# FeverPhone: Accessible Core-Body Temperature Sensing for Fever Monitoring Using Commodity Smartphones

JOSEPH BREDA, Paul G. Allen School of Computer Science & Engineering, University of Washington, USA
MASTAFA SPRINGSTON, Department of Emergency Medicine, University of Washington School of Medicine, USA
ALEX MARIAKAKIS, Department of Computer Science, University of Toronto, USA
SHWETAK PATEL, Paul G. Allen School of Computer Science & Engineering, University of Washington, USA



Smartphones contain thermistors that ordinarily monitor the temperature of the device's internal components; however, these sensors are also sensitive to warm entities in contact with the device, presenting opportunities for measuring human body temperature and detecting fevers. We present FeverPhone — a smartphone app that estimates a person's core body temperature by having the user place the capacitive touchscreen of the phone against their forehead. During the assessment, the phone logs the temperature sensed by a thermistor and the raw capacitance sensed by the touchscreen to capture features describing the rate of heat transfer from the body to the device. These features are then used in a machine learning model to infer the user's core body temperature. We validate FeverPhone through both a lab simulation with a skin-like controllable heat source and a clinical study with real patients. We found that FeverPhone's temperature estimates are comparable to commercial off-the-shelf peripheral and tympanic thermometers. In a clinical study with 37 participants, FeverPhone readings achieved a mean absolute error of 0.229 °C, a limit of agreement of ±0.731 °C, and a Pearson's correlation coefficient of 0.763. Using these results for fever classification results in a sensitivity of 0.813 and a specificity of 0.904.

CCS Concepts: • **Applied computing** → **Consumer health**; • **Human-centered computing** → **Smartphones**.

Additional Key Words and Phrases: temperature sensing, thermometer, fever detection, thermistor, mobile health

**ACM Reference Format:**
Joseph Breda, Mastafa Springston, Alex Mariakakis, and Shwetak Patel. 2023. FeverPhone: Accessible Core-Body Temperature Sensing for Fever Monitoring Using Commodity Smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1, Article 3 (March 2023), 23 pages. https://doi.org/10.1145/3580850

## 1 INTRODUCTION

Traditionally defined as a core body temperature of greater than 38 °C (100.4 °F), fever is known to be the primary predictive symptom for many viral infections and is the single most commonly cited manifestation of COVID-19 [10, 20, 22, 23, 30, 35]. In many cases, clinicians use a lower threshold of 37.5 °C (99.5 °F) [1, 22] or even 37.3 °C [19] as the cutoff for a low-grade fever, triggering the early detection of illness and longitudinal monitoring to defend against community spread. A host of thermometer models are available for at-home patient use, ranging from in-ear to skin-surface to oral form factors, but these devices are often not available in people's homes. A study by Parmar et al. [43] found that out of a cohort of 141 parents who required frequent temperature monitoring for their child with febrile convulsions, only 21 parents (15%) actually had access to a thermometer at

Authors' addresses: Joseph Breda, joebreda@cs.washington.edu, Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, USA; Mastafa Springston, Department of Emergency Medicine, University of Washington School of Medicine, Seattle, USA, mastafa@uw.edu; Alex Mariakakis, Department of Computer Science, University of Toronto, Seattle, USA, mariakakis@cs.toronto.edu; Shwetak Patel, Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, USA, shwetak@cs.washington.edu.


2474-9567/2023/3-ART3
https://doi.org/10.1145/3580850

home. The lacking ubiquity of thermometers may be due to some combination of their non-trivial cost ($15–$300 USD) or the fact that they are only typically needed a few times per year. Thermometers can also sell out during times of need, further reducing access [6].

More readily available thermometry with comparable error could have a substantial impact on personal health screening, identifying nascent cases of illness and enabling more informative telehealth consultations. Accessible fever monitoring could also facilitate pandemic mitigation and epidemiological modeling [7]. To achieve these capabilities, we propose a method of estimating core body temperature using commodity smartphones. Smartphones have become increasingly ubiquitous in recent years, with nearly 3 billion smartphone in 2019 [46]; many of these users reside in under-served areas due to their proximity to clinical resources or economic constraints. Previous work has explored the use of smartphone-connected thermometers to detect influenza-like illnesses at scale [39], hinting at the potential benefits that a purely smartphone-based solution could have.

