Identifying Child Users via Touchscreen Interactions

YUSHI CHENG and XIAOYU JI, Zhejiang University
XIAOPENG LI, University of South Carolina
TIANCHEN ZHANG, Zhejiang University
SHARAF MALEBARY, King Abdulaziz University
XIANSHAN QU, University of South Carolina
WENYUAN XU, Zhejiang University

With the proliferation of smart devices, children can be easily exposed to violent or adult-only content on the Internet. Without any precaution, the premature and unsupervised use of smart devices can be harmful to both children and their parents. Thus, it is critical to employ parent patrol mechanisms such that children are restricted to child-friendly content only. A successful parent patrol strategy has to be user friendly and privacy aware. The apps that require explicit actions from parents are not effective because a parent may forget to enable them, and the ones that use built-in cameras or microphones to detect child users may impose privacy violations. In this article, we propose `iCare`, a system that can identify child users automatically and seamlessly when users operate smartphones. In particular, `iCare` investigates the intrinsic differences of screen-touch patterns between child and adult users from the aspect of physiological maturity. We discover that one's touch behaviors are related to his or her age. Thus, `iCare` records the touch behaviors and extracts hand geometry, finger dexterity, and hand stability features that capture the age information. We conduct experiments on 100 people including 62 children (3 to 17 years old) and 38 adults (18 to 59 years old). Results show that `iCare` can achieve 96.6% accuracy for child identification using only a single swipe on the screen, and the accuracy becomes 98.3% with three consecutive swipes.

CCS Concepts: • **Security and privacy** → **Authentication**; • **Human-centered computing** → *Ubiquitous and mobile computing*;

Additional Key Words and Phrases: Child identification, touch behavior, automatic, implicit

35

ACM Reference format:
Yushi Cheng, Xiaoyu Ji, Xiaopeng Li, Tianchen Zhang, Sharaf Malebary, Xianshan Qu, and Wenyuan Xu. 2020. Identifying Child Users via Touchscreen Interactions. *ACM Trans. Sen. Netw.* 16, 4, Article 35 (July 2020), 25 pages.
https://doi.org/10.1145/3403574

This study was supported by the National Natural Science Foundation of China (61925109, 61702451, and 61941120) and the Natural Science Foundation of Zhejiang Province (LGG19F020020). This article is an extended version of the work published at ACM HotMobile in Tempe, Arizona, in February 2018 [20].
Authors' addresses: Y. Cheng, X. Ji (corresponding author), T. Zhang, and W. Xu (corresponding author), Zhejiang University, Hangzhou, 310027, CN; emails: {yushicheng, xji, tianchen-zhang, xuwenyuan}@zju.edu.cn; X. Li and X. Qu, University of South Carolina, Columbia, SC 29208; emails: {xl4, xqu}@email.sc.edu; S. Malebary, King Abdulaziz University, jeddah, Saudi Arabia; email: smalebary@kau.edu.sa.

1550-4859/2020/07-ART35 $15.00
https://doi.org/10.1145/3403574

1 INTRODUCTION

Mobile devices such as smartphones have woven into the fabric of daily life by providing always-on services ranging from communication to working to entertainment. For convenience, parents tend to share their own devices with their kids at an early age. Recent studies [19, 39] show that almost 90% of modern children have a moderate ability to use a tablet, and they can be even as young as 2 years old. Without any precaution on sharing, the premature and unsupervised use of smart devices has been shown to be harmful to both children and their parents. First, the Internet hosts a variety of content and children can be easily exposed to violent or adult-only images and videos [31, 38], if not guided properly. Actually, children are more likely to be attracted by games and become addicted due to their limited self-control ability, resulting in both physical and mental illness. It is reported that today, children with anxiety are more common than before [35] due to the influence of online content. Second, when children use their parents' smart devices, they may leak private information such as photos unintentionally, or purchase goods without parents' consent. According to a report from the Federal Trade Commission (FTC) [8], in-app purchases from children during the past few years are valued at more than $70 million in Amazon, $32.5 million in Apple, and $19 million in Google. Thus, parent patrol of smart devices for child users is urgently needed.

Various regulations from governments and organizations are established to provide guidance on content access, such as the healthy media use guidelines from the American Academy of Pediatrics [30] and the Child Online Protection Initiative from the International Telecommunication Union (ITU) [18]. Based on those regulations, many parent patrol apps are developed [10, 21]. Although such apps are effective in content access control, they require to know in advance that the current users are children. Typically, this is achieved by manual activation from parents. Alternatively, images [4, 26] or voice data [25, 27] can be utilized to infer the age of users, which, however, might be privacy violating.

In this article, we ask the following question: can we identify child users in an automatic and friendly way?—meaning that the identification requires no explicit user intervention and can automatically differentiate a child from an adult transparently while preserving privacy. To this end, we try to find age-related differences between adults and children based on their touchscreen interactions. Our idea is inspired by touch-based user authentication approaches [11, 22, 34, 40], but these approaches cannot be applied to child identification trivially. This is because authentication focuses on finding distinctive characteristics of a *specific* user, yet child identification aims at finding common distinctive characteristics among a *group* of users at similar ages. Identifying group characteristics is challenging for the following reasons. First, the group could contain a large number of users, and the characteristics of users within the same group may vary greatly. Second, the characteristics of one group may overlap with another group. For instance, the film rating system from the Classification and Rating Administration (CARA) [6] requires to differentiate children that are younger than 13 from ones between 13 and 17, which is challenging, and to the best of our knowledge cannot be accomplished using existing touch-based child identification methods [14, 23, 37].

To overcome the aforementioned challenges, we examine the underlying difference between child and adult users from the aspect of physiological maturity, and propose `iCare` to identify child users in an *automatic* and *implicit* manner. We investigate and extract three key age-related features—hand *geometry*, finger *dexterity*, and hand *stability*—which are proven to be age representative. The three macro features are further extended to 53 sub-features to capture the distinct characteristics of children. To obtain them, we carefully select two natural and commonly used interaction gestures—*tap* and *swipe*—as our target touch behaviors. To make the experiment

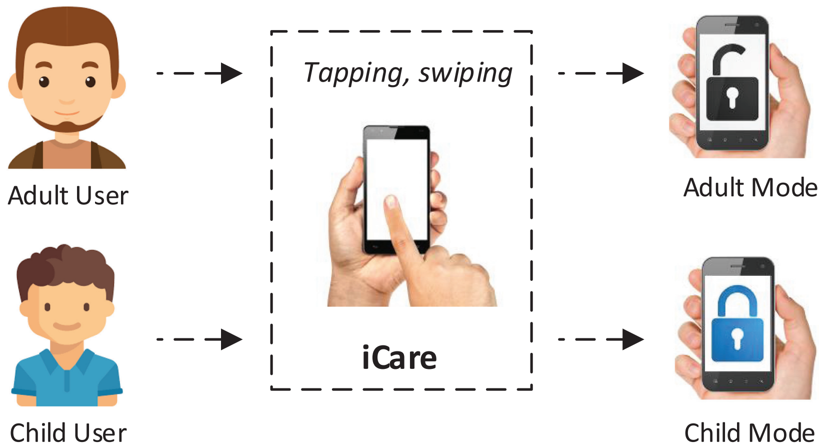


Fig. 1. iCare can identify child users automatically and seamlessly based on their touch behaviors and enable a proper mode correspondingly.

data comprehensive and original, we manually design four tasks for data collection, including two screen-unlocking methods (both password and pattern based) and modifications of two popular applications.

We envision that iCare can support both device- and application-based patrol control. For the former, iCare monitors screen-unlocking operations upon a user's attempt to unlock a device, and the device can then enter the proper mode correspondingly, as shown in Figure 1. For the latter, after the user unlocks a device (or the screen locking is disabled), iCare collects in-app touch behaviors to facilitate a specific application to decide the age of the current user. In the whole process, the user is unaware of the existence of iCare, as no explicit involvement is required from users. A dataset on 100 people including 62 children (3 to 17 years old) and 38 adults (18 to 59 years old) verifies the effectiveness of iCare, and the evaluation results indicate that iCare can achieve an accuracy of 96.6% with a single swipe and 98.3% with multiple consecutive swipes.

To the best of our knowledge, this is the first work on detecting child users older than 11 based on touch patterns. Our main contributions include the following:

- We propose iCare, an effective approach to automatically detect child users on smartphones without compromising user experience and violating privacy.
- We examine the underlying difference between child and adult users in terms of their touch behaviors when operating smart devices, and explore three key age-related features (hand geometry, finger dexterity, and hand stability) that are represented in 53 sub-features to capture the distinct characteristics of children.
- We evaluate iCare using the dataset collected from 62 child users (3 to 17 years old) and 38 adults (18 to 59 years old). Our dataset is collected *in the wild* during users' normal operations without any constraints. Evaluation results demonstrate that iCare can achieve an accuracy of 96.6% using only a single swipe, and the performance can be further improved to 98.3% with three consecutive swipes.

The rest of the article is organized as follows. In Section 2, we first review the work related to our problem. Then, we explain the motivation of this article and introduce the basic idea of iCare in Section 3, then present the system overview and design details of iCare in Section 4. Section 5 evaluates the performance of iCare with extensive experiments, and Section 6 measures its usability and user acceptance rate. Afterward, we discuss the limitations and future work in Section 7 and conclude our article in Section 8.

2 RELATED WORK

In this section, we present studies relevant to ours. Specifically, we discuss the aspects related to child identification and user authentication.

2.1 Child Identification

Much attention has been devoted to recognizing child users from adults. Existing approaches can be divided into two categories: the explicit classification based on extra input information (images, voice, etc.) and the implicit classification based on data captured opportunistically, such as gesture measurements from user behaviors.

Explicit classification. Basaran et al. [4] introduce a classification system (ChildSafe) that exploits human skeletal features collected by 3D depth cameras to identify children. Meinedo and Trancoso [25] propose a paradigm to use acoustic and prosodic features to differentiate age and gender. Although those approaches show proper detection and classification of children (e.g., greater than 97%), utilizing sensors such as built-in cameras and microphones may receive strong resistance from privacy-savvy users because the private information may need to be uploaded to a cloud server.

