are detailed as follows.

UWB activity dataset. A research team from the University of Quebec gathered this dataset, encompassing 15 distinct activities [95]. Ten volunteers provided data, each contributing ten cases for different activities. These activities were conducted in an apartment using three UWB radars. Access to this dataset is available with the author's permission.

UWB-HA4D-1.0. It was collected at the National University of Defense Technology in China and comprises 2,757 records, with each record containing 40 frames of 3D radar data [96]. The samples were collected from 11 volunteers in three different scenarios (unobstructed, 3 cm plastic plate shielding, and 27 cm brick wall shielding). During the data collection process, volunteers were asked to perform the following ten actions: open arms, punch, sit, kick, sit, stand, walk forward, walk left, walk right, and wave. To validate this dataset, a CNN-based activity recognition method was developed, achieving accuracy rates of 92.25% on unobstructed data, 90.00% on 3 cm plastic plate shielding, and 77.00% on 27

Table VII. UWB-based activity recognition

Dataset	Author	Introduction	Gestures	Typical works	Access
UWB activity dataset	Maitre [95]	This dataset contains 1,500 UWB signals in 15 activities.	15 activities (cook pasta, drink, sleep, put on a jacket, do the housework, do the dishes, wash your hands, brush teeth, read a book, eat, walk, put on shoes, take the medication, make tea, and use a computer)	[95]	https://github.com/julienmaitre/Activity-Recognition-with-UWB
UWB-HA4D-1.0	Tian [96]	This dataset contains totally 2,757 UWB signals in 10 different activities.	10 activities (open your arms, punch, sit, kick, sit, stand, walk forward, walk left, walk right, and wave)	[96]	https://radars.ac.cn/web/data/getData?dataType=UWB-HA4D
IR-UWB-Through-wall-Radar-Human-Motion-Status-Dataset	Zhengliang [97]	This dataset contains 4,320 UWB signal data in two scenes and three different motion statues.	Three motions(standing, walking, and empty)	[97]	https://github.com/ZhengliangZhu-2020/IR-UWB-Through-wall-Radar-Human-Motion-Status-Dataset

Table VIII. UWB-based gesture recognition

Dataset	Author	Introduction	Gestures	Typical works	Access
UWB-Gestures	Ahmed [98]	This dataset contains 9,600 samples data with twelve gestures.	twelve categories gestures (horizontal swipes: left to right and right to left, vertical swipes: up to down and down to up, diagonal swipes in four directions, rotational swipes: clockwise and opposite clockwise, direct push, no movement)	[98]	https://doi.org/10.6084/m9.figshare.12652592

cm brick wall shielding. The dataset is available for direct download from the official website.

IR-UWB-Through-wall-Radar-Human-Motion-Status-Dataset. It was collected at Xiamen University, China [97]. It included 4,320 UWB signal records (2,160 signals through the wall and 2,160 without the wall), each of which lasted for 4 seconds. Three adults were asked to perform two motion behaviors (standing and walking). A CNN model was used for verification, and the accuracies were 100% and 99.7% for walking and standing scenes, respectively. This dataset can be downloaded directly from the website.

B. UWB-based gesture recognition

The UWB technology has also been widely used in gesture recognition due to its unique advantages. Open datasets for gesture recognition using UWB signals are listed in Table VIII and described in detail as follows.

UWB-Gestures. The UWB-Gestures dataset was collected at Hanyang University, Korea [98]. Three UWB sensors have been deployed in the gesture area in three different locations (left, top, and right). This dataset contains 12 categories of gestures (horizontal swipes: left to right and right to left; vertical swipes: up to down and down to up; diagonal swipes in four directions; rotational swipes: clockwise and opposite clockwise; direct push, no movement) performed in front of the human body. Eight volunteers participated in the dataset, and 9,600 samples were collected. They used a CNN-based algorithm to classify gestures, and the accuracy achieved was 94%. This dataset can be downloaded from the website.

V. MMWAVE-BASED CONTACTLESS SENSING

This section focuses on mmWave radar, where the detection process typically entails transmitting electromagnetic wave

signals and subsequently receiving the signals reflected by the target. The basic task is to detect surrounding targets and estimate their relevant parameters, such as distance, radial velocity, azimuth angle, etc. In recent years, mmWave radar has been widely used to monitor physical and physiological signals. The availability of data for different tasks is as follows:

- Tasks have open datasets: activity recognition, gesture recognition, and heart rate monitoring.
- Tasks don't have open datasets: respiration monitoring, emotion recognition, sleep monitoring, and localization and tracking.

The following subsections will review the open datasets in mmWave-based contactless sensing works.

A. mmWave-based gesture recognition

Taking advantages of high resolution and low power consumption, mmWave has been exploited for fine-grained tasks such as gesture recognition. The primary principle is to apply radar beam pulsing to the hand and then transform the received signal to obtain features for gesture recognition [59]. Table IX lists a dataset for mmWave-based gesture recognition.

mmWave-gesture-dataset. It was collected by the Beijing University of Posts and Telecommunications [99]. The mmWave radar was placed less than 0.5 meters and 2 to 5 meters away from the human body in the first and second scenarios, respectively. The mmWave radar consisted of four receiving antennas and two transmitting antennas. A total of 144 volunteers produced a total of 56,420 gesture samples. Each sample contained raw signal data, raw Range-Doppler images, PRM (pseudo representative model) information, and its ground truth. These instances lasted a total of 1357 minutes, and they contained five gestures (i.e., swiping left, swiping

Table IX. mmWave-based gesture recognition

Dataset	Author	Introduction	Gestures	Typical works	Access
mmWave-gesture-dataset	Liu <i>et al.</i> [99]	This dataset was collected in two scenarios containing raw signal data, raw Range-Doppler images, PRM (pseudo representative model) information.	five gestures (i.e. swiping left, swiping right, tapping, turning, and some unforeseen movements)	[99]	https://github.com/fengxudi/mmWave-gesture-dataset

Table X. mmWave-based activity recognition

Dataset	Author	Introduction	Activities	Typical works	Access
MMActivity dataset	Singh <i>et al.</i> [58]	This dataset contains 93 minutes data which contains information of spatial coordinates, speed, distance, intensity, and angle, performed by two volunteers.	five actions (i.e. walking, jumping, jumping jacks, squats and boxing)	[58]	https://github.com/nesl/Ra-dHAR
MMWave walking dataset	Gambi <i>et al.</i> [100]	This dataset contains 231 samples of data collected from 29 volunteers.	Six activities (slow walk, fast walk, and slow walk with hands in pockets, walking with a bottle between arms, walking slowly with waving arms slowly, and limping)	[100]	mmWaveRadarWalkingDataset-part1: https://doi.org/10.5281/zenodo.3824534 mmWaveRadarWalkingDataset-part2: https://doi.org/10.5281/zenodo.3897234

Table XI. mmWave-based heart rate monitoring

Dataset	Author	Introduction	Typical works	Access
Mbeats dataset	Zhao <i>et al.</i> [60]	This dataset contains millimeter radar wave signals in eight different postures lasted about 3 hours.	[60]	https://github.com/zhaoyan/mbeats

right, tapping, turning, and some unforeseen movements). This dataset can be obtained directly from the website provided. They used this dataset's PRM information for gesture recognition, and the achieved accuracy was 99.76%.

B. mmWave-based activity recognition

Because mmWave radar can penetrate clothing, it is able to perform long-distance monitoring without requiring images of the subject. Therefore, it has application prospects in activity recognition. Some open datasets for mmWave-based activity recognition are listed in Table X.

MMActivity dataset. It was collected at the University of California, Los Angeles, USA [58]. A mmWave radar with four receiving antennas and three transmitting antennas was employed. Two volunteers stood in front of a radar and performed five actions (i.e., walking, jumping, jumping jacks, squats, and boxing). Each piece of data lasted about 20 seconds, and a total of 93 minutes of data were collected. Each piece of data contains spatial coordinates, speed, distance (distance between the volunteer and the radar), intensity, and angle. The provided website allows for the direct download of this dataset.

MMWave walking dataset. It was collected by Universit Politecnica delle Marche, Italy [100]. They collected data with mmWave radar with one transmitter and four receivers in a room without obstacles. The volunteers first walked 10 meters away from the radar according to different walking postures and then approached the radar. Each piece of data lasted about 16 seconds. In the first part, 19 volunteers walked in three postures (i.e., slow walk, fast walk, and slow walk with hands

in pockets), and a total of 171 pieces of data were collected. In the second part, 10 volunteers walked in three walking postures (i.e., walking with a bottle between arms, walking slowly with waving arms, and limping), and a total of 60 pieces of data were collected. This dataset can be directly downloaded. Wang *et al.* [101] proposed a 3D Orthogonally-Projected EfficientNet for human walk recognition, and it achieved 91.6% average accuracy on this dataset.

C. mmWave-based heart rate monitoring

Compared with wearable devices, mmWave radar-based methods can achieve better heart rate monitoring results because they do not require charging or wearing. The tiny displacements of human skin were detected by mmWave radar to monitor heart rate [60]. The mmWave-based rate monitoring dataset is listed in Table XII.

Mbeats dataset. This dataset was collected at the University of Oxford, United Kingdom [60]. Two volunteers collected millimeter radar wave signals in eight different postures (i.e., lie down straight, lie on the left side, lie on the right side, lie down but the upper body is sitting, sit up and put hands on legs, sit up and cross arms, sit up and cross legs, sit up and cross feet), and the collection lasted about 3 hours in total. A heart rate detection chest strap was used to collect the ground truth. From the provided website, this dataset can be downloaded directly. Researchers used FFT, PK, XCORR, and DNN models on this dataset, and DNN performed best, achieving 95.26% average accuracy.

