Learning Heuristically-selected and Neurally-guided Feature for Age Group Recognition using Unconstrained Smartphone Interaction

Yingmao Miao[1] , **Qiwei Tian**[1] , **Chenhao Lin**[1] , **Tianle Song**[1] , **Yajie Zhou**[1] , **Junyi Zhao**[1] , **Shuxin Gao**[1] , **Minghui Yang**[2] and **Chao Shen**[1]
[1]School of Cyber Science and Engineering, Xi'an Jiaotong University
[2]Guangdong Oppo Mobile Telecommunications Corp., Ltd
{mym2017, michaeltqw, tianlesong, zyj1168, yanyan806333088, gaosx}@stu.xjtu.edu.cn, {linchenhao, chaoshen}@xjtu.edu.cn, yangminghui@oppo.com

Abstract

Owing to the boom of smartphone industries, the expansion of phone users has also been significant. Besides adults, children and elders have also begun to join the population of daily smartphone users. Such an expansion indeed facilitates the further exploration of the versatility and flexibility of digitization. However, these new users may also be susceptible to issues such as addiction, fraud, and insufficient accessibility. To fully utilize the capability of mobile devices without breaching personal privacy, we build the first corpus for age group recognition on smartphones with more than 1,445,087 unrestricted actions from 2,100 subjects. Then a series of heuristically-selected and neurally-guided features are proposed to increase the separability of the above dataset. Finally, we develop *AgeCare*, the first implicit and continuous system incorporated with bottom-to-top functionality without any restriction on user-phone interaction scenarios, for accurate age group recognition and age-tailored assistance on smartphones. Our system performs impressively well on this dataset and significantly surpasses the state-of-the-art methods.

1 Introduction

With the comprehensive development of the functionality of smartphones, the usage of these devices significantly increases, and the span of smartphone users' ages has also been drastically expanded in the recent decade. People other than adults have become a significant part of smartphone users. With the increasing base of mobile phone users, the versatility and flexibility of digitization can be further explored, pushing the boundary of smartphone-based services further toward covering all ages. It is predictable that many fields, such as the medical and educational industries, would profoundly benefit from such an expansion.

Despite the benefits given by smartphones, the larger span of user ages also leads to more potential hazards that users face when using their devices. For the younger age group (under 18)[WIKI, 2023], smartphone addiction is becoming prevalent as more and more teenagers spend a considerable amount of time on social media, mobile games, and other applications. While for the older age group (above 59)[WHO, 2022], fraud and insufficient accessibility have become dominating issues. Therefore, there is an urgent demand for the supervision of users from special age groups to provide tailored protection and guidance against the aforementioned issues.

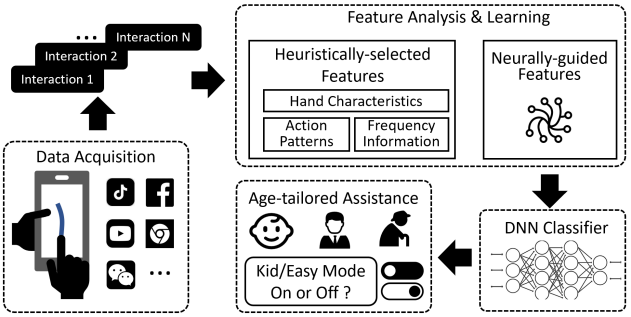

Figure 1: Overview of our AgeCare system. Non-sensitive interaction data are firstly acquired through user devices. Then, **HSF**s and **NGF**s are extracted by corresponding modules, which are then fed into the DNN-based classifier. An age-tailored assistance service will be activated automatically according to the prediction results.

However, current solutions mainly fall into domains such as government policy and guardians' responsibility [Nyamadi *et al.*, 2020]. For the elder group, despite the accessibility functions on most cell phones, some old people are unable to access these functions and have to turn to their siblings for help because of their unfamiliarity with smartphone operations. In other words, the capability of mobile phones as a smart device itself to detect user ages is relatively neglected and unexplored, not to mention the difficulty in determining users' ages without breaching privacy also obstacles an efficient protective deployment for special age groups.

To resolve these issues, we develop an implicit and continuous system called *AgeCare* to provide age-tailored assistance under unconstrained interaction scenarios (*i.e.*, multi-Apps, multi-postures, etc.) through age group recognition. However, there are several challenging obstacles to building such a system. Firstly, to avoid ethical issues in privacy, none personal or sensitive information (including the App-related information from which the current action comes, etc.) can be accessed. Thus, the available data should only be action-

related and sensor-related. Secondly, given the first constraint, the difference among age groups w.r.t. actions and sensors could be subtle. In other words, improperly extracting action and sensor features could lead to an extremely entangled and hard-to-separate data distribution. To address all these challenges, we first build a corpus containing only action data and sensor data from smartphones without breaching individual privacy. Subsequently, by closely investigating and understanding the common sense about the intrinsic difference among all-age users, we heuristically propose *hand characteristics*, *action patterns*, and *frequency information* as our heuristically selected features (HSF) to increase data separability. Furthermore, we take advantage of the expressive power of LSTM and couple HSF with neurally-guided features (NGF), efficiently capturing human-imperceptible but model-friendly characteristics. Finally, we propose a DNN architecture consisting of three modules to analyze heuristically-selected features, extract neurally-guided features, and fuse them for discriminative feature learning and accurate age group recognition.