Implicit classification. Pental [31] infers computer users' age and gender based on keystrokes from a keyboard and mouse. In the areas of touch-based devices, Vatavu et al. [37] utilize touch coordinates to divide users into age groups. Hernandez-Ortega et al. [14] utilize the Sigma-Lognormal theory to capture users' neuromotor skills, which can distinguish children from adults. These two methods [14, 37] only identify children younger than 6 years. Our previous work utilizes the touch behaviors and extracts related features to capture the age information and pushes the boundary to 11 years [23]. However, it is still insufficient to satisfy the requirement of some regulations, such as the CARA film rating system's distinct age of 17. In this article, we improve our previous work by investigating the theoretical foundation of `iCare` from the aspect of physiological maturity, exploring new features that capture the hand stability information of user interactions to achieve higher identification accuracy, and validating the usability and efficiency of `iCare` with more users and extensive experiments. In this way, `iCare` achieves identifying children of various ages who may behave similarly to adults. In addition, existing work [14, 37] uses particular gestures (e.g., dragging two targets simultaneously with two fingers). `iCare`, however, utilizes a wider variety of simple operations, such as swiping and tapping.

2.2 User Authentication

In addition to child identification, our work is inspired by the literature of touch-based user authentication. Frank et al. [11] investigate the applicability of using touchscreen inputs to continuously authenticate users. They extract 30 behavioral features from raw touchscreen logs and justify that simple touch movements are sufficient to authenticate a user. Li et al. [22] describe a continuous user authentication mechanism for smartphones by checking the user's finger movement patterns. Zheng et al. [40] extract four features (acceleration, pressure, size, and time) from smartphone sensors and implement a verification mechanism to validate whether an authenticating user has the true ownership of a smartphone. In HMOG, Sitová et al. [34] introduce a new set of biometric behavioral features (hand movement, orientation, and grasp) for continuous authentication on smartphones. These features can unobtrusively capture subtle micro-movements generated when users tap a screen. These studies mainly focus on finding distinctive characteristics of a specific user for identification, whereas our work aims at exploring the common distinctive

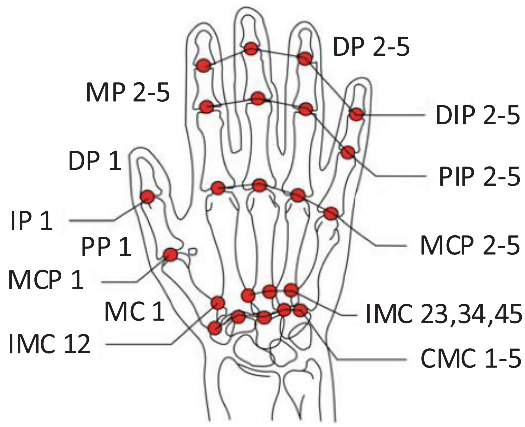


Fig. 2. Skeleton of a human hand. Fingers are numbered as: (1) thumb, (2) index, (3) middle, (4) ring, and (5) little.

characteristics among a group of users at similar ages. Identifying group characteristics can be challenging because the characteristics of users in a group may have a large variance and the ones from two groups may overlap. Our work overcomes such challenges and manages to identify child users.

3 MOTIVATION AND BASIC IDEA

`iCare` targets to identify child users on smartphones in a continuous and natural way by utilizing the features extracted from user operations. Specifically, the identification should be continuous, user friendly, and privacy preserving (i.e., without requiring extra operations from users or violating privacy). To this end, we investigate the following two questions. First, what kind of features should be squeezed to represent the age-relevant characteristics of children? Second, what actions or operations should be utilized to obtain the features?

To investigate, we resort to the essential difference of maturity between children and adults. According to human engineering and kinesiology, common characteristics can be found among people of similar age groups, such as children and adults [28, 36]. The underlying principle is that physiological maturity is different for people of various age groups and their interactions with smartphones, therefore, show discrepancies across age groups. Specifically, human hands develop with age, and the hand development then affects the user's interaction with smartphones.

Hand kinematics of humans. Research in anatomy and biomechanics has shown that the human hand is a highly complex and elegant biomechanical device capable of both gross and fine motor skills [1]. This dexterity emerges from the unique configuration of bones, joints, and muscles. A human hand includes 27 connected bones divided into five groups: 8 carpal bones, 5 metacarpal (MC) bones, 5 proximal phalanges (PP), 4 medial phalanges (MP), and 5 distal phalanges (DP), as shown in Figure 2. They are further connected by six types of joints: the carpometacarpal (CMC) joints, the intermetacarpal (IMC) joints, the metacarpophalangeal (MCP) joints, the proximal interphalangeal (PIP) joints, the distal interphalangeal (DIP) joints, and the interphalangeal (IP) joint of the thumb [13]. Muscles, such as the extensor digitorum communis (EDC) in the forearm, have insertions in multiple joints. The activation of such muscles results in the excursion of multiple tendons, which enables the abundant functions of human hands. As a result, the developmental levels of bones and muscles impact the kinematics of the human hand.

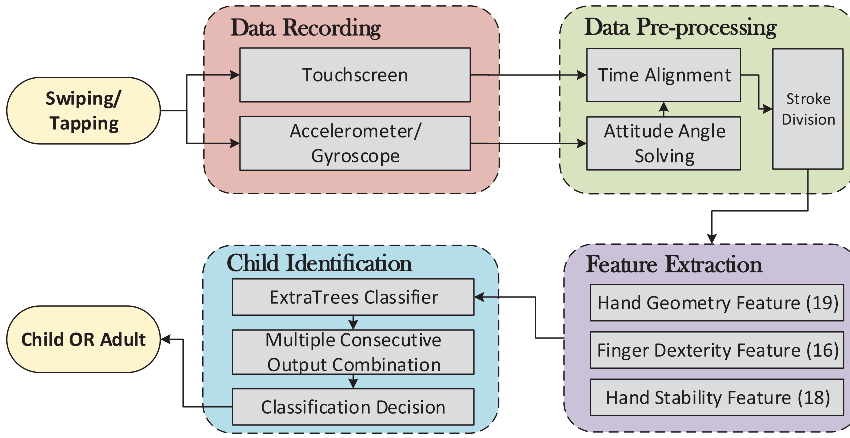


Fig. 3. System overview of iCare.

Hand development of children. The immature child body develops with age. In forensic science, child age can be estimated by measurements of carpals and epiphyses [5]:

$$Age = c_0 + c_1 g + c_2 Bo/Ca + c_3 N_0 - c_4 s - c_5 N_0 s, \tag{1}$$

where g is a variable equal to 1 for boys and 0 for girls, Bo/Ca is the ratio between carpal bones area and carpal area, N_0 is the number of teeth with complete root development, s is the sum of the normalized open apices, and $c_0 - c_5$ are non-negative constants. Therefore, the length and breadth of children's hand bones increase with age, and adults have relatively larger hands and their sizes are stable [16]. For instance, adults (18 to 40 years) have hands 1.4 times longer and 1.3 times wider than those of children (5 to 12 years) [17]. Similarly, the speed of hand dexterity improves with age among children [20, 32]. A test in a group of 237 normal, right-handed children between 5 and 8 years also shows that the rapidity of finger movements increases with age [9]. In the meanwhile, the children's grasp strength becomes steady and grows stronger with age due to the development of muscles. As a result, older children can generally hold things more stably than younger ones [17, 24].

How does hand development affect a user's interactions with mobile devices? As children grow, they improve their abilities on touchscreens in terms of both task completion time and accuracy [3, 38]. For instance, they can tap and drag faster on touchscreens. In particular, children older than 5 years perform taps 4 times faster and drag items 1.3 times faster than those younger than 4 years. In addition, their operations become more accurate with the increase in age. For example, children older than 5 years can press approximately 25% closer to the centers of a target area than those younger ones. It is foreseeable that the discrepancy between adults and children can be larger, especially for young children.

We envision that the aforementioned physiological differences are intrinsic and can be useful in identifying child users, and we extract these age-related characteristics of child users from their normal touch operations on smartphones. Specifically, we elaborate on child-relevant characteristics from three aspects: hand geometry, finger dexterity, and hand stability, as we will show in detail in Section 4.3.

4 DESIGN

The system overview of `iCare` is shown in Figure 3. In general, `iCare` takes swiping and tapping as input and records the corresponding data from the touchscreen, accelerometer, and gyroscope.

Data pre-processing is then performed on the accelerometer and gyroscope data to solve the attitude angles, which are merged with the touchscreen data based on timestamps before being divided into strokes. Three key features (hand geometry, finger dexterity, and hand stability), which are represented in 53 sub-features, are extracted to capture the age-related characteristics embedded in each stroke. Those features are then fed into an ExtraTrees (ET) classifier, from which multiple consecutive outputs are combined to reach a final decision (i.e., whether the current user is a child or not).

4.1 Gesture Choices

To implicitly capture the behavior differences between child and adult users on mobile devices, we have considered the following three criteria when selecting interaction gestures.

Generality and easy-to-use. Generality ensures that an interaction gesture can be widely adopted by both child and adult users. The gesture should also be supported by most mobile applications for implementing abundant functionality so that the data can be easily acquired for classification. Additionally, the gesture should be easy to use for both children and adults. A complicated gesture may cause child users to be frustrated and lead to unsuccessful gestures.

Unobtrusive. To collect data in an unobtrusive way, the gesture should be performed naturally without requiring users to perform extra and abnormal operations or interrupting their existing interactions with smartphones. In other words, the gesture should be a part of users' daily operations on smartphones.

Distinguishable. Furthermore, the gesture should be able to contain sufficient behavior information such that it can distinguish children from adults. Note that representativeness is more important than richness. For example, some gestures may contain rich information and behave well for user authentication, but they may fail to be qualified for child user identification due to the poor generalization capability of extracting group characteristics.