Table XII. Visible-light-based heart rate monitoring

Dataset	Author	Introduction	Typical works	Access
MAHNOB-HCI	Soleymani <i>et al.</i> [102]	This dataset contains 27 people ECG signals, and each subject have 20 video clips for heart rate estimate.	[102], [103]	https://mahnob-db.eu/

Table XIII. Visible-light-based gesture recognition

Dataset	Author	Introduction	Gestures	Typical works	Access
IPN Hand	Benitez <i>et al.</i> [104]	This dataset contains 4,218 human-computer interaction gestures video clips and 1431 no gesture video clips.	13 gestures(one finger point, two fingers point, one finger click, two fingers click, throw up, throw down, throw left, throw right, open hand twice, one finger double click, two fingers double click, zoom in, zoom out, and no gesture)	[104]	https://github.com/GibranBenitez/IPN-hand
HaGRID	Kapitanov <i>et al.</i> [105]	This dataset contains 552,992 RGB images which were divided into 19 categories.	19 categories (have call, thumb up, thumb down, fist, mute gesture, ok gesture, rock gesture, stop gesture, inverse stop gesture, peace gesture, peace inverse gesture, one finger up, two fingers up, inverse two fingers up, three fingers up, three fingers up in second way, four fingers up, five fingers up, and no gesture)	[105]	https://github.com/hukenvs/hagrid

Table XIV. Visible-light-based activity recognition

Dataset	Author	Introduction	Activities	Typical works	Access
HMDB51	Kuehne <i>et al.</i> [106]	This dataset contains 6,766 video clips divided into 51 activities.	51 activities in five categories: facial movements including smile, laugh, speak, etc; facial movements with objects including smoke, eat, drink, etc; body movements including walk, run, jump, etc; body movements with objects including shoot, ride a horse, throwing ball, etc; interaction movements including kiss, hug, box, etc	[107], [108]	https://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/#Downloads
UCF101	Soomro <i>et al.</i> [109]	This dataset contains 13,320 video clips with 101 activities.	101 activities in five categories: body movements including jump, TaiChi, walk, etc; body movements with objects including apply eye makeup, apply lipstick, etc; interaction movements including hair cut, head massage, etc; sport movements including bench press, ride a bike, ride a horse, etc; play instruments movements including playing guitar, playing piano, playing violin, etc	[109], [108]	https://www.crcv.ucf.edu/data/UCF101.php
Kinetics400	Kay <i>et al.</i> [110]	This dataset contains 306,245 video clips containing 400 activities.	400 activities summarized as sports actions, repair actions, interaction actions, personal hygiene actions, cooking actions, eating actions, driving actions, etc	[110], [108]	https://github.com/cvdfoundation/kinetics-dataset

VI. VISIBLE-LIGHT-BASED CONTACTLESS SENSING

Visible light in the wavelength range of 400 to 700 nanometers can be captured by ordinary cameras. It has been exploited for the detection of physical and physiological indicators. Considering the large numbers of research achievements and datasets published in this area, this paper exclusively enumerates some popular and well-known datasets. For more comprehensive details, please refer to [111], [112]. The availability of datasets is as follows:

- Tasks have open datasets: heart rate monitoring, emotion recognition, depression detection.
- Tasks don't have open datasets: sleep monitoring.

The following subsections will review the open datasets.

A. Visible-light-based heart rate monitoring

Visible-light-based heart rate monitoring includes three methods: ballistocardiography (BCG), photoplethysmography (PPG), and remote PPG [36], [37]. Table XII lists a visible-light-based heart rate monitoring dataset.

MAHNOB-HCI. It was a multi-modal dataset collected by a research team at the University of Geneva [102]. It contained recordings of various data streams, including face videos, audio signals, eye gaze, and four physiological data types (GSR, ECG, skin temperature, and respiration amplitude). A total of 27 people sat 0.4 meters from the camera, and 20 frontal face videos were recorded for each subject. Li *et al.* [103] conducted a deep learning algorithm on this dataset and achieved a 6.87% average error. Researchers who want to use

Table XV. Visible-light-based emotion recognition

Dataset	Author	Introduction	Emotions	Typical works	Access
ENTERFACE '05	Martin <i>et al.</i> [113]	This dataset contains 1,166 video sequences of six different emotions.	Six (fear, surprise, anger, happiness, sadness, and disgust)	[113], [114]	http://www.enterface.net/enterface05/
AFEW	Dhall <i>et al.</i> [115]	This dataset contains 1,809 video clips collected from TV series or movies.	Seven (happiness, surprise, anger, disgust, fear, sadness, and neutral)	[115], [116]	https://cs.anu.edu.au/few/AFEW.html
AffectNet	Mollahosseini <i>et al.</i> [117]	This dataset collects 420,299 facial images in 11 emotions searched on the Internet.	Eleven (calm, happy, sad, surprised, afraid, disgusted, angry, disrespectful, blank expression, uncertain, and no face)	[116]	http://mohammadmahour.com/affectnet/
RAF-DB	Li <i>et al.</i> [118]	This dataset contains 29,672 facial images of 7 basic expressions and 11 compound expressions.	Seven basic (fear, neutral, happiness, anger, surprise, disgust, and sadness) and 11 compound (fearfully surprised, sadly fearful, disgustedly surprised, sadly disgusted, fearfully angry, angrily surprised, happily disgusted, sadly surprised, angrily disgusted, sadly angry, and happily surprised)	[118], [119]	http://www.whdeng.cn/raf/model1.html
FER+	Goodfellow <i>et al.</i> [120]	This dataset contains 32,298 facial images with 7 emotions.	Seven (neutral, disgust, anger, surprise, fear, sadness, and happiness)	[120]	https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data
SFEW	Dhall <i>et al.</i> [121]	This dataset contains 700 images selecting frames from the AFEW.	Seven (surprise, happiness, fear, anger, sadness, disgust, and neutral)	[121]	https://cs.anu.edu.au/few/AFEW.html

Table XVI. Visible-light-based depression detection

Dataset	Author	Introduction	Length	Typical works	Access
DAIC-WOZ	Ringeval <i>et al.</i> [122]	This dataset contains 189 video clips.	50hours	[122], [123]	https://dcapswoz.ict.usc.edu/daic-woz-database-download/
AVEC2013	Valstar <i>et al.</i> [124]	This dataset contains 340 video clips to test depression degree.	240hours	[124], [125]	http://avec2013-db.sspnet.eu/

this dataset need to submit an application at the website.

B. Visible-light-based gesture recognition

Visible-light-based gesture recognition datasets are listed in Table XIII.

IPN Hand. IPN Hand was collected by The University of Electro-Communications, Japan [104]. They designed 13 types of gestures, including 12 human-computer interaction gestures (one finger point, two finger point, one finger click, two fingers click, throw up, throw down, throw left, throw right, open hand twice, one finger double click, two fingers double click, zoom in, and zoom out) and no gesture. 50 volunteers participated in 28 different real-world scenarios that were recorded by an RGB camera. In total, they collected 4,218 human-computer interaction gesture video clips and 1,431 no-gesture video clips. Each video clip's duration ranges from 9 frames to 650 frames. They implemented 3D-CNN, Resnet-50, and ResNeXt-101 on this dataset. ResNeXt-101 performed with 86.32% accuracy, which is the best among them.

HaGRID. HaGRID was collected by SberDevices, Russia [105]. This dataset contains 552,992 RGB images, which

were divided into 19 categories (18 categories of gestures and no gestures). The 18 gestures were: call, thumb up, thumb down, fist, mute gesture, ok gesture, rock gesture, stop gesture, inverse stop gesture, peace gesture, peace inverse gesture, one finger up, two fingers up, inverse two fingers up, three fingers up, three fingers up in a second way, four fingers up, and five fingers up. Subjects stood within a range of 0.5 meters to 4 meters from the camera, made gestures and recorded photos. A total of 34,730 individuals collected gesture photos in various indoor environments. The provided website allows for the direct download of this dataset. They used 8 popular architectures to evaluate this dataset. ResNeXt-101 performed best and achieved 99.28% accuracy.

C. Visible-light-based activity recognition

Visible-light-based activity recognition datasets are listed in Table XIV.

HMDB51. This dataset was collected by the Karlsruhe Institution of Technology, Germany [106]. It contains a total of 6,766 video clips, which were classified into 51 actions. The dataset encompasses a total of 51 activities, which have

been grouped into five categories. These categories include facial movements (such as smiling, laughing, speaking), facial movements with objects (involving activities like smoking, eating, drinking), body movements (encompassing actions like walking, running, jumping), body movements with objects (comprising activities like shooting, riding a horse, throwing a ball), and interaction movements (covering actions like kissing, hugging, boxing). Each action contains at least 101 video clips. A group of volunteers watched video clips and provided the ground truth. This dataset can be obtained directly from the website provided.

UCF101. This dataset was collected at the University of Central Florida, USA [109]. It contains a total of 13,320 video clips. A total of 101 actions were included, and each action contained at least 100 video clips. The average video clip length is about 7 seconds. The actions were broadly categorized into five groups: body movements (such as jumping, TaiChi, walking), body movements with objects (involving activities like applying eye makeup, applying lipstick), interaction movements (including actions like hair cutting, head massage), sport movements (covering activities like bench press, riding a bike, riding a horse), and play instrument movements (encompassing actions like playing guitar, playing piano, playing violin). From the provided website, this dataset can be downloaded directly. They provided baseline action recognition results using the standard bag-of-words approach on this new dataset, with an overall performance of 44.5%.

Kinetics400. The Kinetics400 dataset was collected by Google [110]. This dataset contains a total of 306,245 video clips. It contains 400 actions, and each action contains at least 400 fragments. The actions were generally summarized as sports actions, repair actions, interaction actions, personal hygiene actions, cooking actions, eating actions, driving actions, etc. Volunteers recruited from the Internet label the data. The dataset can be obtained directly from the website provided. They updated and expanded this dataset, which is named Kinetics600 and Kinetics700, respectively. These two datasets can also be downloaded from the provided website. Carreira *et al.* [108] proposed a two-stream 3D-ConvNet for image classification. It achieved 78.7% accuracy on Kinetics400, 80.2% on HMDB-51 and 97.9% on UCF-101.

D. Visible-light-based emotion recognition

Human internal emotions can be conveyed by facial expressions [117]. Face acquisition, facial feature extraction, and feature classification are three common steps in the problem of facial expression recognition [118]. Datasets for Visible-light-based sensing for emotion recognition are listed in Table XV and are described in detail as follows.