In summary, our main contributions are given as follows:

- We develop *AgeCare*, the first implicit and continuous system incorporated with bottom-to-top functionality without any restriction on user-phone interaction scenarios, for accurate age group recognition and age-tailored assistance on smartphones.
- We propose a series of heuristically-selected features to integrate *action patterns* and *frequency information* and a number of neurally-guided features to capture human-imperceptible and model-friendly features through the help of LSTM layers, all of which boost model performance in the age group recognition task by a considerable amount.
- We build the first and the largest corpus from unrestricted scenarios for age group recognition on smartphones. It consists of 1,445,087 operations from 2,100 subjects when using more than 50 popular apps on multiple Android smartphones. Our approach evaluated on this challenging dataset achieves an impressive AUC value of 0.91 and F1 score of 0.778 and significantly surpasses the current SOTA methods.

2 Related Works

2.1 Age Group Recognition

Age recognition and age-related identification have recently raised attention in the research field as an increasing number of younger and older users have witnessed the prosperity of digitization. Many age-related issues, including addiction and fraud, have longed for an age group recognition system to provide further solutions. Early studies relying primarily on facial and voice information [Basaran *et al.*, 2014; Savchenko, 2019] often required additional input devices such as cameras or microphones, which may lead to leaking privacy. Anthony et al.[Anthony *et al.*, 2012] research showed large differences between children's and adults' touch-gesture interactions on mobile devices, proving that gesture interactions can be used to distinguish between children and adults. Without compromising privacy, Vatavu et al. [Vatavu *et al.*, 2015a] used touch coordinates to identify the age group of users, especially for children aged between 3 to 6. Nguyen et al. [Nguyen *et al.*, 2019] built a dataset containing 50 users' data from smartphone sensors to more precisely capture action characteristics and then conducted age group recognition using traditional machine learning methods (SVM and RF) on this dataset. Cheng et al. [Cheng *et al.*, 2020] proposed three hand-related characteristics: hand geometry, finger dexterity, and hand stability. The authors then extracted 53-dimensional features on action data from four specific tasks and trained conventional machine learning models to recognize young users.

However, existing works on age group recognition focus mainly on only the children group, with limited or flawed approaches ranging from using sensitive data (faces, voices) to constrained interaction (specifically designed Apps). In addition, the adopted database in their experiments is not large or diverse enough for algorithm validation in realistic scenarios.

2.2 User Authentication through Deep Learning on Smartphones

Recent studies turned to take advantage of the capability of deep neural networks (DNN) to seek reliable user authentication on smartphones [Li *et al.*, 2021; Deng *et al.*, 2021]. DNN models have been proven to be capable of capturing more complicated features under more diverse interacting scenarios and have been used for authentication [Parkhi *et al.*, 2015; Sundararajan and Woodard, 2018]. Hu et al. [Hu *et al.*, 2018] introduced a two-stream CNN for continuous user authentication by using accelerometer and gyroscope sensors on smartphones. Li et al. [Li *et al.*, 2020a] introduced frequency and temporal difference domain information into training features and presented a sensor-based continuous authentication system using a two-stream CNN. And it is one of the first studies to utilize two feature fusion strategies to combine the designed features, using the data from three sensors[Li *et al.*, 2020b]. In [Abuhamad *et al.*, 2020], the authors developed a DNN model using both touchscreen-based and sensor-based features for user authentication. [Lin *et al.*, 2022] designed a novel temporal-aware learning mechanism with DNN for cross-scenario identity authentication.

More recently, the authors [Stragapede *et al.*, 2023] first built a behavioral human-computer interaction database called BehavePassDB under 8 different tasks, and then trained an LSTM model to realize continuous authentication. However, BehavePasssDB contains much privacy-sensitive information and suffers from insufficient tasks to cover realistic smartphone-using scenarios.

Although deep learning (DL) has demonstrated its capability in user authentication, few studies attempt to explore DL deployment in age group recognition on smartphones, not to mention the scarcity of work that takes elderly people into account. Our system becomes the first DL-based practice that recognizes the user ages of children, adults, and elders under unrestricted human-smartphone interaction scenarios.