4.1.1 Gestures on Smartphones. Figure 4 shows eight commonly used gestures on smartphones. The first three are single-touch gestures, and the others are multi-touch gestures that can be used to implement augmented functionality (e.g., zoom in/out in a map application to view nearby places of a location). According to our observations, the last two gestures—*press and swipe* and *rotate*—are difficult for children younger than 9 years to perform independently, and thus they cannot satisfy the criterion of easy-to-use. Moreover, these two gestures are not well supported by most applications (i.e., not general). The gesture of *press and tap* is not frequently used by children due to their poor finger dexterity, although it is adopted by most adult users. The gestures of *pinch* and *spread* are commonly used when viewing maps or images, but they are rarely performed in other applications.

Using tap and swipe. Considering all of the criteria, we finally select *tap* and *swipe* as our gestures. *Tap* is one of the most essential interactions with mobile devices. A single tap is usually performed with one of the fingertips for pressing a button, selecting an image, or typing on an on-screen keyboard. As it is simple and well practiced, tapping has been used in almost all mobile applications and performed by users across all age groups. In particular, tap gestures are necessary to access applications, locally store information on smartphones, or place orders online.

Swipe is another fundamental gesture to interact with smartphones. Users perform this gesture by moving one of their fingertips (usually a thumb or index finger) across the screen without losing contact. Swipe gestures can enable many basic functions, such as sliding the current page horizontally or vertically for browsing videos/images. In particular, when a child uses entertainment applications, swipe will be the basic operation for switching views in games or browsing videos in a list, making it a promising candidate for child identification.

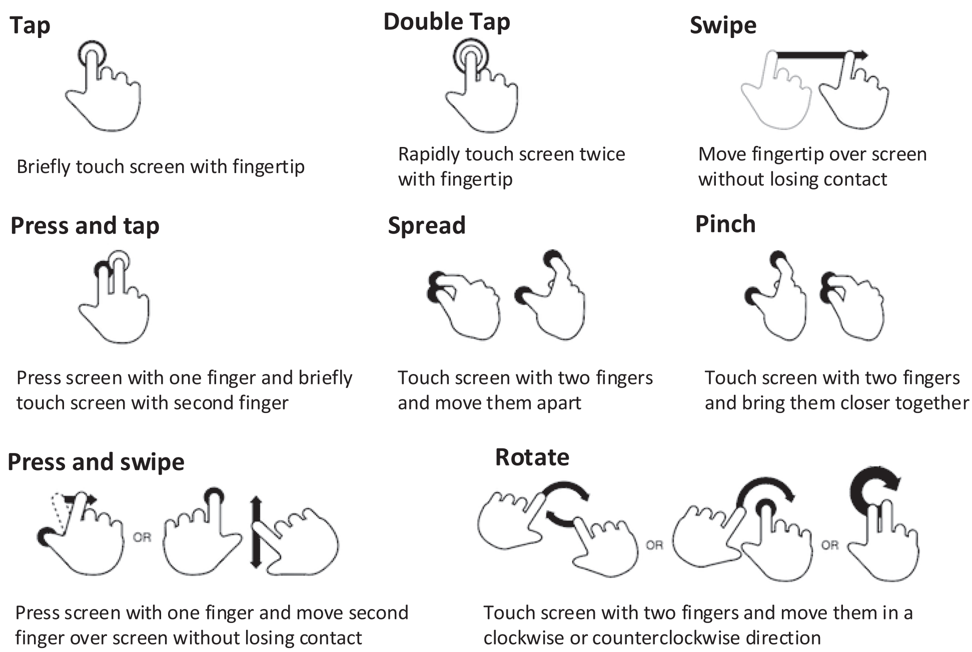


Fig. 4. Eight commonly used gestures on smartphones. The first three can be conducted with one finger, whereas the rest require two fingers to perform.

Table 1. Sample of Swipe and Tap Data

Time	Action	X	Y	Pressure	Size	Finger ID
276416631	0	712	1257	0.775	0.339	0
276416644	1	710	1262	0.763	0.321	0
⋮	⋮	⋮	⋮	⋮	⋮	
276416702	1	668	1414	0.638	0.286	0
276416710	2	—	—	—	—	—
276381043	0	131	112	0.925	0.429	0
276381108	2	—	—	—	—	—

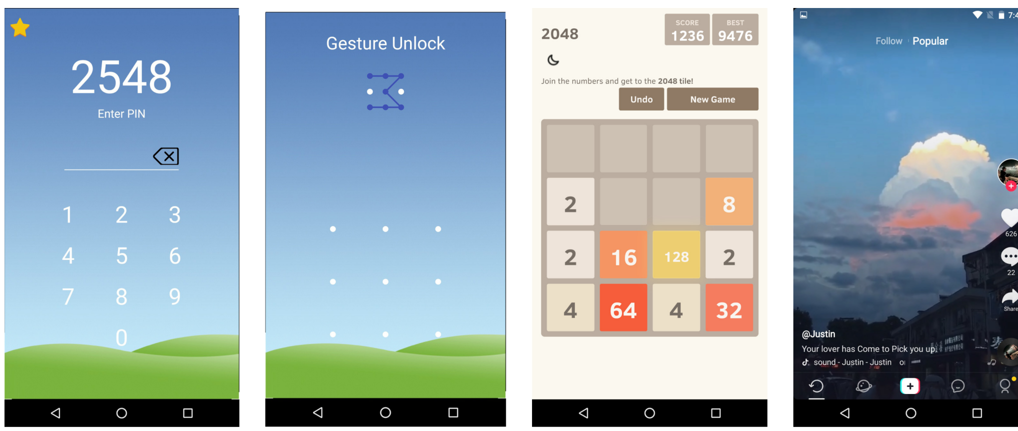
4.1.2 Gesture Data. To extract as much behavior information as possible from the selected gestures, we collect data from the touch sensor, the embedded accelerometer and gyroscope that depict a gesture from different aspects.

Touch data. The touchscreen records a touch action in seven dimensions: the occurred time, the action type, its X-Y coordinates, the pressure and size of the touch area, and the finger ID. Table 1 gives a sample of the swipe and tap data. A tap generally consists of two actions, touch down ("0") and up ("2"), whereas a swipe has one more action, touch move ("1"). A swipe consists of a sequence of touch points, which starts from touching the screen and ends with finger lifting. The event time is recorded in milliseconds and based on the smartphone's non-sleep uptime since boot. Both the pressure and size values are normalized to a range between 0 and 1, where 0 means no pressure or no size at all.

Accelerometer and gyroscope data. The accelerometer and gyroscope record a sampling point in five dimensions: the occurred time, the sensor type, and the amplitude in the X-, Y-, Z axis. Table 2

Table 2. Sample of Sensor (acce. and gyro.) Data

Time	Sensor Type	X	Y	Z
276391548	0	0.05269043	7.2640944	6.662945
276391558	1	−0.19813229	0.15116069	−0.048850477



(a) Numerical Unlock (b) Graphical Unlock (c) 2048 (d) TikTok

Fig. 5. Four tasks inserted with gesture capture function, including two phone-unlocking methods and two popular entertainment apps.

gives a sample of the accelerometer and gyroscope data, where the sensor type “0” indicates the accelerometer and “1” refers to the gyroscope.

For the sake of convenience, we collectively call the selected *swipe* and *tap* gestures *strokes*, and refer to the data collected by the touch sensor as the *touch data*, and the data collected by the accelerometer and gyroscope as the *sensor data* hereafter.

4.2 Data Collection Tasks

`iCare` supports both device- and application-based patrol control. In other words, `iCare` identifies child users upon their phone unlocking and interactions with a specific application. To fulfill the preceding goal and collect data close to smartphones’ real-world usage, we design four tasks on the Android platform: (1) for phone unlocking, we focus on two popular screen-unlocking methods (i.e., numerical and graphical unlocking),[1] and (2) for applications, we focus on game playing and video watching as representative applications since they are popular among children.

Task 1 (Numerical Unlock). This task requires users to unlock the phone with a given numeric PIN code. As shown in Figure 5(a), a PIN number appears at the top of the screen and users are required to input it. Our app generates two four-digit and two six-digit PIN numbers, and each PIN randomly appears twice. Each user unlocks the phone eight times if they succeed every time. We set the maximum number of unlocking attempts to be 10. As an incentive, we reward the user a star each time the correct PIN is entered. In this task, tap gestures introduced by number entering are collected.

Task 2 (Graphical Unlock). This task requires users to unlock the phone with graphical patterns. Figure 5(b) shows an example where a graphical pattern appears at the top of the screen and users

[1] The facial recognition and fingerprint methods are beyond the scope of this article.

are required to draw the same pattern but with no constraint on the direction. Our app generates six graphical patterns, and each pattern randomly appears once. Compared with Task 1, this task requires users to draw each pattern successfully no matter how many times they try. In this task, swipe gestures introduced by pattern drawing are collected.

Task 3 (Games). This task is based on a popular puzzle game named *2048* as shown in Figure 5(c). Users swipe vertically or horizontally to combine two attachable blocks with the same number (e.g., swipe up or down to merge the two vertical "2"s in the left-most column in Figure 5(c)). Then, a single one containing the sum will replace them automatically and a new number (2/4) randomly appears in one of the blank blocks after each valid swipe. The game is over when all spaces are filled with numbers and no move is available to merge the same two blocks. Users are allowed to swipe in any direction to play the game. In this task, swipe and tap gestures introduced by merging numbers and touching the screen, respectively, are collected.

Task 4 (Watching Videos). This task requires users to play a phenomenal video social network application named *TikTok* [33], as shown in Figure 5(d). TikTok is a platform for sharing videos that usually last for 15 seconds or less. It is popular among the young generation, and similar applications can be found in many countries. Users swipe vertically to switch videos (i.e., swipe up to watch new videos and down to look back). The videos are auto-played and users can pause/resume by tapping the screen. Users can also follow the uploader, give likes, leave comments, or share the video by tapping the buttons on the right side. Thus, swipe and tap are the main interactions of TikTok, and swipe gestures introduced by switching videos and tap gestures introduced by pausing/resuming videos or giving likes are collected in this task.