In this paper, we leverage the thermistors embedded in smartphones to estimate a person's core body temperature. We do this by collecting thermistor and touchscreen readings over a 90-second interaction where the smartphone touchscreen is held against the user's forehead. We then define a data-driven model that maps features describing the rate of heat transfer between the user's forehead and the device to the ground-truth temperature collected by an oral thermometer. We deployed this technique in a smartphone app called FeverPhone, and we evaluated this app both in lab and in a clinical setting. First, we used a simulated heat source in the form of a temperature controlled bag of water to ensure the temperature of a source in contact with the device can be recovered from the thermistor signal. Beyond evaluating the core working principle of FeverPhone, we evaluated its robustness across multiple confounds: phone model, phone accessories, pressure of device application, and ambient temperature. We then deployed FeverPhone in a clinical trial with 37 patients, 16 of whom had a low-grade or standard fever. We found that FeverPhone achieves a mean absolute error of 0.229 °C, a limit of agreement of ±0.731 °C, and a Pearson's correlation coefficient of 0.763.

In summary, our primary contribution is a demonstration of how the thermistors in a smartphone can be used to estimate a person's core body temperature without any hardware modifications to the device itself. We demonstrate this capability through the following components:

- A smartphone application and user interaction capable of estimating core body temperature using just the built-in device hardware,
- A lab validation study using a simulated heat source to validate the working principle of FeverPhone and many potential confounds against its generalizability, and
- A preliminary clinical study with 37 patients in a local emergency room.

## 2 RELATED WORK

In this section, we first review literature on the use of mobile devices for monitoring influenza-like illnesses. We then enumerate the multitude of dedicated hardware form factors that people have used to assess core body temperature. Finally, we report on past efforts undertaken by researchers to leverage smartphone thermistors for temperature sensing.

### 2.1 Mobile Health Sensing for Influenza-Like Illness

Mobile devices and wearables are increasingly becoming useful of health monitoring, both at personal scale for preliminary health screening [36] and at population scale for tracking the spread of illness and informing epidemiological models [21]. With respect to influenza-like illness, these smart devices have often been used to record people's sleep, heart rate, and mobility as physiological indicators of symptoms. For example, Hirten et al. [25] found that heart rate variability differed for the seven days before and after a COVID-19 diagnosis with noticeable changes in the signal observed the first day of reported symptoms. Merrill and Althoff [38] leveraged

data on sleep, heart rate, and mobility from smartwatches and fitness trackers to predict cases of COVID-19. Finally, Venkatramanan et al. [48] used anonymous location history collected from smartphones to forecast the spread of the flu. Despite the utility of core body temperature as a signal for influenza-like illness and the existence of temperature sensors on-device, these works and others have yet to incorporate it into their mobile health sensing systems.

## 2.2 Body Temperature Sensing with Smart Devices

Consumer-grade peripheral thermometers come in multiple form factors that can measure temperature at different parts of the body. These locations include but are not limited to the ear canal, the mouth, the armpit, and the temporal artery. Many purpose-built at-home thermometers show high degree of error, demonstrating that thermometry is still a challenging task.

*2.2.1 Limitations of Consumer-grade Thermometers.* In a review of 75 different studies on the accuracy of peripheral thermometers, Niven et al. [41] found that the studied thermometers had pooled limits of agreement outside the predefined clinically acceptable range of ±0.5 °C, with errors being as large as ±1.46 °C for febrile adults. These findings have been corroborated by Farnell et al. [18] and Lawson et al. [32], among others. Similarly, Pecoraro et al. [44] performed a meta-analysis of consumer-grade infrared temporal thermometers for fever cut-off classification and found that of 9 different studies these devices anywhere from 0.41 to 0.91 sensitivity with 0.85 to 1.00 specificity. Another study by Lai et al. [31] found the AUC for non-contact infrared thermometers to be between 0.8 and 0.9 when screening at the forehead site, but also found this value to fall below 0.8 for colder ambient air temperatures (14 °C–16 °C). Nevertheless, consumer-grade peripheral thermometers are currently the standard for at-home temperature monitoring and are used for screening by many businesses.