INTERFACE '05. The INTERFACE '05 dataset was collected by a team at the University of Leuven in Belgium [113]. A collection of 1,166 video clips, featuring 6 different emotions (fear, surprise, anger, happiness, sadness, and disgust), were obtained from 42 volunteers. Muhammad *et al.* used active learning to distinguish images, achieving an accuracy of 84.12% on this dataset [114], [126]. This dataset is available directly from the provided website.

AFEW. It was a facial expression dataset, which contained short video clips of facial expressions close to the real environment [115]. This dataset had a total of 957 video clips collected from 37 categories of TV series or movies. The facial expression information in the videos is divided into seven expressions (neutral, surprise, fear, disgust, happiness, anger, and sadness). Avots *et al.* [116] used the information from the audio for emotion recognition. It performed with 50.2% accuracy on INTERFACE '05 and 46.6% accuracy on AFEW. Researchers who want this dataset can submit an application to the accessed website to get permission.

SFEW. It was collected by a research team at the University of Canberra in Australia and contained a total of 700 images [121]. The images were divided into six different emotions (surprise, happiness, fear, anger, sadness, disgust, and neutral). Borgalli *et al.* [127] proposed a deep learning-based facial expression recognition framework which achieved 40.78% accuracy. Researchers who want to use that dataset need to submit an application through the accessed website.

AffectNet. The AffectNet Facial Expression Dataset was collected by the University of Denver [117]. It included more than 420,299 facial images searched on the Internet. Images in this dataset were divided into 11 categories (calm, happy, sad, surprised, afraid, disgusted, angry, disrespectful, blank expression, uncertain, and no face). She *et al.* [128] proposed an end-to-end algorithm to recognize human facial expressions, and it achieved 63.11% accuracy on this dataset. It is available after applying to the author for permission.

RAF-DB. The RAF-DB dataset contained human facial expression images, which were collected by a research team at Peking University [118]. It contained 29,672 facial images in 7 basic expressions (fear, neutrality, happiness, anger, surprise, disgust, and sadness) or 11 compound expressions (fearfully surprised, sadly fearful, disgustedly surprised, sadly disgusted, fearfully angry, angrily surprised, happily disgusted, sadly surprised, angrily disgusted, sadly angry, and happily surprised). Apart from the data collection, Li *et al.* introduced a deep learning algorithm with 74.20% accuracy for basic expressions and 44.55% for compound expression recognition. In addition, Li *et al.* [119] proposed an attention-based CNN to recognize facial expression images being blocked, and the accuracy was 80.54%. The authors who want this dataset need to apply for permission on the accessed website.

FER+. The FER+ dataset had a total of 32,298 images of the human face, which were composed of 7 different facial expressions (neutral, disgust, anger, surprise, fear, sadness, and happiness) [120]. The dataset was split into two parts: 28,709 images for training, and the remainder for testing. It is readily available for download from the website.

E. Visible-light-based depression detection

A patient's mental condition is usually evaluated by psychologists and psychiatrists by observing their facial expressions and sound cues [124], [125]. Table XVI lists the datasets for video-based sensing for depression detection.

AVEC 2013. The AVEC 2013 dataset comprised 340 video clips with participants answering questions or reading aloud.

The AVEC Inhibitory Challenge predicted depression using BDI-ii scores, which consisted of 21 standardized questions [124]. The AVEC 2014, a subset of the AVEC 2013, included 150 video clips for depression level detection [125]. Williamson *et al.* [129] proposed a machine learning-based algorithm, and the result achieved 7.42 on root-mean-squared error (RMSE) and 5.75 on mean absolute error (MAE). Both datasets can be used after submitting an application for permission.

DAIC-WOZ. The DAIC-WOZ dataset, for the AVEC 2017 challenge, comprises 189 clinical interviews conducted at the University of Southern California [122]. These interviews, conducted in the greater Los Angeles metropolitan area, aim to identify depression symptoms. The dataset spans 50 hours, with interviews ranging from 7 to 33 minutes. It includes both training and testing datasets, involving 107 and 35 participants, respectively. Yang *et al.* [123] proposed a deep-learning algorithm for depression prediction, while Gong *et al.* [130] introduced a context-aware topic model using audio, video, and text data to predict depression levels. Access to this dataset requires an application through the website.

VII. IR-BASED CONTACTLESS SENSING

The IR camera is able to measure both the heat and movement of the objects. To the best of our knowledge, here is the disclosure of datasets used for various tasks:

- Tasks have open datasets: activity recognition, emotion recognition, and face recognition.
- Tasks don't have open datasets: sleep monitoring, and respiration monitoring.

The following subsections will discuss the open datasets in IR-based contactless sensing work.

A. IR-based activity recognition

Human activity analysis is a very important field in computer vision. It aims to generate a 3D human model through images [131]. Datasets for IR-based activity recognition are listed in Table XVII and detailed as follows.

MAHAD-D-MOCAP. The MAHAD-D-MOCAP dataset was collected by a research team at Oklahoma State University, and it included 12 participants performing 11 different movements of IR images [131]. Researchers can get it directly from the website listed on the table.

InfAR dataset. They collected 12 human activities using an IR camera [107]. The activities were waving with one hand, waving with two hands, clapping hands, walking, jogging, jumping, skipping, shaking hands, hugging, pushing, punching, and fighting. Every activity contained 50 video clips, which were completed by 40 volunteers. Each video clip is approximately 4 seconds. They used CNN on this dataset, and it achieved 76.66% average precision. The provided website allows for the direct download of this dataset.

B. IR-based gesture recognition

A dataset for IR-based gesture recognition is listed in Table XVIII and detailed as follows.

Multi-Modal Hand Gesture dataset. This dataset was collected at the Universidad Politecnica de Madrid, Spain [132]. It contains two modalities of information, including infrared images of the hand and skeleton information. Since this section mainly focuses on infrared-based datasets, only the infrared image modalities are introduced. There are a total of 16 static gesture poses and 4 dynamic gestures, completed by 25 volunteers. For dynamic gestures, each volunteer repeated each gesture 30 times, and for static gestures, 200 frames of images are obtained for each gesture, and then divided into 20 groups of static gesture sequences. From the provided website, this dataset can be downloaded directly.

C. IR-based emotion recognition

Distinguishing different human expressions is very useful for biometrics and human-computer interaction [133]. Datasets of IR-based sensing for emotion recognition are listed in Table XIX and described in detail as follows.

Oulu-casia. Oulu-casia was collected by a collaborative research team at the University of Oulu in Finland and the Chinese Academy of Sciences [133]. It contained 80 (50 Finns and 30 Chinese) people from 23 to 58 years of age performing 6 different expressions (happiness, sadness, surprise, disgust, anger, and fear). The total count of images is 2,400, and all the photos were taken in three different situations (weak, dark, and normal). The verification algorithm by the author achieved 73.79% accuracy in normal, 70.63% accuracy in weak, and 69.66% accuracy in dark. This dataset is available after applying to the author to get access.

CK+. The CK+ (Cohn-Kanade) dataset was collected by a research team at Carnegie Mellon University [134]. It had a total of 593 video sequences, and the number of volunteers was 123. These video sequences were divided into seven feelings (surprise, contempt, fear, happiness, disgust, sadness, and anger). The authors extracted facial visual features from the images and classified emotions to test their dataset. Shaees *et al.* [135] proposed a transfer learning method for facial expression recognition, achieving 98.3% accuracy on the CK+ dataset. This dataset is available after submitting an application on the website for permission.

D. IR-based face recognition

Datasets for IR-based sensing for face recognition are listed in Table XX and described in detail as follows.

CBSR NIR. It was a near-IR face dataset collected by the Chinese Academy of Sciences [136]. It contained 3,940 near-IR face images of 197 people. Researchers verified the dataset in their university office building, and achieved a 0.3% error recognition rate. Anyone who wants to use this dataset can download it from the website.

CASIA NIL-VIS 2.0. This dataset was collected by the Chinese Academy of Sciences from 725 people using visible and near-IR cameras, with vision and near-IR facial images of each person [137]. The website provides a download link for this dataset, which can be downloaded.

Terravic. The Terravic Face IR Database was a public IR face dataset [139]. It provided a collection of 21,676 facial

Table XVII. IR-based activity recognition

Dataset	Author	Introduction	Activities	Typical works	Access
MAHAD-D-MOCAP	Lannan <i>et al.</i> [131]	This dataset collects 12 people performing 11 different movements of IR images.	11 activities (motions with both upper and lower limbs, activities with the upper limbs, actions with the lower limbs)	[131]	http://vcip1-okstate.org/pbvs/bench/
InfAR dataset	Gao <i>et al.</i> [107]	This dataset contains 24,000 IR video clips which were divided into 12 activities.	12 activities (one hand, wave with two hands, clap hands, walk, jog, jump, skip, shake hands, hug, push, punch, and fight)	[107]	http://scie.cqupt.edu.cn/ivp/Publications.html

Table XVIII. IR-based gesture recognition

Dataset	Author	Introduction	Gestures	Typical works	Access
Multi-Modal Hand Gesture dataset	Mantecon <i>et al.</i> [132]	This dataset contains 16 static gesture poses IR images and 4 dynamic gestures videos performed by 25 volunteers.	16 static gestures(English letter gestures L and C, fist moved, index, ok gesture, heavy, hang, two, three, four, five, palm, down, palm moved, palm up, up), 4 dynamic gestures(go down, go left, go right, go up)	[132]	http://www.gti.ssr.upm.es/data/MultiModalHandGesture_dataset

Table XIX. IR-based emotion recognition

Dataset	Author	Introduction	Emotions	Typical works	Access
Oulu-Casia	Zhao <i>et al.</i> [133]	This dataset contains 2,880 video sequences performed by 80 people.	Six emotions (happiness, sadness, surprise, disgust, anger, and fear)	[133]	http://www.cse.oulu.fi/MVG/Downloads
CK+	Lucey <i>et al.</i> [134]	This dataset contains 593 video sequences performed by 123 volunteers.	Seven emotions (surprise, contempt, fear, happiness, disgust, sadness, and anger)	[134], [135]	http://vasc.ri.cmu.edu/idb/html/face/facial_expression/