4.3 Feature Extraction

To capture the underlying touch behavior differences among children and adults resulted from physiological maturity, we examine three types of age-related characteristics: hand geometry, finger dexterity, and hand stability. We represent the first two with 35 fine-grained features inferred from the touch data and the last one with 18 features inferred from the sensor data. Table 3 describes the features extracted from the touch and sensor data. Since a tap usually lasts for a short time, we extract an 8-feature subset from the 35 touch features for the tap gesture. In total, we design 53 features for a swipe (35 touch features, 18 sensor features) and 26 features for a tap (8 touch features, 18 sensor features).

Hand geometry. Due to the different developmental stages in physiology among children and adults, their hand geometry tends to be distinctive in terms of hand size, finger length, and strength: children have smaller hands, as well as shorter and weaker fingertips, than adults in general. Consequently, children tend to touch the screen in a narrower range and swipe for a shorter trajectory length. An interesting finding is that the weaker fingers of children do not necessarily result in the lower touch pressure. In contrast, they press the screen harder. This is possibly because they are aware of their weak strength and meanwhile are less confident about their operations. Overall, the variances of hand geometry result in differences in terms of the touch range, touch distance, touch pressure, and size. Features numbered from 1 to 19 in Table 3 are extracted based on hand geometry, with *touch size at mid-stroke*, *pressure at mid-stroke*, and *average touch size* being the three most important features.

Finger dexterity. The finger dexterity associated with children's interactions with smartphones is poor compared to that of adults. According to our observation, children perform each touch stroke on smartphones more slowly and less flexibly than adults do. This is the same case when they try to switch between two touch actions, such as from tap to swipe. This can be attributed to the fact that children's bodies and fingers have not matured and become fully developed, and they have less contact with electronic devices compared with adults. As a result, it impacts their reactions when

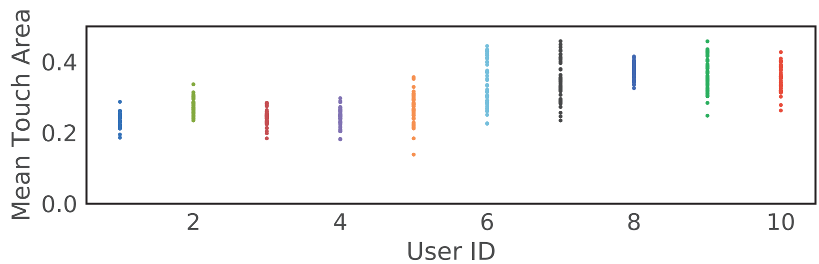
Table 3. Extracted Features from Touch (nos. 1–35) and Sensor Data (nos. 36–53) Based on Hand Geometry (nos. 1–19), Finger Dexterity (nos. 200035), and Hand Stability (nos. 36–53)

No.	Feature Description	No.	Feature Description
1	relative start position of x (Tap)	11	std of touch size
2	relative stop position of x	12	pressure at down (Tap)
3	relative start position of y (Tap)	13	pressure at mid-stroke
4	relative stop position of y	14	average pressure
5	direct end-to-end distance	15	std of pressure
6	trajectory distance	16	x displacement of two consecutive downs (Tap)
7	direction of end-to-end line	17	y displacement of two consecutive downs (Tap)
8	touch size at down (Tap)	18	x displacement of down and last up
9	touch size at mid-stroke	19	y displacement of down and last up
10	average touch size		
20	average velocity	28	median velocity of last 3 points
21	maximum pairwise velocity	29	velocity at mid-stroke
22	relative time of feature 21	30	median acceleration of first 3 points
23	std of pairwise velocity	31	median acceleration of last 3 points
24	maximum pairwise acceleration	32	acceleration at mid-stroke
25	relative time of feature 24	33	mean resultant length
26	std of pairwise acceleration	34	stroke duration (Tap)
27	median velocity of first 3 points	35	inter stroke time (Tap)
36	average pitch	45	RMS of pitch
37	average yaw	46	RMS of yaw
38	average roll	47	RMS of roll
39	std of pitch	48	minimum of pitch
40	std of yaw	49	minimum of yaw
41	std of roll	50	minimum of roll
42	average deviation of pitch	51	maximum of pitch
43	average deviation of yaw	52	maximum of yaw
44	average deviation of roll	53	maximum of roll

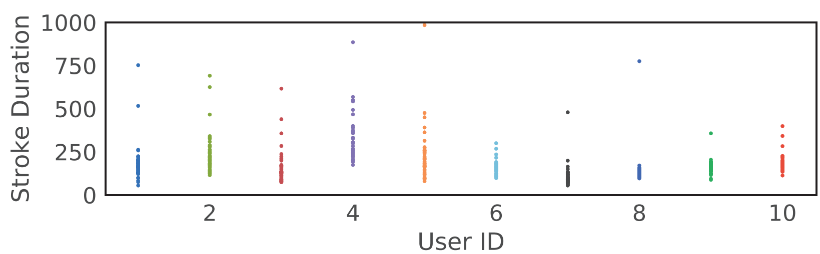
implementing a task using fingers. Considering finger dexterity, we extract 16 features numbered from 20 to 35 in Table 3. The finger dexterity mainly impacts the velocity, acceleration, and task duration, with *inter stroke time*, *median acceleration of last 3 points*, and *stroke duration* being the three most important features.

Hand stability. As a joint effect of hand geometry and finger dexterity, the hand stability of children is poor compared to that of adults when holding smartphones. Due to the ever-increasing screen size, children with small hands have difficulty holding smartphones smoothly and stably in the palm of their hand. The situation is even worse when they are asked to perform tasks on smartphones. We observe that children are more likely to wobble phones with large angles and be attracted by surroundings, resulting in frequent lift up and down. Considering the discrepancy in hand stability, we extract 18 features numbered from 36 to 53 in Table 3 in terms of the triaxial attitude angles derived from the accelerometer and gyroscope, with *maximum of yaw*, *yaw*, and *minimum of yaw* being the three most important features.

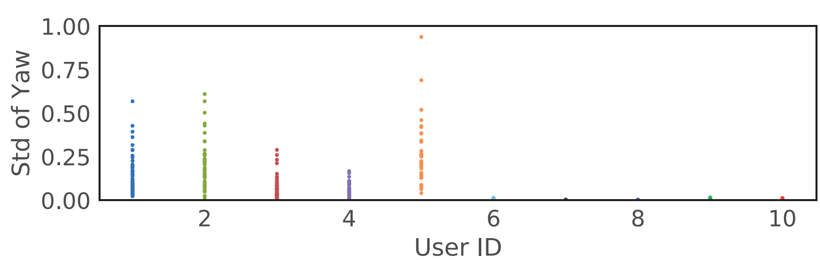
We take one feature from each feature category as an example and have a close examination. Figure 6 shows the distributions of *average touch size*, *stroke duration*, and *std of yaw* among 10



(a) Average touch sizes for 5 children and 5 adults. Adults have larger touch areas compared with children as a result of hand-size discrepancy.



(b) Stroke durations for 5 children and 5 adults. Children spend more time performing the same gesture due to slow reaction and poor finger dexterity.



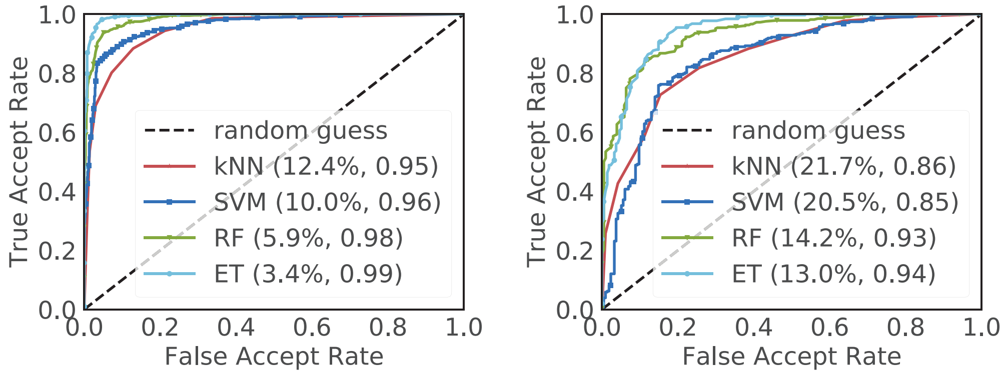
(c) Standard deviations of yaw for 5 children and 5 adults. Children shake phones more frequently since they are active and have poor hand stability.

Fig. 6. Average touch sizes (hand geometry feature), stroke durations (finger dexterity feature), and standard deviations of yaw (hand stability feature) for 10 users. Users 1 through 6 are children younger than 10 years, and the rest are adults older than 20 years.

users, including five adults older than 20 years and five children younger than 10 years. These features are extracted from 500 swipes (50 swipes per user). The results confirm our preceding analysis that children have relatively smaller touch areas due to their smaller hands, spend more time on each operation due to their poor finger dexterity, and wobble phones more frequently due to their bad hand stability.

4.4 Classifier Choices

To classify children and adults based on the above features, we implement four machine learning classifiers, namely k-Nearest-Neighbors (kNN), Support Vector Machine (SVM), Random Forests (RF), and ET.



(a) Based on swipes

(b) Based on taps

Fig. 7. ROC curves of four classifiers. Legends are in the form “label (EER, AUC),” where *label* indicates one of the four classifiers, *EER* is the equal error rate, and *AUC* is the area under the ROC.

kNN. *k*NN is a non-parametric method that has no underlying assumption of data, and each sample is assigned to one class based on a majority vote of its neighbors [2]. *k*NN does not do any generalization based on the training data, and therefore it is simple, fast, and yet with highly competitive results.