*2.2.2 Emerging Alternative Thermometer Designs.* Researchers have explored alternative hardware designs to perform fever detection. For example, dedicated torso- and wrist-mounted sensors have been developed to capture the wearer's skin temperature [8, 26]; this data can later be used to infer core body temperature by accounting for basic information like the wearer's height and weight. Researchers have also explored the use of contactless thermal cameras for fever detection. Shinde et al. [45] and Wei et al. [50] leveraged forward-looking infrared (FLIR) cameras to identify fevers from infrared video data. The Huawei Honor Play 4 smartphone released in 2020 includes an infrared temperature sensor which can be used as a thermometer [2], but this sensor has not seen pervasive use in other smartphone models. On the whole, these approaches to temperature sensing require dedicated or specialized hardware that is not yet ubiquitous, limiting the ability for researchers and clinicians to acquire temperature data during telehealth scenarios for personal screening. Our work seeks to alleviate this issue by taking advantage of devices that are already widely adopted, namely smartphones, for this purpose.

## 2.3 Temperature Sensing Using Smartphone Thermistors

Smartphones are equipped with many temperature-sensitive resistors called thermistors for monitoring the integrity of the device's other internal components, namely the battery. This technology is the exact same hardware present in state-of-the-art clinical-grade thermometer probes. Peripheral thermometers that leverage thermistors often do not measure temperature directly but rather predict body temperature based on the change in temperature experienced by the thermistor when it comes into contact with the body. On top of this temporal prediction, proprietary algorithms are used to map peripheral body temperature to core body temperature. Although smartphone thermistors are designed to monitor the temperature of components in the device itself, they have been repurposed for two main applications, which are listed below.

*2.3.1 Ambient Air Temperature.* Past work has shown that it is possible to passively measure the ambient air temperature for the purposes of crowdsourcing temperature estimates within buildings or cities [42, 47].

Breda et al. [12] demonstrated ambient air temperature estimation by combining battery temperature data with information about the smartphone's software usage and physical context. This work leveraged a hierarchy of linear and exponential models to characterize the different thermal states of the device, and then it used activity recognition to select the appropriate model. He et al. [24] later developed a similar technique for ambient air temperature estimation, utilizing battery current usage for infer the state of the device. Both of these studies specifically modeled the heat generated by smartphone activity and decoupled this information from ambient air temperature. Finally, Chau [13] focused on opportunistically estimating ambient air temperature when the smartphone had minimal software utilization and, therefore, minimal heat being generated within itself. The authors go on to describe how the position of the device (e.g., within the pocket, in the user's hand) impacts the temperature estimates.

*2.3.2 Core Body Temperature.* The notion of using smartphones to estimate core body temperature has been explored to a limited degree in prior work. Breda and Patel [11] ran a simulated study using a sous-vide machine and exhibited the feasibility of using smartphone thermistors to discriminate high and low water temperatures after making contact with the simulated heat source for 3 minutes. Jun et al. [28] replicated this research with similar results after 8 minutes of contact between the device and simulated heat source. Neither of these studies evaluated the corresponding systems for measuring core body temperature in live human subjects.

FeverPhone represents the first example of a system that leverages smartphone thermistors for core body temperature estimation and fever monitoring in real people. Prior work related to ambient air temperature estimation has focused on passive sensing scenarios, requiring information about the software being run on the phone and how the phone is being operated by the user to apply corrections to the data. Prior work also entirely focuses on the use of the battery thermistor which is the only publicly accessible thermistor on Android. FeverPhone makes use of all the thermistors on the smartphone by leveraging Android phones with root-level access to read each thermistors system files. These thermistors were found to update at a higher frequency and perform better at estimating core body temperature than the battery thermistor. FeverPhone is not a passive system but rather relies on an active interaction to estimate core body temperature. This means that we are able to both control the software running on the device and the way the device is held during the interaction to improve accuracy.

## 3 SYSTEM DESIGN

Fig. 1 illustrates the overall design of FeverPhone. In short, we treat the smartphone and its built-in thermistors as a temperature probe. The user is asked to place the phone's touchscreen in direct contact with their forehead in order to control the data collection process in a repeatable and convenient manner. FeverPhone then analyzes not only the temperature signal measured by a single thermistor, but also the capacitive data from the device's touchscreen to capture features summarizing the type of contact made with the user's skin. In the subsections that follow, we describe this design decisions and algorithm behind FeverPhone in greater detail.

### 3.1 Working Principle

*3.1.1 Newton's Law of Cooling.* Newton's Law of Cooling states that that the rate of heat loss of an object is directly proportional to the difference between the temperatures of the object and its environment. This concept is mathematically expressed as follows:

$$T(t) = (T_0 - T_{env})e^{-kt} + T_{env} \tag{1}$$

where $T_0$ represents the initial temperature of the object, $T_{env}$ is the temperature of the environment, and $k$ is a constant that dictates the cooling rate of the object.