Table XX. IR-based face recognition

Dataset	Author	Introduction	People	Typical works	Access
CBSR NIR	Li <i>et al.</i> [136]	This dataset contains 3,940 near IR face images.	197 people	[136]	http://vcip1-okstate.org/pbvs/bench/Data/07/download.html
CASIA NIL-VIS 2.0	Li <i>et al.</i> [137]	This dataset contains 17,580 near IR and visible-light face images.	725 people	[137], [138]	http://www.cbsr.ia.ac.cn/english/NIR-VIS-2.0-Dataset.html
Terravic	Benamara <i>et al.</i> [139]	This dataset contains 21,676 near IR face images.	20 people	[139]	http://vcip1-okstate.org/pbvs/bench/Data/04/download.html

Table XXI. Audio-based emotion recognition

Dataset	Author	Introduction	Emotions	Typical works	Access
Berlin Emo-db	Burkhardt <i>et al.</i> [140]	This dataset contains 535 German voice data in seven different moods.	Seven emotions (bored, happy, disgusted, angry, sad, afraid, and neutral)	[140], [141]	https://www.kaggle.com/datasets/piyushagni5/berlin-database-of-emotional-speech-emodb

Table XXII. Audio-based speech recognition

Dataset	Author	Introduction	Length	Typical works	Access
LibriSpeech	Panayotov <i>et al.</i> [142]	This dataset which contains English speech data is collected from English audio books.	982.1 hours	[142], [143]	https://www.openslr.org/12/
Common Voice	Ardila <i>et al.</i> [144]	This dataset has 29 languages which contain 58,250 pieces voice data with 40,000 contributors.	2,454 hours	[144]	https://voice.mozilla.org/en/datasets

images from 20 people. Researchers proposed a human face recognition method and tested it on their dataset. This dataset can be downloaded from the accessed website.

VIII. AUDIO-BASED CONTACTLESS SENSING

Audio signals are primarily collected through microphones, which work by converting sound wave signals into electrical signals. Acoustic signals carry a large amount of usable information during propagation, which can be used for contactless sensing tasks. Some related datasets are as follows:

- Tasks have open datasets: emotion recognition, and speech recognition.
- Tasks don't have open datasets: respiration monitoring.

The following subsections will review the open datasets in audio-based contactless sensing works.

A. Audio-based emotion recognition

Extracting emotional patterns from voice messages has become a hot topic, especially in the field of human-computer interaction. The process of emotion recognition from speech mainly contains two steps: feature extraction and feature classification. The first step refers to extracting some voice features, and then a classifier is used to distinguish different emotions [145]. An open dataset for audio-based sensing for emotion recognition is listed in Table XXI.

Berlin Emo-db. The Berlin Emo-db dataset was collected by a research team at the University of Berlin [140]. This dataset contained 535 German voice recordings presented by 5 male and 5 female professional actors posing 7 different emotions (bored, happy, disgusted, angry, sad, afraid, and neutral). Volunteers simulated emotions and engaged in daily conversations within an anechoic chamber, while researchers recorded the audio. Zhao *et al.* [141] introduced a deep learning framework for speech emotion recognition, achieving 95.33% accuracy in speaker-dependent experiments and 95.89% accuracy in speaker-independent experiments. Anyone who wants to use this dataset can download it directly from the website.

B. Audio-based speech recognition

The progress of the computer algorithm converting the speech signal into words is the definition of speech recognition [146]. There are two open datasets for acoustic-based speech recognition, as listed in Table XXII.

LibriSpeech. The LibriSpeech dataset, collected by a research team at Johns Hopkins University, comprises 982.1 hours of English speech data [142]. The dataset is categorized into clean (475 hours of speech) and other (507.1 hours of speech) based on speaker presentation levels. Baevski *et al.* [143] achieved a Word Error Rate (WER) of 1.8 in clean and 3.3 in other cases using a self-supervised speech recognition method. Researchers who want to use this dataset can get it directly from the website.

Common Voice. The Common Voice dataset was collected by a research team at Indiana University [144]. It contained about 2,454 hours (1,965 hours verified) of recorded voice data from more than 40,000 contributors in 29 different languages. Researchers need to submit an application to get permission.

IX. DISCUSSION

In recent years, technology based on contactless sensing has made great progress. Some open datasets have been released for the community, yet there are some emerging challenges and problems to be solved [9], as demonstrated in Fig. 2 and will be detailed in the subsections.

A. Dataset availability

Previous works have discussed the reasons for the absence and unavailability of public datasets [23], [24]. First, some datasets contain sensitive personal information, such as ID numbers, sex, etc. In order to protect the privacy of the parties involved and prevent data from being misused or stolen, researchers may choose not to release these datasets. Second, some countries or regions have strict legal and regulatory restrictions on certain types of data. These restrictions may include data usage purposes, storage requirements, transfer restrictions, etc. For compliance with relevant regulations, the dataset may not be fully available. Last but not least, some datasets may involve the protection of copyright or intellectual property, such as music, movies, literary works, etc. Making these datasets public without authorization may involve copyright infringement issues.

The reasons of lacking datasets for sleep monitoring can be considered in the following aspects. The sleep monitoring devices are difficult to wear and may cause discomfort to the subject, so the collected data may not reflect the subject's natural sleep data [30]. At the same time, the subjects' sleep data contains a lot of personal information, possibly because of which some regulations prevent the data from being disclosed.

There are several factors that have limited the availability of respiratory monitoring datasets. First, collecting data on breathing is challenging due to its inherent variability, which can lead to misinterpretation of the ground truth [147]. While monitoring breathing in controlled environments like hospitals is relatively straightforward, there are significant challenges in monitoring breathing over extended periods in daily environments [148]. Without the guidance of medical professionals, the accuracy of breathing monitoring may be compromised. In addition, the lack of standardization for scenarios requiring respiratory monitoring leads to variations in monitoring protocols [148]. As a result, the performance metrics obtained from different studies are not directly comparable as they are acquired under varying testing conditions.

Having practical, reliable, and accessible datasets is very important for the scientific progress of contactless sensing. One solution is to encourage users to contribute their own collected data from in-home devices, creating larger and more diverse datasets that are close to real-world scenarios. Incentivizing authors to share their datasets for validation purposes is also helpful in creating a more balanced and comprehensive dataset. However, it is also very important to consider factors like privacy and security when using open datasets.

B. Robustness and accuracy

Robustness is a very important performance metric for contactless sensing technology. Strong robustness determines

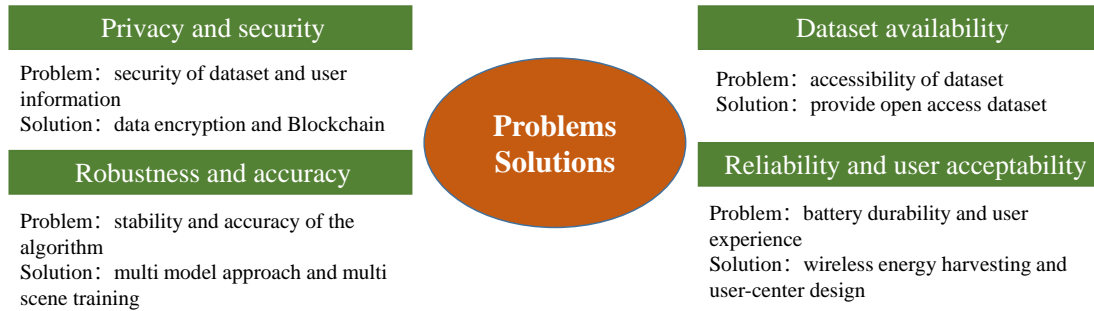


Fig. 2. Potential problems and related solutions.

whether the method or system can be applied with stable performance in different application scenarios [9]. The performance of an algorithm is generally relevant to the dataset's collection scenario. For example, some video datasets were collected in an experimental environment, so the background was very monotonic. It may have performed perfectly in an experimental scene but not be suitable for a realistic application. In addition, for different contactless sensing technologies, there are different factors that may affect the accuracy and robustness [6]. It is obvious that acoustic-based sensing technology will be affected by sound signals, and vision-based sensing technology will be affected by the light. For wireless signals, low sample rate, narrow channel bandwidth, movement of the user, and impact of environmental noise were listed as the factors that affect the wireless sensing performance [149]. Dealing with more complicated situations, multiple scenes, and different backgrounds is another challenge. Several multimodal approaches have been proposed, and they are able to promote robustness and accuracy [150].

The authenticity and availability of the public datasets for contactless sensing are also worth discussing. Transparency regarding data authenticity and collection methods is essential for ensuring the credibility and reproducibility of research findings. While there may be limitations in detailing the dataset in academic papers due to space constraints, data repositories should include comprehensive descriptions and annotations for each sample. Paying particular attention to noisy or anomalous data instances will aid future researchers in understanding and interpreting the dataset accurately. Moreover, by documenting the authenticity of the data and providing clear information on the methods used for dataset collection, researchers can enhance transparency and uphold the integrity of scientific research. Overall, documenting the authenticity and collection methods of data promotes transparency and contributes to the integrity of scientific research.

C. Privacy and security

There are many papers on contactless sensing that discuss privacy and security issues [6], [9]. The process of obtaining datasets inevitably contains a large amount of privacy information. Hacking of these devices may lead to the disclosure of personal information and personal privacy. For example, WiFi devices used for surfing the Internet may be hacked by some password-cracking software. Vision-based and acoustic-based

technologies contain personal images and sound information, and the loss of these informations may be used for illegal purposes. The security of the data also deserves attention. If the data is attacked or changed maliciously during transmission, these malicious acts of destruction will influence or delay the diagnostic process and frustrate the patient. For example, the data and parameters captured by contactless sensing technology, such as heart rate [102], [151] or respiratory rate [82], may be changed in the process of transmission, which may lead users to misjudge their health condition, and the doctor may give incorrect treatment. Recently, blockchain has been reported as a promising method to store or transmit data because of its decentralized system and shared ledger functionalities [152].