SVM. SVM is a popular and powerful tool for binary classification, which outputs an optimal hyperplane that maximizes the margin between two classes [7]. More importantly, it can solve the non-linearly separable problem by mapping data into a higher-dimensional space using the kernel trick. Here, we use the radial basis function (RBF) as our non-linear kernel.

RF. RF is an efficient algorithm that can classify large amounts of data with accuracy [15]. It is an ensemble method that constructs a number of decision trees during training time and outputs the class label with the most votes from all models. It also estimates the importance of each feature in classification.

ET. ET is another tree-based ensemble classification approach that fits a number of randomized decision trees on various sub-samples of the dataset and uses averaging to improve prediction accuracy and avoid over-fitting [12]. Compared with RF, this method drops the idea of using bootstrap copies of the learning sample and tests random splits over a fraction of features instead of all possible splits. As a result, ET is generally cheaper to train from a computation point of view.

We compare these classifiers in Section 5, and the results are shown in Figure 7. For the sake of high classification accuracy and low computation cost, we employ ET as our classification algorithm.

5 EVALUATION

To evaluate the performance of `iCare`, we conduct experiments with 68 child users and 32 adult users across 6 months, and our study obtained the approval of the Institutional Review Board (IRB). In addition to the overall performance with a single stroke, we investigate the performance improvement with multiple consecutive strokes, the scalability of `iCare`, and the performance discrepancy among various tasks, age/gender/race groups, and age boundaries. In summary, the performance of `iCare` can be summarized as follows:

- `iCare` can achieve an equal error rate (EER) of 3.4% with a single swipe and 13.0% with a single tap.

Table 4. Summary of Experimental Smartphones

No.	Manufacturer	Model	OS	Size	Resolution	PPI*
1	LG	Google Nexus 5X	Android v.7.1.1	5.2"	1080 × 1920	424
2	LG	Google Nexus 5	Android v.6.0.1	4.95"	1080 × 1920	445
3	HUAWEI	P10 Plus	Android v.8.0.0	5.5"	2560 × 1440	540
4	Samsung	Galaxy S6 Edge+	Android v.7.0	5.7"	2560 × 1440	518

*Pixels per inch.

Table 5. Participants and Collected Data

Group	Age	No. of Users	No. of Males	No. of Females	No. of Whites	No. of Asians	No. of Taps	No. of Swipes
G-Children	3~5	12	7	5	7	5	1,214	1,506
	6~9	23	13	10	7	16	1,351	4,766
	10~13	14	7	7	2	12	874	2,737
	14~17	13	5	8	0	13	945	2,792
G-Adults	18~29	27	18	9	1	26	1,279	10,282
	30~39	7	6	1	1	6	469	2,077
	40~49	3	2	1	0	3	201	582
	50~59	1	1	0	1	0	54	156
Total	3~59	100	59	41	19	81	6,387	24,898

- The accuracy of `iCare` can be further improved by multiple consecutive strokes. `iCare` achieves an EER of 1.7% with three consecutive swipes and 3.9% with three consecutive taps.
- `iCare` shows no obvious performance discrepancy among different tasks, which demonstrates its generality.
- `iCare` is able to scale to new users and phones, which validates its scalability for real-world deployment.
- The identification performance across various age, gender, and race groups has a slight discrepancy.
- `iCare` is able to classify adults and children with different boundaries (e.g., boundaries other than 17) and multiple age boundaries (e.g., 13 and 17).

5.1 Experimental Setup

5.1.1 Device Setup. During the experiment, we use four smartphones: (1) Google Nexus 5X, (2) Google Nexus 5, (3) HUAWEI P10 Plus, and (4) Samsung Galaxy S6 edge+. The detailed information of each device is revealed in Table 4, among which the Nexus 5X is the main device used for data collection, and the other three phones with various sizes are used to evaluate the scalability of `iCare` to new devices.

We implement Tasks 1 through 4 on the Android platform. The first two are based on self-developed apps, whereas the latter are modified based on existing apps. The data collection module of `iCare` runs in the background of the experimental devices. We turn off the rotation function during data collection to eliminate biases associated with various holding ways.

5.1.2 Participant. We recruit two groups of users in our experiment: *G-Children* with 62 children from 3 to 17 years old and *G-Adults* with 38 adults from 18 to 59 years old. Table 5 summarizes the demographics of the participants. We then divide them into eight fine-grained sub-groups given their different growth stages. In *G-Children*, we have more children aged 6 to 9 years since they are more active and more likely to cause privacy or monetary damage to their parents, compared to younger and elder children. In addition, each group contains both male and female users from two races: Asian and White [29].

All participants are required to remain seated while holding the phone in their hands to play. All child participants report that they have previously used touchscreen phones. Before the start of the experiment, we briefly explain to participants how to play with the apps and let them become familiar with them.

5.1.3 Datasets. All of the participants finish the four tasks. In Task 1, we collect 3,469 taps among the *G-Children* group and 1,504 taps among the *G-Adults* group. Note that although the number of children aged 3 to 5 years is not the largest, we collect the most taps per child compared to older children and adults. That is because children in this age group are more likely to enter PINs incorrectly and hence have more attempts. In Task 2, we collect 520 and 268 swipes from the group of *G-Children* and *G-Adults*, respectively. In Task 3, we collect 138 taps and 9,200 swipes from the group of *G-Children*, and 103 taps and 11,684 swipes from the group of *G-Adults*. The *G-Adults* group does better in the 2048 game in general, and thus more swipes are collected among this group. However, a fair amount of children state that they have played this particular game or a similar one before. Most of the children between 3 and 5 years are able to reach 64 or larger, whereas the ones between 6 and 11 years reach at least 128, and the highest block reached is 512 by a 5th grader (10 years old). In Task 4, we collect 777 taps and 2,801 swipes from the group of *G-Children*, and 396 taps and 1,145 swipes from the group of *G-Adults*. Among all tasks, Task 4 is the easiest to perform for most children in the *G-Children* group. As not all children between 3 and 5 years are sensitive to numbers or patterns, some of them are slow in completing former tasks, especially Task 1. However, nearly all children younger than 18 years develop an immediate attraction to Task 4. Most children between 6 and 17 years state that they have used TikTok or similar applications at home and are able to use the search function expertly. Children between 3 and 9 years show undivided attention when watching videos and giving likes, which validates a comment made by a kindergarten teacher that young children are easily attracted to live videos and audios. Overall, we collect 6,387 samples for the tap dataset and 24,898 samples for the swipe dataset.

5.2 Metrics

To evaluate the performance of the binary classifiers, we choose the commonly used performance metrics: the area under the curve (AUC) and the EER. The receiver operating characteristic (ROC) curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied, and is created by plotting the true accept rate (TAR) against the false accept rate (FAR). The TAR is the probability of correctly identifying a child, whereas the FAR is the probability that the classifier incorrectly accepts a child. The AUC is a value between 0 and 1, and a large value indicates better performance. The EER is the rate when both accept and reject errors are equal, and the lower this value, the better the classifier.

5.3 Results

5.3.1 Impact of Classifier Choices. In this section, we evaluate the performance of the aforementioned four classification methods (i.e., *k*NN, SVM, RF, and ET) in identifying child users.

Setup. We label our datasets based on which group a stroke sample is from. All stroke samples from *G-Adults* are labeled as class 0, and the strokes from *G-Children* are labeled as class 1. Then, we randomly divide the labeled tap and swipe datasets into two parts: 80% for model training and

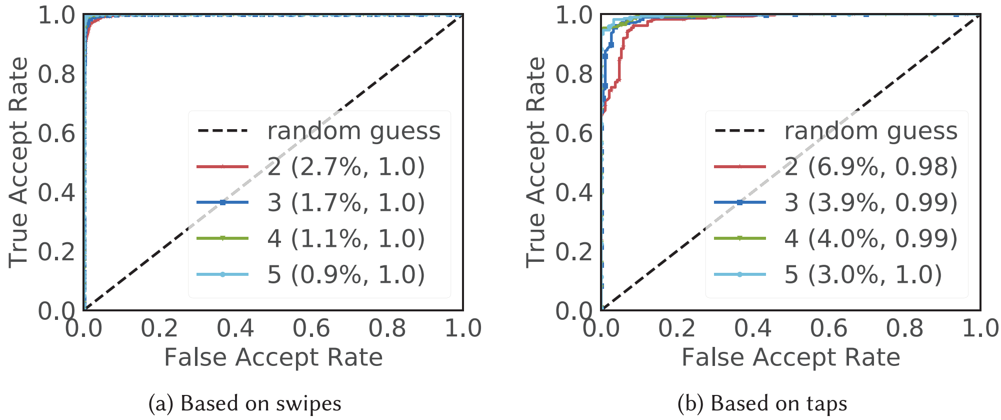


Fig. 8. ROC curves with different numbers of swipes/taps. Legends are in the form "label (EER, AUC)," where *label* indicates the number of consecutive swipes/taps used.

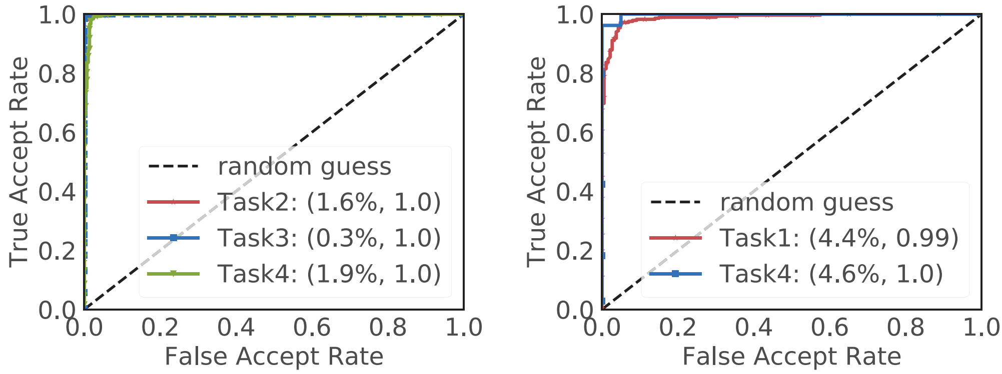
20% for model testing. For the swipe dataset, we set the parameter $k = 7$ for kNN, the complexity parameter $c = 1.4$, the gamma parameter $\gamma = 0.15$ for SVM, and the number of trees to 200 for RF/ET. For the tap dataset, $k = 9$, the choice of $c = 1.4$ and $\gamma = 0.7$, and the tree number is 100. All parameters are determined and optimized by the grid search with the stratified fivefold cross validation using the training datasets. As the training and testing datasets are both small, the time cost is less than 1 second for both training and classification with all four classifiers.