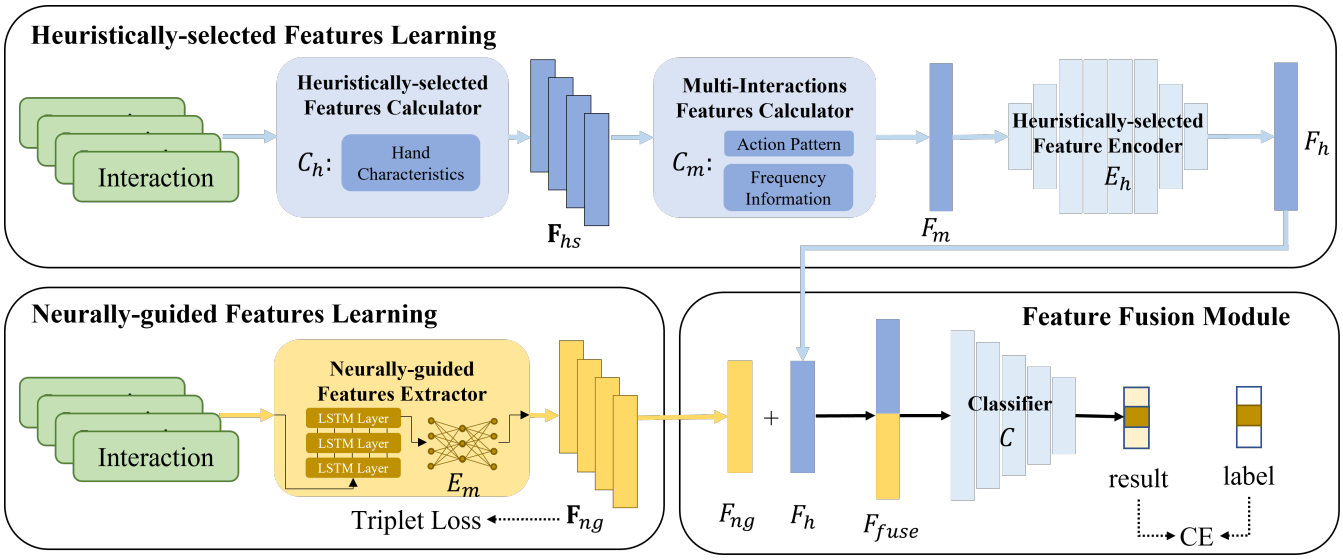

Figure 2: Network Architecture of AgeCare. It consists of three modules: (1) A heuristically-selected features learning module to extract hand characteristics, action patterns, and frequency information; (2) A neurally-guided features learning module to extract NGFs; (3) A feature fusion module to fuse HSFs and NGFs to be fed into a DNN-based classifier for training.

3 Methodology

3.1 Overview

Our system AgeCare, as shown in Figure 1, is the first system incorporated with bottom-to-top functionality, including data acquisition and pre-processing, feature extraction, age detection, and age-tailored assistance. The main approach proposed in our system consists of three consecutive phases: a data acquisition method responsible for collecting and processing data for subsequent feature extraction; a series of heuristically-selected and neurally-guided features that help models to capture key behavioral information when using smartphones; a specifically designed network to learn from user behavior and discriminate the age group of the user accordingly, as shown in Figure 2.

3.2 Data Acquisition and Pre-processing

The data acquisition is subject to the consent of all the subjects, and children's data are collected under the authorization and supervision of their guardians. The whole process strictly follows the guidelines of the privacy act in our country, and the study has been approved by our institution's ethics boards.

We develop a generalized built-in SDK in Android smartphones to achieve real-time and non-conscious interaction data(*i.e.*, action data and sensor data) acquisition. More than 50 popular Apps of various types are pre-installed in 14 different models of smartphones for data acquisition.During collection, subjects are first asked to input their age group, user ID, gender, and posture as user attribute data. This attribute collection process is only executed during the training phase.

Then, they are allowed to use any App and operate smartphones normally for more than ten minutes without any constraints. Meanwhile, our SDK will record the device attributes, including the device brand, model, and screen resolution. And the interaction data, including multi-finger touchscreen action data and sensor data, will also be collected automatically. The touchscreen action data mainly includes timestamp, finger contact area size, X\Y coordinates, action type, and finger number. The sensor data contains the information from the three-axis accelerometer, gyroscope, gravimeter, linear acceleration, orientation, rotation, and magnetometer data at each timestamp. In addition, the subjects are asked to operate the smartphones in four different postures, including putting smartphones on the table, holding smartphones while sitting with/without propping arms on the table, and holding smartphones while standing. Finally, our database includes 2,100 subjects and 1,445,087 operations, with a nearly balanced distribution w.r.t ages, genders, and body postures. The summary of our database is shown in Tab.1. The whole collection process lasts for about three months.

Since the data is collected without any restriction, the existence of abnormal data is inevitable. In order to obtain clean data, it is necessary to eliminate invalid data in the original data after data acquisition, such as missing key fields. Then, we convert the multi-finger action data into multiple single-finger action data because the multi-finger action data are mixed together at the same timestamp, and the data length in each timestamp is different, making it not difficult for further processing and feature extraction. Finally, the single-finger action vector sequence is obtained, in which each unit represents a complete sliding action.

3.3 Feature Extraction and Analysis

This section introduces the heuristically-selected features (HSF) that demonstrate distinctiveness and thus provide separability among different age groups and neurally-guided features (NGF), which can be human-imperceptible and be captured by deep learning models.

Database	Subjects	Operations	Age Group	Multi Finger	Multi Posture	Gender	Scenario	Devices Used
AgeCare	2,100	1,445,087	3-8,9-14,15-17,18-59,60-64,65-69,70+	yes(3)	yes(4)	M/F	unrestricted	14
[Cheng *et al.*, 2020]	100	31,285	3-17,18-59	no	no	M/F	extra specific tasks	4
[Vatavu *et al.*, 2015b]	119	4,069	3-6,18+	yes(2)	no	M/F	extra specific tasks	2
[Anthony *et al.*, 2014]	74	10,300	6-17,18-33	no	no	UNK	extra acquisition task& gesture interaction task	1
[Syed *et al.*, 2019]	31	19,373	14-38	no	no	UNK	extra matching game	4
[Nguyen *et al.*, 2019]	50	14,383	3-12,24-66	no	no	UNK	3 specific tasks	1

Table 1: The summary of different databases. Our database collects the multi-finger (3 fingers) interaction data from 7 age groups, in 4 body postures, on 14 devices, and in completely unrestricted scenarios.