There are also some ethical concerns regarding the retrieval and usage of data from the Internet. In this scenario, it is difficult to contact the involved people and obtain their explicit consent. As a result, there is a risk of unintentionally infringing upon individuals' privacy rights. Researchers should consider alternative approaches, such as anonymizing data, to mitigate these risks. In addition, utilizing robust data protection measures and adhering to relevant legal frameworks is crucial for safeguarding personal privacy. While open access to datasets can facilitate research progress, it is imperative to consider the long-term implications. In an information-driven society, vast amounts of readily accessible data, including sensitive personal information, pose potential risks for malicious exploitation and unauthorized access. Researchers using online data sources should prioritize ethical considerations. Striking a balance between promoting innovation and protecting privacy is crucial when deciding whether to publicly release the datasets. Addressing these ethical concerns ensures responsible data practices and promotes knowledge advancement while respecting individuals' privacy.

D. Reliability and user acceptability

Constant contactless monitoring of human physical and physiological signs needs a reliable battery system. Devices such as WiFi devices, cameras, and microphones need batteries or to be plugged in. The sudden run-out of the battery will cause inconvenience to the users, and it will destroy the monitoring process. Wireless energy harvesting was mentioned in [6] to deal with battery capacity problems. On the other hand, reducing the consumption of energy is also an effective way to

increase the reliability of the system. Although higher accuracy requires more energy, in the future, the low energy cost and better performance of contactless sensing technology will be in high demand to promote reliability and user satisfaction [39]. In addition, user acceptability for contactless sensing technology also deserves our attention. It is a very important factor in judging whether the recognition or monitoring system is helpful. Satisfying user needs will allow the technology to be widely used [153]. Recently, there have been a variety of personal-custom types of products emerging. In [6], user-centered design was mentioned to satisfy user requirements, and they suggested researchers consider the acceptability of users during the design of a healthcare system.

E. Suggestions for establishing an open dataset

Benefiting from the above discussion, we come up with suggestions for establishing an open dataset. First, a public dataset should provide detailed documentation or metadata that describes the structure, fields, and meaning of the dataset. Such descriptions can help users understand how data is organized and be able to quickly find the information they need. The public open dataset is also suggested to be verified and audited to ensure data accuracy and reliability. This can be achieved through multi-source cross-validation, data cleaning, error correction, and other methods. This is the most basic guarantee of dataset quality. Then, publicly available download links should be provided and the links are suggested to be maintained in a timely manner. High-quality public datasets should be maintained by relevant agencies or groups to ensure that the dataset has been updated and is available. Whether adding new data or fixing errors, regular maintenance and updates provide convenience to researchers working with these datasets. What's more, the baseline of the dataset under the classic models and some performance metrics should also be provided for reference. In addition, in order to protect the privacy information of volunteers, the creators of public datasets should obtain the volunteers' permission or certification from relevant agencies. This can be achieved by complying with privacy protection regulations and using appropriate authorization mechanisms. The volunteer-related information can also be deleted or anonymized to ensure that nothing in the data directly or indirectly reveals the identity of an individual. Last but not least, cross-platform compatibility is also important, and websites that expose the open datasets should provide download options in multiple formats to ensure that users of different operating systems (mainly including Linux and Windows) can easily access and use the provided samples.

X. CONCLUSIONS

Contactless sensing has attracted increasing concern in recent years, as it is able to acquire physical and physiological signs in a remotely noninvasive way and has great potential for multi-target monitoring. This paper extensively collected the public datasets in this field, i.e., contactless sensing methods for physical and physiological sign monitoring. The collection

equipment, scenarios, sample counts, data formats, accessible links, related works and other issues of these datasets were comprehensively summarized. We also discussed the issues of availability, privacy, and accuracy of the publicly available datasets that should attract increasing attention from the community and outlined some potential solutions. This paper further summarized several required conditions that an expected dataset is suggested to be satisfied. Our summary of open datasets is expected to be beneficial for the research of contactless-based sensing.

XI. CONFLICT OF INTEREST

The authors declare that they don't have conflict of interest.

REFERENCES

- [1] J. Shang and J. Wu, "Fine-grained vital signs estimation using commercial Wi-Fi devices," in *Proceedings of the Eighth Wireless of the Students, by the Students, and for the Students Workshop*, 2016, pp. 30–32.
- [2] S. M. Lee and D. Lee, "Opportunities and challenges for contactless healthcare services in the post-COVID-19 Era," *Technological Forecasting and Social Change*, vol. 167, p. 120712, 2021.
- [3] L. P. Malasinghe, N. Ramzan, and K. Dahal, "Remote patient monitoring: a comprehensive study," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 57–76, 2019.
- [4] S. Huynh, H.-P. Tan, and Y. Lee, "Towards unobtrusive mental well-being monitoring for independent-living elderly," in *Proceedings of the 4th International on Workshop on Physical Analytics*, 2017, pp. 1–6.
- [5] H. Rohmetra, N. Raghunath, P. Narang, V. Chamola, M. Guizani, and N. R. Lakkaniga, "Ai-enabled remote monitoring of vital signs for COVID-19: methods, prospects and challenges," *Computing*, pp. 1–27, 2021.
- [6] M. Nouman, S. Y. Khoo, M. P. Mahmud, and A. Z. Kouzani, "Recent advances in contactless sensing technologies for mental health monitoring," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 274–297, 2021.
- [7] M. Kebe, R. Gadhaifi, B. Mohammad, M. Sanduleanu, H. Saleh, and M. Al-Qutayri, "Human vital signs detection methods and potential using radars: A review," *Sensors*, vol. 20, no. 5, p. 1454, 2020.
- [8] F. Al Hossain, A. A. Lover, G. A. Corey, N. G. Reich, and T. Rahman, "Flusense: a contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–28, 2020.
- [9] Y. Ge, A. Taha, S. A. Shah, K. Dashtipour, S. Zhu, J. M. Cooper, Q. Abbasi, and M. Imran, "Contactless WiFi sensing and monitoring for future healthcare-emerging trends, challenges and opportunities," *IEEE Reviews in Biomedical Engineering*, 2022.
- [10] A. A. Alnuaim, M. Zakariah, P. K. Shukla, A. Alhadlaq, W. A. Hatamleh, H. Tarazi, R. Sureshbabu, and R. Ratna, "Human-computer interaction for recognizing speech emotions using multilayer perceptron classifier," *Journal of Healthcare Engineering*, vol. 2022, 2022.
- [11] T. Vuletic, A. Duffy, L. Hay, C. McTeague, G. Campbell, and M. Grealy, "Systematic literature review of hand gestures used in human computer interaction interfaces," *International Journal of Human-Computer Studies*, vol. 129, pp. 74–94, 2019.
- [12] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sensors Journal*, vol. 15, no. 3, pp. 1321–1330, 2014.
- [13] G. Sarapata, Y. Dushin, G. Morinan, J. Ong, S. Budhdeo, B. Kainz, and J. O'Keeffe, "Video-based activity recognition for automated motor assessment of parkinson's disease," *IEEE Journal of Biomedical and Health Informatics*, 2023.
- [14] R. Sadeghian, J. D. Schaffer, and S. A. Zahorian, "Towards an automatic speech-based diagnostic test for Alzheimers disease," *Frontiers in Computer Science*, vol. 3, p. 624594, 2021.
- [15] B. Yu, Y. Wang, K. Niu, Y. Zeng, T. Gu, L. Wang, C. Guan, and D. Zhang, "Wifi-sleep: sleep stage monitoring using commodity Wi-Fi devices," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 13900–13913, 2021.