Results. Figure 7 shows the ROC curves of the four classifiers based on the swipe and tap datasets, respectively. The results of kNN, SVM, RF, and ET are colored in red, blue, green, and cyan, respectively. For both datasets, RF and ET classifiers outperform kNN and SVM. All of the classifiers achieve better performance on the swipe dataset than on the tap dataset because a swipe usually contains more data points, and thus more features can be extracted for classification. Overall, RF achieves a 5.9% EER and a 0.98 AUC on the swipe dataset, whereas ET shows better results with a 3.4% EER and a 0.99 AUC. Even with a single tap, the ET algorithm can achieve a 13.0% EER and a 0.94 AUC. The result indicates that it is promising to identify children through only a single stroke. Given the accuracy performance and the computation cost, we choose ET as our classification algorithm and use the aforementioned optimized parameters by default in the following experiments.

5.3.2 Multiple Strokes. Although with only one swipe `iCare` can achieve an EER of 3.4%, combining multiple consecutive strokes for classification is likely to further improve the accuracy. In this section, we explore the appropriate number of consecutive strokes used for classification.

Setup. In this set of experiments, we prepare the testing datasets by randomly choosing 10% labeled consecutive samples from participants at each age. Overall, the swipe testing dataset consists of 2,500 samples and the tap testing dataset consists of 660 samples, which are evenly distributed among children and adults. Then, we train an ET model using the remaining swipe and tap samples, respectively. Instead of classifying all swipes/taps individually and reaching a final decision by the majority vote, we combine multiple consecutive outputs at an earlier stage by their probabilities and average them as our final predicted probability.

Results. Figure 8 shows the ROC curves by varying the number of swipes/taps taken for a classification decision. The classification error can be significantly decreased by increasing the number of strokes. The EER decreases to 0.9% for swipe and 3.0% for tap, as we increase the number of



(a) Based on swipes (b) Based on taps

Fig. 9. ROC curves with different tasks. Legends are in the form "label (EER, AUC)," where *label* indicates one of the four tasks.

strokes to five. The results have clearly demonstrated that using multiple strokes improves the accuracy. To strike the balance between usability and accuracy, we choose the number of consecutive strokes used for classification to be 3 in the following experiments.

5.3.3 Performance of Different Tasks. Since various tasks involve different numbers and types of strokes, we evaluate the performance of each task, respectively.

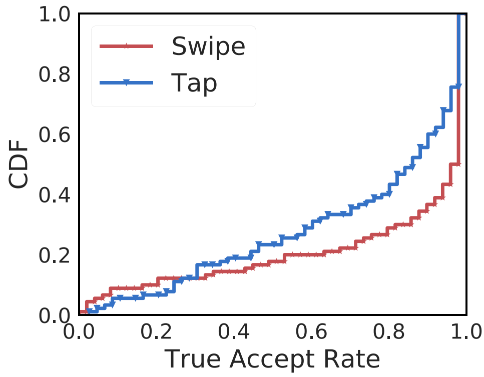
Setup. In this set of experiments, we first separate the labeled swipe/tap dataset into sub-datasets associated with each task. Then, we randomly divide the sub-datasets into two parts: 80% for model training and 20% for model testing. During experiments, swipes are collected in Tasks 2, 3, and 4, and taps are collected in Tasks 1, 3, and 4. However, because not all users tap the screen during the 2048 game and the tap samples collected are quite limited, we evaluate the performance of Task 3 based on swipes only.

Results. As shown in Figure 9(a), with three consecutive swipes, Tasks 2, 3, and 4 can achieve an EER below 1.9% and an AUC of 1.0. Task 3 (i.e., the 2048 game) shows the best accuracy since it has a much more constant pattern and we collect the most swipes from it. Similarly, with three consecutive taps, Task 1 achieves an EER of 4.4% and an AUC of 0.99, and Task 4 achieves an EER of 4.6% and an AUC of 1.0, as shown in Figure 9(b). In general, `iCare` shows no obvious performance discrepancy on different tasks based on either swipes or taps. Since the gestures we select (i.e., swipe and tap) are the most basic interactions in almost all of the applications, we believe `iCare` is generic and likely to scale to other applications.

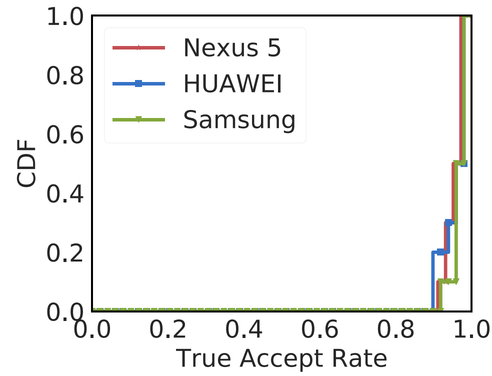
5.3.4 Scalability to New Users. In real life, `iCare` may need to cooperate with new users. To investigate the scalability of `iCare` to new users, we conduct the following experiments.

Setup. From the 100 users, we choose 99 users for training and designate the remaining one to be introduced as a new user to the trained system. In other words, we use the labeled stroke samples from 99 users to train the ET classifiers, and we use the stroke samples from the remaining user to test the scalability to new users. The classification threshold is set to 0.5 for swiping and 0.575 for tapping according to the EERs in Figure 8. To eliminate the random error, we repeat the experiment 100 times until each user serves as the new user once.

Results. The CDF of the TARs in Figure 10(a) reveals that, based on swipes, `iCare` can successfully classify 83% of new users with a TAR over 0.8, 64% of new users with a TAR over 0.9, and, more remarkably, 42% of new users with a TAR of 1.0. `iCare` shows a slightly lower accuracy when



(a) Scalability to new users



(b) Scalability to new phones

Fig. 10. Scalability of iCare to new users (a) and new smartphones (b).

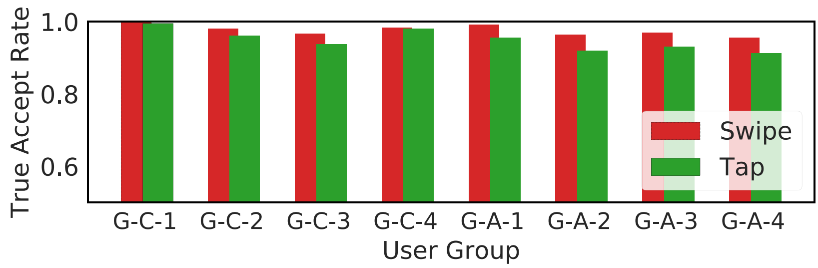
based on taps, but half of the new users can still be classified with a TAR over 0.87. We believe that the scalability of iCare is likely to be limited by the small user size at present. In real-world deployment, we can build iCare with a large user dataset and combine multiple types of gestures. More importantly, the pre-trained iCare model can be updated during user usage to adapt to new users and scenarios. Specifically, the pre-trained model can only be used when iCare initializes on new devices. During usage, iCare can adjust its model with the new user data and build a more specific model for concrete scenarios such as the in-home scenario. With the update technique, we believe that iCare can scale to new users better in practice.

5.3.5 Scalability to New Phones. Users may access more than one phone in real life, and as a result, iCare shall classify users across devices (i.e., users enrolled with one phone shall be able to be classified on other phones). To investigate the scalability of iCare to new phones, we conduct the following experiments.

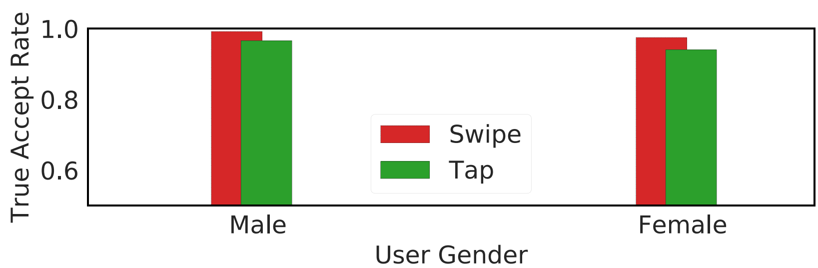
Setup. From the 100 users, we randomly choose 10 users including both child and adult users to conduct Task 3 on four phones besides the Nexus 5X. The detailed information of each phone is revealed in Table 4. To mimic the real cases, the four used smartphones are of various sizes given that different screen sizes may impact the way users hold and use the phone. We train the ET classifier with swipe samples collected from the Nexus 5X, and we test it with data collected from the other three phones, respectively. The classification threshold is set to 0.5 according to the EERs in Figure 8.

Results. We plot the CDF of the TARs for each phone in Figure 10(b), which reveal that iCare can successfully classify users across devices with an average TAR of 96.5% for Nexus 5, 96.6% for HUAWEI, and 97.9% for Samsung. The results indicate that different devices and screen sizes do not necessarily impact the identification of iCare. Together with the ability to identify new users, we believe iCare is scalable in real-world deployment.