Heuristically-selected Features and Analysis

Intuitively, users across different ages exhibit a varying pattern of using smartphones. To capture these behavioral variances among age groups, we heuristically craft a 231-dimension feature based on our investigation and common knowledge called Heuristically-Selected Features (HSF). HSF can be categorized into hand characteristics, action patterns, and frequency information. The explicit explanations for how these features are crafted and how they can capture the differences among all age groups are as follows.

Hand characteristics. Because of the different developmental stages in hands among different age groups, current work [Cheng *et al.*, 2020] characterizes hand-related features mainly as hand stability, finger flexibility, and palm structure. As these features are intuitively crucial for depicting the innate variance of children, adults, and elders, we keep these hand characteristics as our benchmark features. Hand stability depicts the capability of a user to keep hands steady and thus cause subtle output changes in sensors. We believe that teenagers and elderly people have weaker hand stability than adults. Hence, pitch, yaw, and roll calculated using data from the accelerometer are extracted. Finger dexterity demonstrates how fast and dexterous a user's action is when using an APP. Since adults are believed to exhibit higher flexibility while aged people are prone to slower actions, we thus capture such features using the velocity and acceleration of fingers. Palm structure is defined as the geometry property of users' hands, including hand sizes, finger length, palm width, etc. Such features are incorporated because the growth of human hands is stage-wise and significantly different across different ages. Consequently, we use features such as swiping distance and the largest deviation point [Nguyen *et al.*, 2019] to capture these characteristics.

Action patterns. Due to innate physiological differences, users from different age groups may exhibit various habits and patterns when using smartphones. This could result from the difference in a combination of reading speed, cognitive ability, and reaction speed among children, adults, and elders. Besides, the variance in the choice of Apps also accounts for different patterns across age groups. Action patterns, including standard deviation of hand characteristics, action duty ratio, and action area distribution, are proposed to characterize the aforementioned differences. For example, the action duty ratio is defined by the percentage of an action duration compared with its complete action period (*i.e.*, its duration plus its time interval between the last action). A larger action duty ratio indicates a more frequent and denser action pattern. Given the action duration T_d, action interval T_i, action coordinates on screen X, Y, we now give the expressions of four representative features in action patterns. The duty ratio dr, children likelihood p_c, elder likelihood p_e, and action area distribution I_{act} can be calculated as follows:

$$dr = \frac{T_d}{T_d + T_i} \times 100\% \tag{1}$$

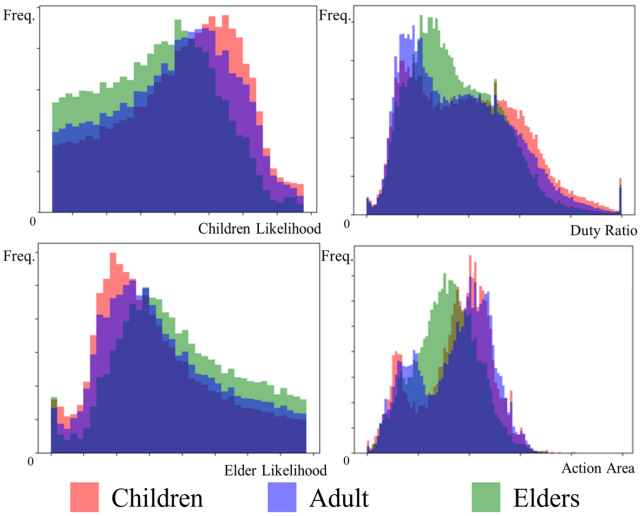

Figure 3: Visualization of data distribution and separability using histograms for action patterns features: **children likelihood**(upper left), **elder likelihood**(lower left), **duty ratio**(upper right) and **action area distribution**(lower right), respectively.

$$p_c = e^{-\frac{T_i - \beta_c}{1000}} \tag{2}$$

$$p_e = e^{-\frac{\beta_e - T_i}{1000}} \tag{3}$$

$$I_{act} = \frac{X}{w} \times (\frac{Y}{h} + 1) \tag{4}$$

where β_c, β_e, w, h are pre-determined constants. Note that all time-related data are given in milliseconds (ms). A visualization of feature separability is given in Figure 3. Details about other features in action patterns can be found in the supplementary document.

Frequency information. Inspired by the method in [Amini *et al.*, 2018], which uses frequency information of mobile phone sensors for authentication, we further extract frequency domain features from multiple sensors and integrate them into heuristically selected features. Action data in the frequency information are not considered as many of the features in action patterns have already integrated such information (e.g., duty ratio). Specifically, we use the time-domain data from sensors such as the accelerometer, gyroscope, magnetometer, etc. Data are recorded as per a complete window of the user's action w.r.t. the X, Y, and Z axis. To reduce the computational cost, Fast Fourier Transform is used to extract high-frequency information as frequency information. Given an action sequence(*i.e.*, from the start to the end of the action) in the X-axis of length N, the n-th data of this sequence x_n, the corresponding FFT of the total sequence X_k is:

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{-j2\pi kn}{N}} \qquad k = 0, 1, 2..., N-1 \tag{5}$$

where k stands for the frequency bin in the frequency domain. Note that when $k = N/2$, $X_{N/2} = \sum_{n=0}^{N-1} x_n e^{-j\pi n}$ is chosen as the high-frequency feature of the sensor sequence.