- [16] Y.-G. Park, S. Lee, and J.-U. Park, "Recent progress in wireless sensors for wearable electronics," *Sensors*, vol. 19, no. 20, p. 4353, 2019.
- [17] T. R. Ray, J. Choi, A. J. Bhandokar, S. Krishnan, P. Gutruf, L. Tian, R. Ghaffari, and J. A. Rogers, "Bio-integrated wearable systems: a comprehensive review," *Chemical Reviews*, vol. 119, no. 8, pp. 5461–5533, 2019.
- [18] A. Yadav, A. Singh, M. K. Dutta, and C. M. Travieso, "Machine learning-based classification of cardiac diseases from PCG recorded heart sounds," *Neural Computing and Applications*, vol. 32, pp. 17 843–17 856, 2020.
- [19] J. Chen, Z. Guo, X. Xu, L.-b. Zhang, Y. Teng, Y. Chen, M. Woźniak, and W. Wang, "A robust deep learning framework based on spectrograms for heart sound classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2023.
- [20] J. Chen, S. Sun, L.-b. Zhang, B. Yang, and W. Wang, "Compressed sensing framework for heart sound acquisition in internet of medical things," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 2000–2009, 2022.
- [21] Y. Tian, G.-H. Lee, H. He, C.-Y. Hsu, and D. Katabi, "Rf-based fall monitoring using convolutional neural networks," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–24, 2018.
- [22] K. T. Nahiyan and M. A. R. Ahad, "Contactless monitoring for healthcare applications," in *Vision, Sensing and Analytics: Integrative Approaches*. Springer, 2021, pp. 243–265.
- [23] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: a big data-ai integration perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328–1347, 2019.
- [24] D.-T. Dang-Nguyen, L. Zhou, R. Gupta, M. Riegler, and C. Gurrin, "Building a disclosed lifelog dataset: challenges, principles and processes," in *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, 2017, pp. 1–6.
- [25] M. Feidakis, T. Daradoumis, and S. Caballé, "Emotion measurement in intelligent tutoring systems: what, when and how to measure," in *2011 Third International Conference on Intelligent Networking and Collaborative Systems*. IEEE, 2011, pp. 807–812.
- [26] K. Byron, S. Terranova, and S. Nowicki Jr, "Nonverbal emotion recognition and salespersons: Linking ability to perceived and actual success," *Journal of Applied Social Psychology*, vol. 37, no. 11, pp. 2600–2619, 2007.
- [27] A. Dziedzickis, A. Kaklauskas, and V. Bucinskas, "Human emotion recognition: Review of sensors and methods," *Sensors*, vol. 20, no. 3, p. 592, 2020.
- [28] C. Massaroni, A. Nicolo, M. Sacchetti, and E. Schena, "Contactless methods for measuring respiratory rate: A review," *IEEE Sensors Journal*, vol. 21, no. 11, pp. 12 821–12 839, 2020.
- [29] K. A. I. Aboalayon, M. Faezipour, W. S. Almuhammadi, and S. Moslehpour, "Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation," *Entropy*, vol. 18, no. 9, p. 272, 2016.
- [30] X. Li, Y. Gong, X. Jin, and P. Shang, "Sleep posture recognition based on machine learning: A systematic review," *Pervasive and Mobile Computing*, p. 101752, 2023.
- [31] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–40, 2021.
- [32] R. Grzeszick, J. M. Lenk, F. M. Rueda, G. A. Fink, S. Feldhorst, and M. Ten Hompel, "Deep neural network based human activity recognition for the order picking process," in *Proceedings of the 4th International Workshop on Sensor-based Activity Recognition and Interaction*, 2017, pp. 1–6.
- [33] K. Murakami and H. Taguchi, "Gesture recognition using recurrent neural networks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1991, pp. 237–242.
- [34] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 3, pp. 311–324, 2007.
- [35] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "SignFi: Sign language recognition using WiFi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–21, 2018.
- [36] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, "Video-based heart rate measurement: Recent advances and future prospects," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 10, pp. 3600–3615, 2018.
- [37] M. A. Hassan, A. S. Malik, D. Fofi, N. Saad, B. Karasfi, Y. S. Ali, and F. Meriaudeau, "Heart rate estimation using facial video: A review," *Biomedical Signal Processing and Control*, vol. 38, pp. 346–360, 2017.
- [38] W. Wang, X. Yu, B. Fang, Y. Zhao, Y. Chen, W. Wei, and J. Chen, "Cross-modality LGE-CMR segmentation using image-to-image translation based data augmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 4, pp. 2367–2375, 2023.
- [39] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.
- [40] H. Obeidat, W. Shuaieb, O. Obeidat, and R. Abd-Alhameed, "A review of indoor localization techniques and wireless technologies," *Wireless Personal Communications*, vol. 119, pp. 289–327, 2021.
- [41] M. Morales, S. Scherer, and R. Levitan, "A cross-modal review of indicators for depression detection systems," in *Proceedings of the fourth workshop on computational linguistics and clinical psychology From linguistic signal to clinical reality*, 2017, pp. 1–12.
- [42] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A survey," *Sensors*, vol. 20, no. 2, p. 342, 2020.
- [43] W. Ghai and N. Singh, "Literature review on automatic speech recognition," *International Journal of Computer Applications*, vol. 41, no. 8, 2012.
- [44] W. Taylor, Q. H. Abbasi, K. Dashtipour, S. Ansari, S. A. Shah, A. Khalid, and M. A. Imran, "A review of the state of the art in non-contact sensing for COVID-19," *Sensors*, vol. 20, no. 19, p. 5665, 2020.
- [45] D. Xu, Y. Wang, B. Xiong, and T. Li, "MEMS-based thermoelectric infrared sensors: A review," *Frontiers of Mechanical Engineering*, vol. 12, pp. 557–566, 2017.
- [46] D. Hollosi, G. Nagy, R. Rodigast, S. Goetze, and P. Cousin, "Enhancing wireless sensor networks with acoustic sensing technology: use cases, applications & experiments," in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*. IEEE, 2013, pp. 335–342.
- [47] D. Yang, Z. Zhu, J. Zhang, and B. Liang, "The overview of human localization and vital sign signal measurement using handheld IR-UWB through-wall radar," *Sensors*, vol. 21, no. 2, p. 402, 2021.
- [48] J. Zhang, R. Xi, Y. He, Y. Sun, X. Guo, W. Wang, X. Na, Y. Liu, Z. Shi, and T. Gu, "A survey of mmwave-based human sensing: Technology, platforms and applications," *IEEE Communications Surveys & Tutorials*, 2023.
- [49] T. G. Mondal and M. R. Jahanshahi, "Applications of computer vision-based structural health monitoring and condition assessment in future smart cities," *The rise of smart cities*, pp. 193–221, 2022.
- [50] B. Yang, X. Li, Y. Hou, A. Meier, X. Cheng, J.-H. Choi, F. Wang, H. Wang, A. Wagner, D. Yan *et al.*, "Non-invasive (non-contact) measurements of human thermal physiology signals and thermal comfort/discomfort poses-a review," *Energy and Buildings*, vol. 224, p. 110261, 2020.
- [51] C. Cai, R. Zheng, and M. Hu, "A survey on acoustic sensing," *ArXiv Preprint ArXiv:1901.03450*, 2019.
- [52] J. Chen, W. Wang, B. Fang, Y. Liu, K. Yu, V. C. M. Leung, and X. Hu, "Digital twin empowered wireless healthcare monitoring for smart home," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 11, pp. 3662–3676, 2023.
- [53] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2. 0: Passive human tracking with a single Wi-Fi link," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 350–361.
- [54] X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-sleep: Contactless sleep monitoring via WiFi signals," in *2014 IEEE Real-Time Systems Symposium*. IEEE, 2014, pp. 346–355.
- [55] L. Cheng, A. Zhao, K. Wang, H. Li, Y. Wang, and R. Chang, "Activity recognition and localization based on ubw indoor positioning system and machine learning," in *2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2020, pp. 0528–0533.
- [56] Y. Lee, J.-Y. Park, Y.-W. Choi, H.-K. Park, S.-H. Cho, S. H. Cho, and Y.-H. Lim, "A novel non-contact heart rate monitor using impulse-radio ultra-wideband (ir-uw) radar technology," *Scientific Reports*, vol. 8, no. 1, p. 13053, 2018.
- [57] S. Ahmed and S. H. Cho, "Hand gesture recognition using an ir-uw radar with an inception module-based classifier," *Sensors*, vol. 20, no. 2, p. 564, 2020.
- [58] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a

- millimeter-wave radar,” in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, 2019, pp. 51–56.
- [59] J. Lien, N. Gillian, M. E. Karagozler, P. Amihoud, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, “Soli: Ubiquitous gesture sensing with millimeter wave radar,” *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–19, 2016.
- [60] P. Zhao, C. X. Lu, B. Wang, C. Chen, L. Xie, M. Wang, N. Trigoni, and A. Markham, “Heart rate sensing with a robot mounted mmwave radar,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2812–2818.
- [61] S. Kwon, H. Kim, and K. S. Park, “Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone,” in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 2174–2177.
- [62] H. Kaya, F. Gürpınar, and A. A. Salah, “Video-based emotion recognition in the wild using deep transfer learning and score fusion,” *Image and Vision Computing*, vol. 65, pp. 66–75, 2017.
- [63] E. Magdalena Nowara, T. K. Marks, H. Mansour, and A. Veer-araghavan, “Sparseppg: Towards driver monitoring using camera-based vital signs estimation in near-infrared,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1272–1281.
- [64] I. A. Cruz-Albarran, J. P. Benítez-Rangel, R. A. Osornio-Rios, and L. A. Morales-Hernandez, “Human emotions detection based on a smart-thermal system of thermographic images,” *Infrared Physics & Technology*, vol. 81, pp. 250–261, 2017.
- [65] K. Wang, N. An, B. N. Li, Y. Zhang, and L. Li, “Speech emotion recognition using fourier parameters,” *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 69–75, 2015.
- [66] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates *et al.*, “Deep speech: Scaling up end-to-end speech recognition,” *ArXiv Preprint ArXiv:1412.5567*, 2014.
- [67] S. Zhang, S. Zhang, T. Huang, W. Gao, and Q. Tian, “Learning affective features with a hybrid deep model for audio-visual emotion recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 3030–3043, 2017.
- [68] L. Schoneveld, A. Othmani, and H. Abdelkawy, “Leveraging recent advances in deep learning for audio-visual emotion recognition,” *Pattern Recognition Letters*, vol. 146, pp. 1–7, 2021.
- [69] H. Zou, J. Yang, H. Prasanna Das, H. Liu, Y. Zhou, and C. J. Spanos, “Wifi and vision multimodal learning for accurate and robust device-free human activity recognition,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.
- [70] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, “Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi,” in *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2017, pp. 1–10.
- [71] L. Tian, L. Chen, Z. Xu, and Z. Chen, “An angle recognition algorithm for tracking moving targets using WiFi signals with adaptive spatiotemporal clustering,” *Sensors*, vol. 22, no. 1, p. 276, 2022.
- [72] E. Lohan, J. Torres-Sospedra, P. Richter, H. Leppäkoski, J. Huerta, and A. Cramariuc, “Crowdsourced WiFi-fingerprinting database and benchmark software for indoor positioning,” *Zenodo Repos*, 2017.
- [73] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, “Ujiindoorloc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems,” in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2014, pp. 261–270.
- [74] K. S. Kim, S. Lee, and K. Huang, “A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on Wi-Fi fingerprinting,” *Big Data Analytics*, vol. 3, no. 1, pp. 1–17, 2018.
- [75] S. Bozkurt, G. Elibol, S. Gunal, and U. Yayan, “A comparative study on machine learning algorithms for indoor positioning,” in *2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*. IEEE, 2015, pp. 1–8.
- [76] L. E. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.
- [77] J. G. Rohra, B. Perumal, S. J. Narayanan, P. Thakur, and R. B. Bhatt, “User localization in an indoor environment using fuzzy hybrid of particle swarm optimization & gravitational search algorithm with neural networks,” in *Proceedings of Sixth International Conference on Soft Computing for Problem Solving*. Springer, 2017, pp. 286–295.
- [78] R. Gomes, M. Ahsan, and A. Denton, “Random forest classifier in SDN framework for user-based indoor localization,” in *2018 IEEE International Conference on Electro/Information Technology (EIT)*. IEEE, 2018, pp. 0537–0542.
- [79] Z. Chen, M. I. AlHajri, M. Wu, N. T. Ali, and R. M. Shubair, “A novel real-time deep learning approach for indoor localization based on RF environment identification,” *IEEE Sensors Letters*, vol. 4, no. 6, pp. 1–4, 2020.
- [80] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, “A survey on behavior recognition using WiFi channel state information,” *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.
- [81] N. Marques, F. Meneses, and A. Moreira, “Combining similarity functions and majority rules for multi-building, multi-floor, WiFi positioning,” in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2012, pp. 1–9.
- [82] U. M. Khan, Z. Kabir, S. A. Hassan, and S. H. Ahmed, “A deep learning framework using passive WiFi sensing for respiration monitoring,” in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.
- [83] J. Yang, X. Chen, H. Zou, C. X. Lu, D. Wang, S. Sun, and L. Xie, “Sensefi: A library and benchmark on deep-learning-empowered WiFi human sensing,” *Patterns*, 2023.
- [84] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, “WiFi CSI based passive human activity recognition using attention based BLSTM,” *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, 2018.
- [85] L. Guo, L. Wang, J. Liu, W. Zhou, and B. Lu, “HuAc: Human activity recognition using crowdsourced WiFi signals and skeleton data,” *Wireless Communications and Mobile Computing*, vol. 2018, 2018.
- [86] R. Alazrai, A. Awad, A. BahaA, M. Hababeh, and M. I. Daoud, “A dataset for WiFi-based human-to-human interaction recognition,” *Data in Brief*, vol. 31, p. 105668, 2020.
- [87] R. Alazrai, M. Hababeh, A. BahaA, M. Z. Ali, and M. I. Daoud, “An end-to-end deep learning framework for recognizing human-to-human interactions using WiFi signals,” *IEEE Access*, vol. 8, pp. 197 695–197 710, 2020.
- [88] J. Yang, X. Chen, H. Zou, D. Wang, Q. Xu, and L. Xie, “Efficientfi: Toward large-scale lightweight WiFi sensing via CSI compression,” *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 13 086–13 095, 2022.
- [89] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, “Multimodal CSI-based human activity recognition using GANs,” *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17 345–17 355, 2021.
- [90] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo, “Contactless respiration monitoring via off-the-shelf WiFi devices,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 10, pp. 2466–2479, 2015.
- [91] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, “Widar3.0: Zero-effort cross-domain gesture recognition with WiFi,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [92] J. Yang, X. Chen, H. Zou, D. Wang, and L. Xie, “AutoFi: Towards Automatic WiFi Human Sensing via Geometric Self-Supervised Learning,” *ArXiv preprint ArXiv:2205.01629*, 2022.
- [93] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, “Joint activity recognition and indoor localization with WiFi fingerprints,” *IEEE Access*, vol. 7, pp. 80 058–80 068, 2019.
- [94] H. Abdelnasser, M. Youssef, and K. A. Harras, “Wigest: A ubiquitous WiFi-based gesture recognition system,” in *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2015, pp. 1472–1480.
- [95] J. Maitre, K. Bouchard, C. Bertuglia, and S. Gaboury, “Recognizing activities of daily living from UWB radars and deep learning,” *Expert Systems with Applications*, vol. 164, p. 113994, 2021.
- [96] J. Tian, S. Yongkun, D. Yongpeng, H. Xikun, S. Yongping, Z. Xiaolong, and Q. Zhifeng, “UWB-HA4D-1.0: An ultra-wideband radar human activity 4D imaging dataset,” *Journal of Radars*, vol. 11, no. 1, pp. 27–39, 2022.
- [97] Z. Zhengliang, Y. Degui, Z. Junchao, and T. Feng, “Dataset of human motion status using IR-UWB through-wall radar,” *Journal of Systems Engineering and Electronics*, vol. 32, no. 5, pp. 1083–1096, 2021.
- [98] S. Ahmed, D. Wang, J. Park, and S. H. Cho, “UWB-gestures, a public dataset of dynamic hand gestures acquired using impulse radar sensors,” *Scientific Data*, vol. 8, no. 1, p. 102, 2021.
- [99] H. Liu, A. Zhou, Z. Dong, Y. Sun, J. Zhang, L. Liu, H. Ma, J. Liu, and N. Yang, “M-gesture: Person-independent real-time in-air gesture recognition using commodity millimeter wave radar,” *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 3397–3415, 2021.