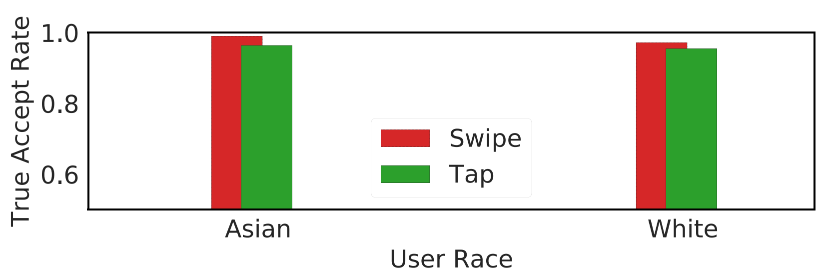
5.3.6 Influence of Different Age Groups, Genders, and Races. Due to the difference in growth stages, users from various age groups may show discrepancies in the performance of iCare. To understand how users from different age groups behave on the classification results, we perform an analysis on the eight fine-grained sub-groups: children 3 to 5 years old (*G-C-1*), children 6 to 9 years old (*G-C-2*), children 10 to 13 years old (*G-C-3*), children 14 to 17 years old (*G-C-4*), adults 18 to 29 years old (*G-A-1*), adults 30 to 39 years old (*G-A-2*), adults 40 to 49 years old (*G-A-3*), and adults 50 to 59 years old (*G-A-4*).



(a) Performance of different age groups



(b) Performance of different genders

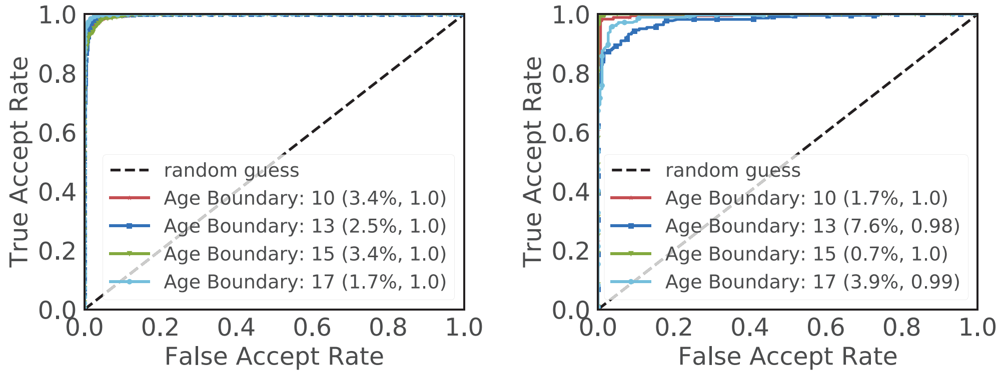


(c) Performance of different races

Fig. 11. Performance of different age groups (a), genders (b), and races (c).

Setup. In this set of experiments, we use 80% of samples of the labeled tap and swipe datasets for training and use the rest for testing. The tree number for the ET classifier is set to be 200 for swiping and 100 for tapping, the number of consecutive strokes used for classification is 3, and the classification threshold is 0.5 for swiping and 0.575 for tapping according to the EERs in Figure 8.

Results. Figure 11(a) shows the TAR for each sub-group. For both swiping and tapping, we have much better performance in classifying the *G-C-1* group. This is reasonable as younger children tend to be more different from adults in terms of hand geometry, finger dexterity, and hand stability. By contrast, `iCare` has slightly lower performance in classifying the *G-C-3* and adult groups. We assume it is a result of individual differences. Children 10 to 13 years old are in a rapid stage of growth and development, and thus may show more individual differences compared with younger or elder child groups. Similarly, as we regard all users older than 18 years as adults and label them as the same class, such a large age range may introduce significant individual differences and thus result in the performance decrease. This finding enlightens us that dividing users into more fine-grained age groups and building corresponding multi-class classifiers may further help improve the performance of `iCare`.



(a) Based on swipes (b) Based on taps

Fig. 12. Performance of `iCare` with different age boundaries.

Another interesting point is whether users from different genders or races behave the same with `iCare`. To investigate this, we group the prediction results based on user gender and race, respectively. Figure 11(b) and (c) show the TARs for the male, female, Asian, and White group. Surprisingly, for both swiping and tapping, male users outperform female users and Asian users outperform White users. We conjecture that one of the reasons is that we have more male and Asian participants. However, this finding indicates that besides age, we can potentially find common characteristics among similar gender or race groups of people.

5.3.7 Influence of Different Age Boundaries. In the preceding analysis, we set the age boundary that partitions children and adults to be 17. However, various countries or districts may have different definitions of the boundary. For instance, American society pays more attention to children younger than 13 years. To investigate the performance of `iCare` with different age boundaries, we conduct the following experiments.

Setup. We set the boundary to be 10, 13, 15 and 17, respectively. Then, we label the swipe/tap dataset according to the boundary (i.e., samples from users aged above it are labeled as class 0, and the rest are labeled as class 1). For each boundary, we use 80% of samples for training and 20% of samples for testing based on either taps or swipes.

Results. The classification results shown in Figure 12 reveal that, based on swipes, `iCare` achieves the best EER with the boundary 17 but the discrepancy is not obvious among all age boundaries. For taps, the best EER occurs with the boundary 15, and the performance is slightly lower given the age boundary 13. Nevertheless, `iCare` shows good performance in general and is feasible to cope with different age boundaries.

5.3.8 Influence of Multiple Age Boundaries. Similar to the CARA film rating system, some countries or districts may rate an application's suitability for users or suggest appropriate screen time for users with multiple age boundaries (i.e., the reasonable screen time might be 1 hour per day for children younger than 14 years and 2 hours for children 14 to 17 years old). To explore the possibility of `iCare` to identify children with multiple age boundaries, we conduct the following experiments.

Setup. We set two age boundaries, 13 and 17, and label the samples from children younger than 14 years as class 0, samples from children 14 to 17 years old as class 1, and the rest as class 2. We use 80% of samples for training and 20% of samples for testing, and we set the number of consecutive

Table 6. Multi-Boundary Identification Based on Swipes

Group	TAR	FAR
G-Children-13	0.972	0.001
G-Children-17	0.991	0.000
G-Adults	0.999	0.024

Table 7. Multi-Boundary Identification Based on Taps

Group	TAR	FAR
G-Children-13	0.962	0.029
G-Children-17	0.951	0.003
G-Adults	0.967	0.033

Table 8. Age Distribution of Surveyed Users and Their Children

Group	Age (years)	Number
Adult (User)	25–29	8 (6.3%)
	30–34	44 (34.9%)
	35–39	41 (32.5%)
	40–44	28 (19.8%)
	45–48	8 (6.3%)
Child	0–2	8 (6.3%)
	3–5	46 (36.5%)
	6–9	41 (32.5%)
	10–13	16 (12.7%)
	14–17	15 (11.9%)

strokes used for classification to be 5 and the classification threshold to be 0.33 for both swiping and tapping.

Results. Table 6 and Table 7 reveal that, with five consecutive strokes, `iCare` is able to accurately classify children with multiple age boundaries. This indicates that `iCare` is beyond the scope of identifying children in binary and is capable of multi-age child classification that can augment more application scenarios.

6 USER STUDY

We measure the usability and user acceptance rate of `iCare` by conducting a questionnaire survey among parents with children younger than 18 years, who are also the target users of `iCare`. The total amount of valid questionnaires is 126, and the participants cover a large age range without losing generality. The age distribution of the parents participating and their children are shown in Table 8, respectively.

During the survey, we ask five yes-no questions for each participant, and the detailed description of each question and its answers is summarized in Table 9. The results show that 108 of 126 parents are reluctant to let their children use smartphones at a young age, and 96 participants report that their children will use their smartphones intentionally or unintentionally. Further analysis

Table 9. Summary of the User Survey

No.	Question	Yes	No
1	Are you willing to let your kid(s) use a smartphone at the present age?	18	108
2	Will your kid(s) use your smartphone at ordinary times?	96	30
3	Do you feel like restricting your kid(s) access to some specific applications, such as games?	104	22
4	Do you recognize the significance of `iCare`?	120	6
5	Are you willing to use `iCare`?	114	12

Table 10. User Selection of Applications of Which They Intend to Restrict Children's Access

Application	User Selections
Game	104 (100%)
Video	97 (93.3%)
Photo	19 (18.3%)
Finance	61 (58.7%)
Others	35 (33.7%)

reveals that the older the children are, the less their parents will care about their smartphone usage. However, even parents with an open mind about children's smartphone usage like to restrict their access to some specific applications, such as games. Among the 104 parents who answer yes to Question 3, we ask an additional multi-choice question about what applications they like to restrict their children' access and offer five options: (1) Game, (2) Video, (3) Photo, (4) Finance, and (5) Others, as shown in Table 10. Surprisingly, all 104 participants choose "Game," and more than 93% of users select "Video." The goal of `iCare` is in line with the survey results, and the gestures we select (i.e., swipe and tap) are common in both games and video applications. In addition, 120 of 126 users recognize the significance of our work, and 114 users (>90%) are willing to use `iCare` for parent patrol.

In summary, parents with children younger than 18 years do have the requirement to restrict their children's access to smartphones, and among applications on the market, games and videos are their major concerns. As an automatic and implicit parent patrol mechanism, more than 95% of surveyed parents recognize the significance of `iCare` and more than 90% of users feel like using it.

7 LIMITATIONS AND FUTURE WORK

Based on the existing framework, several issues remain to be explored.

Diversity of users. As children's dexterity develops significantly every year and adult users cover a large age range, dividing them into finer age groups and building a corresponding multi-class classifier possibly increases overall accuracy. In addition, we observe that some children 3 to 5 years old are slow in reading numbers, yet aged users (i.e., adults older than 70 years) may show similar behaviors. Since we have not tested users older than 59 years yet, it is interesting to include more aged users and analyze how poor numeracy will affect touch and sensor patterns, if any.

Discrepancies across genders/races. Given the performance discrepancy across genders and races, it is worth checking the reasons behind it. One possible cause is the unbalanced training dataset, thus more female and White users can be recruited to verify it. Another assumption is that, similar to age, common characteristics are shared among the same gender or race group of people. In this case, building separate classifiers for different genders or races may further improve the overall accuracy.

Limited gestures. There are many other gestures (e.g., double tap, pinch, and spread) that have not been explored in our study. In reality, users may have to change among different types of gestures back and forth to complete a task on smartphones. Fusing all types of gestures can result in a faster classification decision and possibly improve accuracy.