Neurally-guided Features and Analysis

All the above features related to hand characteristics and action patterns are extracted from user interaction behavior data in an interpretable way based on human intuition and experience. Although these features have the ability to describe user interaction, there are still implicit features of interactive behavior that cannot be captured by heuristically selected features. Therefore, we design a sequential model of the interaction as an additional interaction feature extractor to obtain the deep neural embedding from the raw data of user interaction.

The user's i-th interaction can be represented by a sequence b_i, in which each sampling frame is a time step. The data of the j-th time step b_{ij} consists of the X coordinate, Y coordinate, sliding speed, sliding acceleration of the user's touch screen interaction, as well as the sampling values of all 21 mobile phone sensors. These interaction sequences are directly fed into the neurally-guided feature extractor, which is composed of a sequential feature extractor and an encoder. The output of the neurally-guided feature extractor will be used as the embedding of the interaction data. Meanwhile, the triplet loss is used to optimize the network and make the difference of interaction embedding of different age groups as large as possible. Specifically, the feature extractor consists of 3 LSTM layers stacked together, and the encoder is an MLP with 5 full connection layers.

3.4 Network Architecture

As shown in Figure 2, our network is composed of three primary modules, the HSFs learning module, the NGFs learning module, and the feature fusion modules. The heuristically selected feature learning consists of a heuristically selected features calculator C_h, a multi-interactions features calculator C_m, and an HSF encoder E_h. This module takes several consecutive interactions as input and outputs the HSFs embedding of these interactions. The NGFs learning containing extractor E_m is used for NGFs extraction, which takes the same interactions as input, and outputs the neurally-guided embedding. The feature fusion module is a classifier composed of MLP, which takes the two types of features as input and outputs the prediction results of the model.

Heuristically-selected Features Learning

The HSFs learning module is designed to get multi-interactions embedding code F_h from several consecutive interactions. As described in Section 3.3, we first extract the hand characteristics $F_{hc} \in \mathbb{R}^{46}$ of each interaction through C_h. Since there is also a temporal relationship between every two consecutive interactions, to obtain more information conducive to the model prediction results from the interactions, we extract the features of multiple interactions F_m in addition to the hand characteristics through C_m. Specifically, we calculate several statistical characteristics for each feature component of the consecutive interactions, such as the features' mean, standard deviation, maximum, and minimum. In addition, action pattern features, such as action duty cycle, action coverage area, and frequency information, are also calculated. Finally, we obtain the multiple interactions feature $F_m \in \mathbb{R}^{231}$. After that, E_h takes the HSFs F_m as input and outputs the multi-interactions embedding code $F_h \in \mathbb{R}^{256}$. The HSF Encoder is an MLP with 13 full-connection layers.

Neurally-guided Features Learning

As described in Section 3.3, we also design an NGF extractor E_m to obtain an additional interaction embedding code F_{mf} based on the characteristics that user interaction data is a time series. To obtain the correlation between every two consecutive interactions, we adopt the same strategy of statistical features calculation to get the NGFs. Specifically, after processing interaction data by sequential model, we can obtain an embedding feature $F_{mf,i} \in \mathbb{R}^{64}$ for the i-th interaction. Then we calculate each feature component's mean, standard deviation, and maximum and obtain the multi-interactions embedding code $F_{mf} \in \mathbb{R}^{192}$. The experimental results in Section 4.4 demonstrate that the strategy of multi-interactions statistical feature calculation is necessary and can significantly improve the model performance.

Feature Fusion Module

In the feature fusion module, we directly concatenate the heuristically selected features and NGFs and feed them into the classifier, which is an MLP with 5 full connection layers.

Loss Function

Our model adopts a two-stage training strategy and imposes two losses, cross-entropy loss and triplet loss. Cross entropy loss is a common loss in classification tasks. Due to the large intra-type differentiation in the unconstrained scene, we adopt the triplet loss to minimize the distance between neurally-guided embedding feature F_{ng} of the same age group and expand that of different age groups. The loss function is as follows:

$$L = CE(\hat{l}, l) + L_{triplet}(f_{ng}) \tag{6}$$

4 Experimental Results

This section details the experimental setups, including the dataset overview, evaluation metrics, and implementation details. We then evaluate the proposed AgeCare system by comparing it with the SOTA systems/approaches. In addition, we examine the impact of HSFs, NGFs, multi-interaction strategy, and the data augmentation strategy.

4.1 Datasets and Evaluation Metrics

The self-collected data consisting of 2,100 subjects are used in our experiments. Table 2 illustrates the details of our dataset. As given in the table, our database includes 700 adults, 700 children, and 700 elderly. In addition, the corresponding distribution of subjects w.r.t. four different body postures (described in Sec.3.2) is 514, 511, 536, and 539. The number of male and female subjects is 989 and 1111, respectively. After data preprocessing, the interaction data of each subject will be used for feature extraction and model training. The split ratio of the training set and test set is 7:3.