- [100] E. Gambi, G. Ciattaglia, A. De Santis, and L. Senigagliaesi, "Millimeter wave radar data of people walking," *Data in brief*, vol. 31, p. 105996, 2020.
- [101] Z. Wang, C. Yao, J. Ren, M. Feng, and X. Jiang, "Human activity recognition using 3d orthogonally-projected efficientnet on radar time-range-doppler signature," in *2022 IEEE 2nd International Conference on Software Engineering and Artificial Intelligence (SEAI)*. IEEE, 2022, pp. 26–30.
- [102] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2011.
- [103] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 4264–4271.
- [104] G. Benitez-Garcia, J. Olivares-Mercado, G. Sanchez-Perez, and K. Yanai, "Ipn hand: A video dataset and benchmark for real-time continuous hand gesture recognition," in *2020 25th international conference on pattern recognition (ICPR)*. IEEE, 2021, pp. 4340–4347.
- [105] A. Kapitanov, A. Makhlyarchuk, and K. Kvanchiani, "Hagrid-hand gesture recognition image dataset," *arXiv preprint arXiv:2206.08219*, 2022.
- [106] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "Hmdb: a large video database for human motion recognition," in *2011 International conference on computer vision*. IEEE, 2011, pp. 2556–2563.
- [107] C. Gao, Y. Du, J. Liu, J. Lv, L. Yang, D. Meng, and A. G. Hauptmann, "Infar dataset: Infrared action recognition at different times," *Neurocomputing*, vol. 212, pp. 36–47, 2016.
- [108] J. Carreira and A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6299–6308.
- [109] K. Soomro, A. R. Zamir, and M. Shah, "Ucf101: A dataset of 101 human actions classes from videos in the wild," *arXiv preprint arXiv:1212.0402*, 2012.
- [110] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev *et al.*, "The kinetics human action video dataset," *arXiv preprint arXiv:1705.06950*, 2017.
- [111] P. Pareek and A. Thakkar, "A survey on video-based human action recognition: recent updates, datasets, challenges, and applications," *Artificial Intelligence Review*, vol. 54, pp. 2259–2322, 2021.
- [112] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based methods for action representation, segmentation and recognition," *Computer Vision and Image Understanding*, vol. 115, no. 2, pp. 224–241, 2011.
- [113] O. Martin, I. Kotsia, B. Macq, and I. Pitas, "The eNTERFACE'05 audio-visual emotion database," in *22nd International Conference on Data Engineering Workshops (ICDEW'06)*. IEEE, 2006, pp. 8–8.
- [114] G. Muhammad and M. F. Alhamid, "User emotion recognition from a larger pool of social network data using active learning," *Multimedia Tools and Applications*, vol. 76, no. 8, pp. 10 881–10 892, 2017.
- [115] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Acted facial expressions in the wild database," *Australian National University, Canberra, Australia, Technical Report TR-CS-11*, vol. 2, p. 1, 2011.
- [116] E. Avots, T. Sapiński, M. Bachmann, and D. Kamińska, "Audiovisual emotion recognition in wild," *Machine Vision and Applications*, vol. 30, no. 5, pp. 975–985, 2019.
- [117] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2017.
- [118] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2852–2861.
- [119] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439–2450, 2018.
- [120] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *International Conference on Neural Information Processing*. Springer, 2013, pp. 117–124.
- [121] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expressions in the wild: data and experiment protocol," *CVHCI Google Scholar*.
- [122] F. Ringeval, B. Schuller, M. Valstar, J. Gratch, R. Cowie, S. Scherer, S. Mozgai, N. Cummins, M. Schmitt, and M. Pantic, "Avec 2017: Real-life depression, and affect recognition workshop and challenge," in *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, 2017, pp. 3–9.
- [123] L. Yang, D. Jiang, X. Xia, E. Pei, M. C. Oveneke, and H. Sahli, "Multimodal measurement of depression using deep learning models," in *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, 2017, pp. 53–59.
- [124] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic, "Avec 2013: the continuous audio/visual emotion and depression recognition challenge," in *Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, 2013, pp. 3–10.
- [125] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic, "Avec 2014: 3D dimensional affect and depression recognition challenge," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, 2014, pp. 3–10.
- [126] Y. Fu, B. Li, X. Zhu, and C. Zhang, "Active learning without knowing individual instance labels: a pairwise label homogeneity query approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 4, pp. 808–822, 2013.
- [127] R. A. Borgalli and S. Surve, "Deep learning framework for facial emotion recognition using cnn architectures," in *2022 International Conference on Electronics and Renewable Systems (ICEARS)*. IEEE, 2022, pp. 1777–1784.
- [128] J. She, Y. Hu, H. Shi, J. Wang, Q. Shen, and T. Mei, "Dive into ambiguity: Latent distribution mining and pairwise uncertainty estimation for facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6248–6257.
- [129] J. R. Williamson, T. F. Quatieri, B. S. Helfer, R. Horwitz, B. Yu, and D. D. Mehta, "Vocal biomarkers of depression based on motor incoordination," in *Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, 2013, pp. 41–48.
- [130] Y. Gong and C. Poellabauer, "Topic modeling based multi-modal depression detection," in *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, 2017, pp. 69–76.
- [131] N. Lannan, L. Zhou, and G. Fan, "A multiview depth-based motion capture benchmark dataset for human motion denoising and enhancement research," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 427–436.
- [132] T. Mantecón, C. R. Del-Blanco, F. Jaureguizar, and N. García, "A real-time gesture recognition system using near-infrared imagery," *PloS one*, vol. 14, no. 10, p. e0223320, 2019.
- [133] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, "Facial expression recognition from near-infrared videos," *Image and Vision Computing*, vol. 29, no. 9, pp. 607–619, 2011.
- [134] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 2010, pp. 94–101.
- [135] S. Shaees, H. Naeem, M. Arslan, M. R. Naeem, S. H. Ali, and H. Aldabbas, "Facial emotion recognition using transfer learning," in *2020 International Conference on Computing and Information Technology (ICCIT-1441)*. IEEE, 2020, pp. 1–5.
- [136] S. Z. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 627–639, 2007.
- [137] S. Li, D. Yi, Z. Lei, and S. Liao, "The casia nir-vis 2.0 face database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 348–353.
- [138] R. He, X. Wu, Z. Sun, and T. Tan, "Wasserstein cnn: Learning invariant features for NIR-VIS face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1761–1773, 2018.
- [139] N. K. Benamara, E. Zigh, T. B. Stambouli, and M. Keche, "Towards a robust thermal-visible heterogeneous face recognition approach based on a cycle generative adversarial network," *Int. J. Interact. Multimed. Artif. Intell.*, 2022.
- [140] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, B. Weiss *et al.*, "A database of German emotional speech," in *Interspeech*, vol. 5, 2005, pp. 1517–1520.
- [141] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1D & 2D CNN LSTM networks," *Biomedical Signal Processing and Control*, vol. 47, pp. 312–323, 2019.
- [142] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: an asr corpus based on public domain audio books," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 5206–5210.