Improving accuracy. Given three consecutive swipes/taps and the accuracy, our method can be a good supplement for existing parent patrol apps. However, before real-world deployment, the accuracy shall be improved. We can improve the accuracy by solving the preceding two limitations. Additionally, other information on smartphones, such as device pairing, can be exploited to derive users' characteristics for refining classification results.

8 CONCLUSION

To improve the usability of existing parent patrol applications, this article introduces an implicit approach to detect whether a child is operating a smartphone without special user attention. Specifically, we extract the features of hand geometry, finger dexterity, and hand stability from swipe and tap gestures, and we explore four machine learning algorithms to classify children. Our pilot study with 62 children and 38 adults shows that the trained ET model can achieve an EER of 3.4% with a single swipe and the performance can be improved to 1.7% by using three consecutive swipes. As well, `iCare` shows no obvious performance discrepancy on the applications that are inserted with the gesture capture function and can extend to new users and phones, which demonstrates its generality and scalability for real-world deployment. In addition to the current implementation on smartphones, we believe `iCare` can be easily extended to other smart devices, such as tablets and smart watches. Future work includes exploring `iCare` with more users and devices.

REFERENCES

[1] Ammar Ahmad, Cyrille Migniot, and Albert Dipanda. 2017. Tracking hands in interaction with objects: A review. In *Proceedings of the 13th International Conference on Signal-Image Technology and Internet-Based Systems (SITIS'17)*. IEEE, Los Alamitos, CA, 360–369.

[2] Naomi S. Altman. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *American Statistician* 46, 3 (1992), 175–185.

[3] Lisa Anthony, Quincy Brown, Berthel Tate, Jaye Nias, Robin Brewer, and Germaine Irwin. 2014. Designing smarter touch-based interfaces for educational contexts. *Personal and Ubiquitous Computing* 18, 6 (2014), 1471–1483.

[4] Can Basaran, Hee Jung Yoon, Ho Kyung Ra, Sang Hyuk Son, Taejoon Park, and JeongGil Ko. 2014. Classifying children with 3D depth cameras for enabling children's safety applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'14)*. 343–347.

[5] Roberto Cameriere and Luigi Ferrante. 2008. Age estimation in children by measurement of carpals and epiphyses of radius and ulna and open apices in teeth: A pilot study. *Forensic Science International* 174, 1 (2008), 60–63.

[6] Classification and Ratings Administration. 1968. The Film Rating System. Retrieved July 9, 2020 from https://filmratings.com/.

[7] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20, 3 (1995), 273–297.

[8] Andrew Dalton. 2017. Amazon Refunds $70 Million Your Kids Accidentally Spent in Apps. Retrieved July 9, 2020 from https://tinyurl.com/ybqnpzea.

[9] Martha Bridge Denckla. 1973. Development of speed in repetitive and successive finger-movements in normal children. *Developmental Medicine & Child Neurology* 15, 5 (1973), 635–645.

[10] Mobile Fence. 2018. Mobile Fence Parental Control. Retrieved July 9, 2020 from https://play.google.com/store/apps/details?id=com.mobilefence.family.

[11] Mario Frank, Ralf Biedert, Eugene Ma, Ivan Martinovic, and Dawn Song. 2013. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security* 8, 1 (2013), 136–148.

[12] Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. *Machine Learning* 63, 1 (2006), 3–42.

[13] Agneta Gustus, Georg Stillfried, Judith Visser, Henrik Jörntell, and Patrick van der Smagt. 2012. Human hand modelling: Kinematics, dynamics, applications. *Biological Cybernetics* 106, 11–12 (2012), 741–755.

[14] Javier Hernandez-Ortega, Aythami Morales, Julian Fierrez, and Alejandro Acien. 2017. Predicting age groups from touch patterns based on neuromotor models. In *Proceedings of the 8th International Conference on Pattern Recognition Systems (ICPRS'17)*. 1–6.

[15] Tin Kam Ho. 1995. Random decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition (ICDAR'95)*, Vol. 1. IEEE, Los Alamitos, CA, 278–282.

[16] B. Hohendorff, C. Weidermann, K. J. Burkhart, P. M. Rommens, K. J. Prommersberger, and M. A. Konerding. 2010. Lengths, girths, and diameters of children's fingers from 3 to 10 years of age. *Annals of Anatomy (Anatomischer Anzeiger)* 192, 3 (2010), 156–161.

[17] Sheik N. Imrhan. 1989. Trends in finger pinch strength in children, adults, and the elderly. *Human Factors* 31, 6 (1989), 689–701.

[18] ITU and UNICEF. 2015. Guidelines for Industry on Child Online Protection. Retrieved July 9, 2020 from https://tinyurl.com/ya3s6dz5.

[19] Hilda K. Kabali, Matilde M. Irigoyen, Rosemary Nunez-Davis, Jennifer G. Budacki, Sweta H. Mohanty, Kristin P. Leister, and Robert L. Bonner. 2015. Exposure and use of mobile media devices by young children. *Pediatrics* 136, 6 (2015), 1044–1050.

[20] J. P. Kuhtz-Buschbeck, A. Boczek-Funcke, M. Illert, K. Joehnk, and H. Stolze. 1999. Prehension movements and motor development in children. *Experimental Brain Research* 128, 1–2 (1999), 65–68.

[21] Screen Time Labs. 2018. Screen Time Parental Control. Retrieved July 9, 2020 from https://play.google.com/store/apps/details?id=com.screentime.rc.

[22] Lingjun Li, Xinxin Zhao, and Guoliang Xue. 2013. Unobservable re-authentication for smartphones. In *Proceedings of the 20th Annual Network Distributed System Security Symposium (NDSS'13)*.

[23] Xiaopeng Li, Sharaf Malebary, Xianshan Qu, Xiaoyu Ji, Yushi Cheng, and Wenyuan Xu. 2018. iCare: Automatic and user-friendly child identification on smartphones. In *Proceedings of the 19th International Workshop on Mobile Computing Systems and Applications (HotMobile'18)*. ACM, New York, NY, 43–48.

[24] Lisa Link, Shirley Lukens, and Mary Ann Bush. 1995. Spherical grip strength in children 3 to 6 years of age. *American Journal of Occupational Therapy* 49, 4 (1995), 318–326.

[25] Hugo Meinedo and Isabel Trancoso. 2010. Age and gender classification using fusion of acoustic and prosodic features. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech'10)*. 2818–2821.

[26] Microsoft. 2018. How Old Do I Look? Retrieved July 9, 2020 from https://how-old.net/.

[27] Seyed Mostafa Mirhassani, Alireza Zourmand, and Hua-Nong Ting. 2014. Age estimation based on children's voice: A fuzzy-based decision fusion strategy. *Scientific World Journal* 2014 (2014), 534064.

[28] The Church of Jesus Christ of Latter-Day Saints. 1999. *Lesson 2: Age characteristics of children. In Teaching, No Greater Call: A Resource Guide for Gospel Teaching*. The Church of Jesus Christ of Latter-Day Saints, 110–16.

[29] U.S. Office of Management and Budget. 1997. Race and Ethnicity Classifications. Retrieved July 9, 2020 from http://www.iowadatacenter.org/aboutdata/raceclassification.

[30] American Academy of Pediatrics. 2018. Children and Media Tips from the American Academy of Pediatrics. Retrieved July 9, 2020 from https://tinyurl.com/yd6lej9y.

[31] Avar Pentel. 2017. Predicting age and gender by keystroke dynamics and mouse patterns. In *Proceedings of the 25th Conference on User Modeling, Adaptation, and Personalization (UMAP'17)*. ACM, New York, NY, 381–385.

[32] Janet L. Poole, Patricia A. Burtner, Theresa A. Torres, Cheryl Kirk McMullen, Amy Markham, Michelle Lee Marcum, Jennifer Bradley Anderson, and Clifford Qualls. 2005. Measuring dexterity in children using the nine-hole peg test. *Journal of Hand Therapy* 18, 3 (2005), 348–351.

[33] BYTEMOD PTE. LTD. 2018. TikTok. Retrieved July 9, 2020 from https://www.tiktokv.com/.

[34] Zdeňka Sitová, Jaroslav Šeděnka, Qing Yang, Ge Peng, Gang Zhou, Paolo Gasti, and Kiran S. Balagani. 2016. HMOG: New behavioral biometric features for continuous authentication of smartphone users. *IEEE Transactions on Information Forensics and Security* 11, 5 (2016), 877–892.

[35] Michael Ungar. 2016. Why Are Kids So Anxious These Days? Retrieved July 9, 2020 from https://tinyurl.com/yd8zlr5n.

[36] Willem Van Mechelen, Jos W. R. Twisk, G. Bertheke Post, Jan Snel, and Han C. G. Kemper. 2000. Physical activity of young people: The Amsterdam Longitudinal Growth and Health Study. *Medicine & Science in Sports & Exercise* 32, 9 (2000), 1610–1616.

[37] Radu-Daniel Vatavu, Lisa Anthony, and Quincy Brown. 2015. Child or adult? Inferring smartphone users' age group from touch measurements alone. In *Proceedings of the 15th International Conference on Human-Computer Interaction (INTERACT'15)*. 1–9.

[38] Radu-Daniel Vatavu, Gabriel Cramariuc, and Doina Maria Schipor. 2015. Touch interaction for children aged 3 to 6 years: Experimental findings and relationship to motor skills. *International Journal of Human-Computer Studies* 74 (2015), 54–76.

[39] Elyse Wanshel. 2018. 10 Reasons Why You Shouldn't Give a Child a Smartphone or Tablet. Retrieved July 9, 2020 from https://tinyurl.com/ybuborln.

[40] N. Zheng, K. Bai, H. Huang, and H. Wang. 2014. You are how you touch: User verification on smartphones via tapping behaviors. In *Proceeding of the 22nd IEEE International Conference on Network Protocols (ICNP'14)*. 221–232.

Received September 2019; revised May 2020; accepted May 2020