To evaluate the performance of our approach, following previous works [Cheng *et al.*, 2020; Nguyen *et al.*, 2019; Abuhamad *et al.*, 2020] in behavioral biometric authentication and age group detection, we use Macro-F1, Macro-AUC (Area Under Curve), and Macro-EER (Equal Error Rate) as our evaluation metrics, which are the most commonly used metrics in multi-category classification tasks.

4.2 Implementation Details

We developed our data acquisition SDK by JAVA, which is applicable to various Android-based smartphones. For DNN training, the margin in the triplet loss is 0.1, and the weights of CE loss and triplet loss are 0.1 and 1, respectively. Our algorithm was implemented in the Pytorch deep learning framework [Paszke *et al.*, 2017]. All experiments were run with batch size 16 on a Ubuntu 20.04 with NVIDIA RTX 2080Ti GPU. We used the Adam [Kingma and Ba, 2014] optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$. The learning rates for the two modules were 10^{-3} and 10^{-4}, respectively. StepLR is used to adjust the learning rate. For every 20 epochs, the learning rate decreases to 10%.

4.3 Comparison with State-of-the-arts

To evaluate the proposed approach, we conducted experiments by comparing our method with five state-of-the-art methods for age identification, including the random forest (RF), support vector machine (SVM), and ET classifiers in iCare [Cheng *et al.*, 2020] and the RF and SVM classifiers in [Nguyen *et al.*, 2019]. To achieve the best performance of previous methods for comparison, we tuned the parameters of ET, RF, and SVM classifiers. In addition, we trained heuristically-selected Feature model (HSFE) and neurally-guided Feature model (NGFE), respectively, for comparison. The results are illustrated in Table 3 and Figure 4(a). It can be seen that we can achieve better performance than existing methods by only using the HSF or NGF. Our approach significantly outperforms state-of-the-art methods and can achieve a Macro-F1 of 0.778, a Macro-AUC of 0.91, and a Macro-EER of 0.164. From the results, we can see that the features learned by conventional machine learning methods are broadly separable but not discriminative enough for reliable age group recognition. We attribute the performance improvement to the specifically designed HSF, the NGF extractor, and the multi-interaction strategy to obtain the correlation between continuous interaction.

Age Group	3-7	9-14	15-17	18-59	60-65	66-70	70+	Total
No.	300	300	100	700	100	387	213	2,100

Table 2: Dataset Details.

Method	F1	AUC	EER
RF [Cheng *et al.*, 2020]	0.579	0.79	0.289
ET [Cheng *et al.*, 2020]	0.598	0.79	0.301
SVM [Cheng *et al.*, 2020]	0.500	0.75	0.351
RF [Nguyen *et al.*, 2019]	0.653	0.82	0.259
SVM [Nguyen *et al.*, 2019]	0.543	0.75	0.328
AgeCare(HSFE)	0.761	0.90	0.174
AgeCare(NGFE)	0.731	0.88	0.197
AgeCare	**0.778**	**0.91**	**0.164**

Table 3: Comparison with State-of-the-art Methods.

4.4 Influence of Postures and Genders

To validate the influence of postures and genders, we firstly conducted experiments on different body postures, *i.e.*, putting smartphones on the table (On the Table), holding smartphones while sitting with/without propping arms on the table (Sitting supported/unsupported), and holding smartphones while standing (Standing). The identification performance of different genders was also evaluated.

As shown in Table 4 and Figure 4(b), the identification performance for unsupported postures, *i.e.*, unsupported sitting and standing, is much better than for supported postures, *i.e.* supported sitting and on the table. It indicates that sensor-based features can efficiently capture the discriminative information in the sensor data among different age groups, particularly in unsupported scenarios. We also observe a slightly better performance in female subjects than male, which may be due to the unbalanced gender ratio in our database.

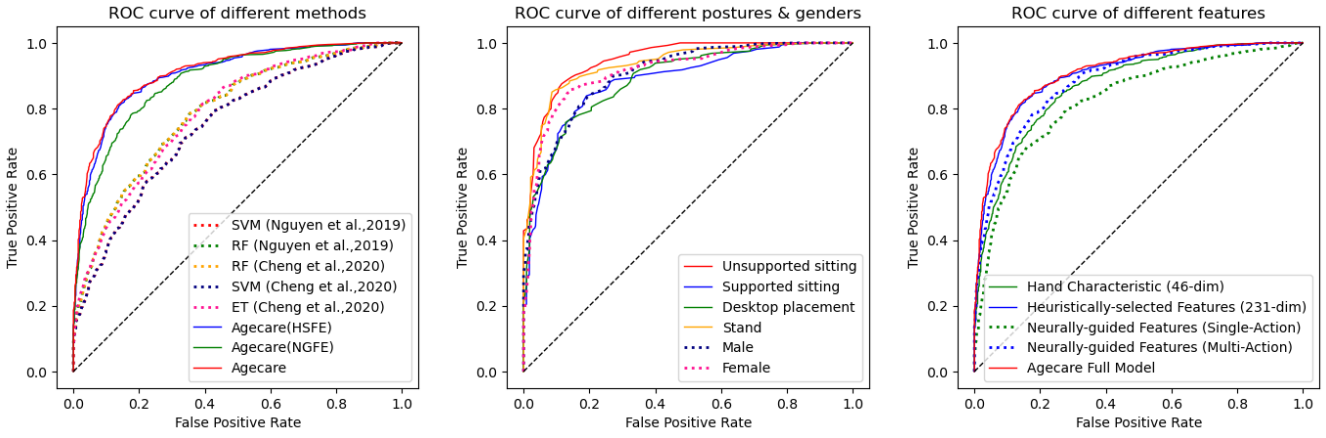

Figure 4: ROC curves illustrate (a) Comparative results with different methods. (b) Comparative results of Agecare under different external conditions. (c) Comparative results using different features.