- [143] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12 449–12 460, 2020.
- [144] R. Ardila, M. Branson, K. Davis, M. Henretty, M. Kohler, J. Meyer, R. Morais, L. Saunders, F. M. Tyers, and G. Weber, "Common voice: A massively-multilingual speech corpus," *ArXiv preprint ArXiv:1912.06670*, 2019.
- [145] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Al-hussain, "Speech emotion recognition using deep learning techniques: A review," *IEEE Access*, vol. 7, pp. 117 327–117 345, 2019.
- [146] S. K. Gaikwad, B. W. Gawali, and P. Yannawar, "A review on speech recognition technique," *International Journal of Computer Applications*, vol. 10, no. 3, pp. 16–24, 2010.
- [147] F. Q. AL-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, "Respiration rate monitoring methods: A review," *Pediatric pulmonology*, vol. 46, no. 6, pp. 523–529, 2011.
- [148] E. Vanegas, R. Igual, and I. Plaza, "Sensing systems for respiration monitoring: A technical systematic review," *Sensors*, vol. 20, no. 18, p. 5446, 2020.
- [149] X. Liu, J. Yin, Y. Liu, S. Zhang, S. Guo, and K. Wang, "Vital signs monitoring with RFID: Opportunities and challenges," *IEEE Network*, vol. 33, no. 4, pp. 126–132, 2019.
- [150] A. Prati, C. Shan, and K. I.-K. Wang, "Sensors, vision and networks: From video surveillance to activity recognition and health monitoring," *Journal of Ambient Intelligence and Smart Environments*, vol. 11, no. 1, pp. 5–22, 2019.
- [151] C. Will, K. Shi, S. Schellenberger, T. Steigleder, F. Michler, J. Fuchs, R. Weigel, C. Ostgathe, and A. Koelpin, "Radar-based heart sound detection," *Scientific Reports*, vol. 8, no. 1, pp. 1–14, 2018.
- [152] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy in blockchain technology: A futuristic approach," *Journal of Parallel and Distributed Computing*, vol. 145, pp. 50–74, 2020.
- [153] R. Steele, A. Lo, C. Secombe, and Y. K. Wong, "Elderly persons perception and acceptance of using wireless sensor networks to assist healthcare," *International Journal of Medical Informatics*, vol. 78, no. 12, pp. 788–801, 2009.



Junxin Chen (Senior Member, IEEE) is currently a Full Professor at the School of Software, Dalian University of Technology, Dalian 116621, China. He received the B. Sc. Degree in Communications Engineering and M.Sc. and Ph.D degrees in Communications and Information System, all from Northeastern University in 2007, 2009, and 2016 respectively. From 2019 to 2020, he was a Postdoc Research Fellow in Department of Computer and Information Science, under the UM Macau Talent Programme (Class A), University of Macau. He worked as an Assistant Professor and Associate Professor in the College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China. Dr. Chen is currently working as a Full Professor at the School of Software, Dalian University of Technology. His research interests include Internet of Medical Things, smart healthcare, artificial intelligence, and information security. He has authored/co-authored over 100+ scientific papers in international peer-reviewed journals and conferences, such as, IEEE T-MM, T-CSVT, T-II, IoT, and IEEE Network, IEEE Photonics Journal, etc. He has an H-index of 33 and a total of 3200 citations. He is an associate editor for Expert Systems (Wiley), the Leading Guest Editor of IEEE Journal of Biomedical and Health Informatics, etc., a Guest Editor of Signal Processing: Image Communication, etc. He is a regular reviewer of extensive top-tier field journals. He has received more than 10 awards from Mainland China and Macau, and has given more than 30 invited talks for universities and companies in China and Australia, such as University of Macau, Deakin University, etc. He is a Senior Member of IEEE.



Tongyue He was born in Shandong, China, in Jan. 1998. He received his B.A. degree in automation, in 2020. He was a master's student in control engineering at Northeastern University, China, from 2020 to 2022. He is currently pursuing the Ph.D. degree in biomedical engineering at Northeastern University, Shenyang, China. His research interests include optical fiber sensor and deep learning based on artificial intelligence.



Kangyue Liang was born in Shijiazhuang, China. He received the bachelors degree in biomedical engineering from Northeastern University. He is currently pursuing the master's degree in electronic information with Northeastern University, Shenyang, China. His research interest includes contactless sensing technologies.



Wei Wang is currently a Professor with the Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, China; and also with the School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China. He had been the UM Macao Research Fellow at University of Macau, Macau SAR. He received PhD degree in software engineering from Dalian University of Technology in 2018. His research interests include computational social science, data mining, internet of things, and artificial intelligence.



Amit Kumar Singh is currently an Associate Professor with the Computer Science and Engineering Department, National Institute of Technology Patna, Bihar, India. He has authored over 250 peer-reviewed journals, conference publications, and book chapters. Dr. Singh have been recognized as "WORLD RANKING OF TOP 2% SCIENTISTS" in the area of Biomedical Research" (for Year 2020) and "Artificial Intelligence & Image Processing" (for the Year 2021, 2022 and 2023) according to the survey given by Stanford University, USA. According to AD Scientific Index 2024, he is listed at 54th position Asia Top 100 Engineering & Technology / Computer Science Scientists in India 2024. Dr. Singh is the Associate Editor of IEEE Trans. On Multimedia, ACM Trans. Multimedia Comput. Commun. Appl., IEEE Trans. Computat. Social Syst., IEEE Trans. Ind. Informat., IEEE J. Biomed. Heal. Informatics, Eng. Appl. Artif. Intell., Elsevier. His research interests include multimedia data hiding, image processing, biometrics, & Cryptography.



Danda B. Rawat (Senior Member, IEEE) is the Executive Director, Research Institute for Tactical Autonomy (RITA) - a University Affiliated Research Center (UARC) of the US Department of Defense, Associate Dean for Research & Graduate Studies, a Full Professor in the Department of Electrical Engineering & Computer Science (EECS), Founding Director of the Howard University Data Science & Cybersecurity Center, and Founding Director of DoD Center of Excellence in Artificial Intelligence & Machine Learning (CoE-AIML). He has secured over \$110 million as a PI and over \$18 million as a Co-PI in research funding from the US National Science Foundation (NSF), US Department of Homeland Security (DHS), US National Security Agency (NSA), US Department of Energy, National Nuclear Security Administration (NNSA), National Institute of Health (NIH), DoD and DoD Research Labs, Industry (Microsoft, Intel, VMware, PayPal, etc.) and private Foundations. He is the recipient of NSF CAREER Award in 2016, Department of Homeland Security (DHS) Scientific Leadership Award in 2017, Provost's Distinguished Service Award 2021, and the Best Paper Awards (IEEE CCNC, IEEE ICII, IEEE DroneCom and BWCA). He has been serving as an Editor/Guest Editor for over 100 international journals including the Associate Editor of IEEE Transactions on Cognitive Communications and Networking, Associate Editor of IEEE Transactions of Service Computing, Editor of IEEE Internet of Things Journal, Associate Editor of IEEE Transactions of Network Science and Engineering and Technical Editors of IEEE Network. He is a Senior Member of IEEE and ACM, a member of ASEE and AAAS, and a Fellow of the Institution of Engineering and Technology (IET). He is an ACM Distinguished Speaker (2021- 2023).



Houbing Song received the Ph.D. degree in electrical engineering from University of Virginia, Charlottesville, VA, in August 2012. He is currently a Tenured Associate Professor, the Director of the NSF Center for Aviation Big Data Analytics (Planning), the Associate Director for Leadership of the DOT Transportation Cybersecurity Center for Advanced Research and Education (Tier 1 Center), and the Director of the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us), University of Maryland, Baltimore County (UMBC), Baltimore, MD. Prior to joining UMBC, he was a Tenured Associate Professor of Electrical Engineering and Computer Science at Embry-Riddle Aeronautical University, Daytona Beach, FL. He serves as an Associate Editor for IEEE Transactions on Artificial Intelligence (TAI) (2023-present), IEEE Internet of Things Journal (2020-present), IEEE Transactions on Intelligent Transportation Systems (2021-present), and IEEE Journal on Miniaturization for Air and Space Systems (J-MASS) (2020-present). He was an Associate Technical Editor for IEEE Communications Magazine (2017-2020). His research interests include cyber-physical systems/Internet of Things, cybersecurity and privacy, and AI/machine learning/big data analytics. Dr. Song is an IEEE Fellow, an ACM Distinguished Member, and an ACM Distinguished Speaker. Dr. Song has been a Highly Cited Researcher identified by Clarivate (2021, 2022) and a Top 1000 Computer Scientist identified by Research.com.



Zhihan Lyu (Senior Member, IEEE) is currently an Associate Professor at Uppsala University in Sweden. He has contributed 300 articles, including more than 50 articles on IEEE/ACM TRANSACTIONS. He has received U.S. \$1.5 million funding as a PI. He has been involved in many European and Chinese projects supported with funding of U.S. \$25 million. He received more than 20 awards from China, Europe, and IEEE. He has given more than 80 invited talks for universities and companies in Europe and China. He has given 15 keynote speeches for international conferences. He is currently the Co-Chair or a TPC of 50 conferences, including ACM MM 2021 and ACM IUI 2015-2022. He has reviewed 400 articles. He is also the Editor-in-Chief of Internet of Things and Cyber-Physical Systems (KeAi); an Associate Editor of 18 journals, including IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, and IEEE TECHNOLOGY POLICY AND ETHICS NEWSLETTER; and a Leading Guest Editor for 40 special issues, including nine IEEE. He is an IEEE Senior Member, BCS Fellow, and ACM Distinguished Speaker.