Posture	Gender	F1	AUC	ERR
On the Table	-	0.735	0.89	0.203
Sitting Supported		0.732	0.88	0.192
Sitting Unsupported		0.817	**0.95**	0.139
Standing		**0.843**	0.93	**0.127**
-	Male	0.749	0.91	0.183
	Female	0.798	0.91	0.155

Table 4: Performance of AgeCare w.r.t. postures and genders.

Module	F1	AUC	EER
HC (46)	0.699	0.87	0.209
All HSF (231)	0.761	0.90	0.174
NGF (single)	0.672	0.83	0.247
NGF (multiple)	0.731	0.88	0.197
AgeCare	**0.778**	**0.91**	**0.164**

Table 5: Ablation experiments of the proposed features.

4.5 Ablation Experiments

We performed ablation experiments to illustrate the effectiveness of the HSFs and NGFs.

HSF and NGF extractor. We conducted a series of experiments, including using hand characteristics only, using full heuristically selected features (*i.e.*, plus action patterns and frequency information), as well as using full HSF and NGF. For our newly proposed heuristically-selected features (HSF), experimental results given in Table 5 and Figure 4 show that features combined with action patterns and frequency information (All HSF in the table) significantly outperform using only hand characteristics (HC in the table) by a notable margin w.r.t. all evaluation metrics (0.06 for F1, 0.03 for AUC and 0.33 for EER, respectively). This result suggests the increased separability incorporated by the proposed feature, demonstrating the superiority of our newly proposed HSF. As for the effectiveness of the NGF extractor, Table 5 also indicates that the overall performance of the model can be further boosted after combining HSFs with the NGF extractor. This suggests that our proposed feature extractor can efficiently capture human-imperceptible features and serve as an efficient complement to HSFs.

Multi-interaction strategy. As described in Section 3.4, in order to obtain the correlation between every two consecutive interactions, we adopt the same strategy of statistical feature calculation to get the model- favored features. We thus trained the network using single interaction features and multiple interactions features, respectively. As shown in Table 5, the model trained with NGFs of multiple interactions performs considerably better than the model trained with a single interactions. These results suggest that consecutive actions indeed contain more information than a single interaction.

5 Conclusion, Implications and Future Works

This paper proposes the AgeCare system and establishes the first user interaction database for age group recognition under unconstrained conditions. We also propose two feature extractors to expand extant hand characteristics into *heuristically-selected* features using action patterns and frequency information and extract novel *neurally-guided* features as complementary features of the former features. Experimental results show that with a combination of HSFs and NGFs, our AgeCare significantly outperforms state-of-the-art approaches in age group recognition.

Our system indicates the potential to identify the age group of smartphone users using unrestricted interaction data without compromising privacy-sensitive information. However, to mitigate the gap between the requirements for commercial applications among millions of smartphone users and the current performance and scalability of our model, further exploration of our work calls for tailored features and more streamlined models.

Ethics Statement

The data collection process in this study conforms to ethical rules, and the study was approved by our institution's and partner organizations' ethics boards. The data collection is subject to the consent of all the subjects, does not involve other personal private information, and all user data are anonymous. To be specific, we cooperated with several kindergartens, primary schools, middle schools, universities, and old people's homes to complete the data collection. Before the data collection, we informed all the children subjects, their guardians (parents and teachers), and other subjects of the relevant collection contents. The collected data will be treated anonymously and used only for this experiment. In the process of collection, children are collected under the authorization and supervision of their guardians. The children's guardians and all the other subjects have authorized all the collection processes, and the awareness and authorization documents have been signed.

Considering the potential ethical concerns associated with predicting user age, we have made many efforts to ensure that user privacy is not violated during actual deployment of AgeCare. Firstly, all aspects of AgeCare are performed locally, without involving any data transmission or cloud services. Both interaction data and age recognition results are kept private and inaccessible to any third-party application. Secondly, AgeCare will be integrated as a built-in feature within the OS of smartphones, requiring users to activate it manually. Meanwhile, the predicted results are only used for child addiction prevention and tailored care service. Finally, as to the normal use of adults, we may accept a lower recall for the target groups to minimize false classification of the adult group and prevent any negative impact on their normal usage. If there are still false positives, other conventional authentication methods, such as passwords, can be used to remove restrictions.

Contribution Statement

Yingmao Miao and Qiwei Tian contributed equally to this work. Chenhao Lin and Chao Shen are corresponding authors.

Acknowledgements

This research is supported by the National Key Research and Development Program of China (2020AAA0107702), the National Natural Science Foundation of China (62006181, 62161160337, 62132011, U21B2018, U20A20177, 62206217), the Shaanxi Province Key Industry Innovation Program (2023-ZDLGY-38, 2021ZD LGY01-02).

References

[Abuhamad *et al.*, 2020] Mohammed Abuhamad, Tamer Abuhmed, David Mohaisen, and DaeHun Nyang. Autosen: Deep-learning-based implicit continuous authentication using smartphone sensors. *IEEE Internet of Things Journal*, 7(6):5008–5020, 2020.

[Amini *et al.*, 2018] Sara Amini, Vahid Noroozi, Amit Pande, Satyajit Gupte, Philip S Yu, and Chris Kanich. Deepauth: A framework for continuous user reauthentication in mobile apps. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 2027–2035, 2018.

[Anthony *et al.*, 2012] Lisa Anthony, Quincy Brown, Jaye Nias, Berthel Tate, and Shreya Mohan. Interaction and recognition challenges in interpreting children's touch and gesture input on mobile devices. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, pages 225–234, 2012.

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

[Basaran *et al.*, 2014] Can Basaran, Hee Jung Yoon, Ho Kyung Ra, Sang Hyuk Son, Taejoon Park, and Jeong-Gil Ko. 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*, pages 343–347, 2014.

[Cheng *et al.*, 2020] Yushi Cheng, Xiaoyu Ji, Xiaopeng Li, Tianchen Zhang, Sharaf Malebary, Xianshan Qu, and Wenyuan Xu. Identifying child users via touchscreen interactions. *ACM Transactions on Sensor Networks (TOSN)*, 16(4):1–25, 2020.

[Deng *et al.*, 2021] Shaojiang Deng, Jiaxing Luo, and Yantao Li. Cnn-based continuous authentication on smartphones with auto augmentation search. In *International Conference on Information and Communications Security*, pages 169–186. Springer, 2021.

[Hu *et al.*, 2018] Hailong Hu, Yantao Li, Zhangqian Zhu, and Gang Zhou. Cnnauth: continuous authentication via two-stream convolutional neural networks. In *2018 IEEE international conference on networking, architecture and storage (NAS)*, pages 1–9. IEEE, 2018.

[Kingma and Ba, 2014] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Li *et al.*, 2020a] Yantao Li, Hailong Hu, Zhangqian Zhu, and Gang Zhou. Scanet: sensor-based continuous authentication with two-stream convolutional neural networks. *ACM Transactions on Sensor Networks (TOSN)*, 16(3):1–27, 2020.

[Li *et al.*, 2020b] Yantao Li, Bin Zou, Shaojiang Deng, and Gang Zhou. Using feature fusion strategies in continuous authentication on smartphones. *IEEE Internet Computing*, 24(2):49–56, 2020.

[Li *et al.*, 2021] Yantao Li, Peng Tao, Shaojiang Deng, and Gang Zhou. Deffusion: Cnn-based continuous authentication using deep feature fusion. *ACM Transactions on Sensor Networks (TOSN)*, 18(2):1–20, 2021.

[Lin *et al.*, 2022] Chenhao Lin, Jingyi He, Chao Shen, Qi Li, and Qian Wang. Crossbehaauth: Cross-scenario behavioral biometrics authentication using keystroke dynamics.

IEEE Transactions on Dependable and Secure Computing, pages 1–1, 2022.

[Nguyen *et al.*, 2019] Toan Nguyen, Aditi Roy, and Nasir Memon. Kid on the phone! toward automatic detection of children on mobile devices. *Computers & Security*, 84:334–348, 2019.

[Nyamadi *et al.*, 2020] Makafui Nyamadi, Richard Boateng, and Immaculate Asamenu. Smartphone addictions: A review of themes, theories and future research directions. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, pages 6093–6102, 2020.

[Parkhi *et al.*, 2015] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. *BMVC*, 2015.

[Paszke *et al.*, 2017] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. *NIPS 2017 Autodiff Workshop*, 2017.

[Savchenko, 2019] Andrey V Savchenko. Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output convnet. *PeerJ Computer Science*, 5:e197, 2019.

[Stragapede *et al.*, 2023] Giuseppe Stragapede, Ruben Vera-Rodriguez, Ruben Tolosana, and Aythami Morales. Behavepassdb: Public database for mobile behavioral biometrics and benchmark evaluation. *Pattern Recognition*, 134:109089, 2023.

[Sundararajan and Woodard, 2018] Kalaivani Sundararajan and Damon L Woodard. Deep learning for biometrics: A survey. *ACM Computing Surveys (CSUR)*, 51(3):1–34, 2018.

[Syed *et al.*, 2019] Zahid Syed, Jordan Helmick, Sean Banerjee, and Bojan Cukic. Touch gesture-based authentication on mobile devices: The effects of user posture, device size, configuration, and inter-session variability. *Journal of Systems and Software*, 149:158–173, 2019.

[Vatavu *et al.*, 2015a] Radu-Daniel Vatavu, Lisa Anthony, and Quincy Brown. Child or adult? inferring smartphone users' age group from touch measurements alone. In *IFIP Conference on Human-Computer Interaction*, pages 1–9. Springer, 2015.

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

[WHO, 2022] WHO. Ageing and health. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health, 2022. Accessed: 2023-05-29.

[WIKI, 2023] WIKI. United nations convention on the rights of the child. https://en.wikipedia.org/wiki/Child, 2023. Accessed: 2023-05-29.