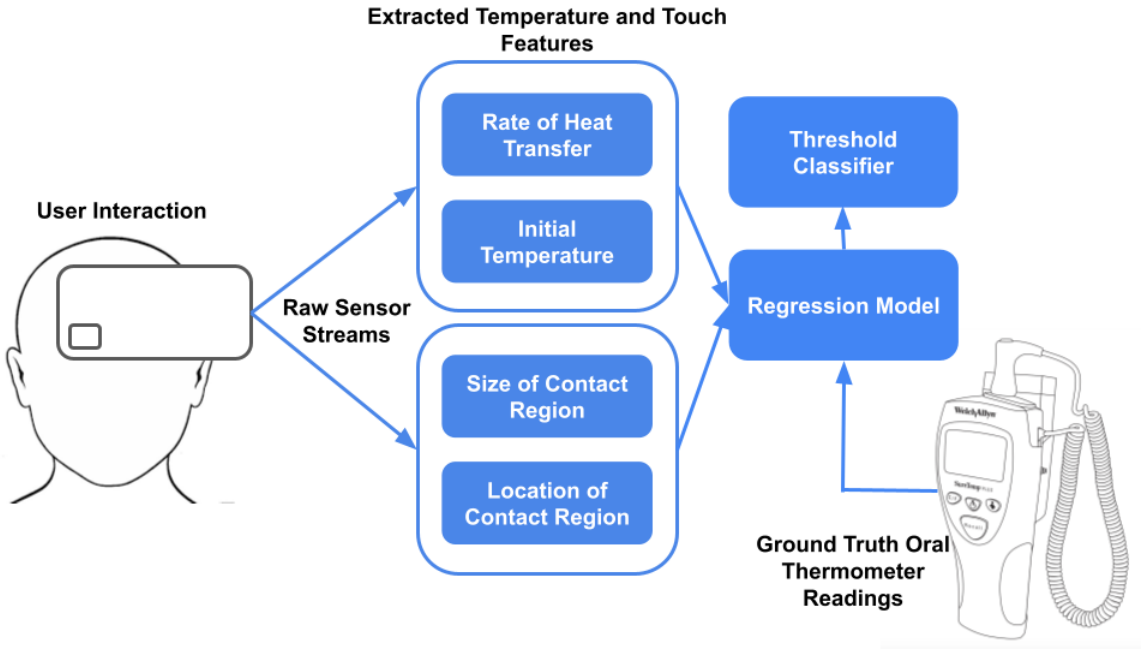


Fig. 1. A box diagram illustrating the algorithmic pipeline for FeverPhone. As the user places the smartphone's touchscreen against their head, FeverPhone records features related to the temperature change experienced by the thermistor and the contact surface as measured by the phone's touchscreen. These features are used in a regression model to predict core body temperature.

For the purposes of FeverPhone, $T_0$ represents the initial temperature measured by a given thermistor before contact is made between the device and the user's forehead. Since the thermal mass of the user is significantly greater than that of the smartphone, the temperature of the device should eventually reach $T_{env}$ — the steady-state temperature reached by the device after it comes to thermal equilibrium with the user's forehead. $k$ describes the rate of heat transfer from the smartphone to the user's forehead, which varies across devices and users. More specifically, device-specific factors include the total mass of the smartphone, the position of the thermistor within the device, and the device's software utilization. User-specific factors include the physical composition of the user and the way that they position the smartphone on their forehead. Although $k$ encompasses many factors, Equation 1 is still idealized as it does not consider the cooling of the smartphone that occurs separately from the heat transfer to the user's forehead (e.g., heat dissipation, heat transfer to the user's hand).

*3.1.2 Preconditions of Thermometry.* Much like existing clinical-grade thermometers, FeverPhone relies on heat being transferred from the user to the thermistor. If the temperature of the thermistor is too similar to the temperature of the user's forehead, there will not be enough heat transfer to make a proper estimate. Further, if the temperature of the thermistor is greater than that of the user, heat will instead be transferred from the device to the user. All of these factors are influenced by the ambient air temperature, which impacts both the resting temperature of the thermistor and the user's peripheral body temperature. Therefore, most thermometers and temperature sensing devices constrain the ambient air temperatures that are suitable for operation. This

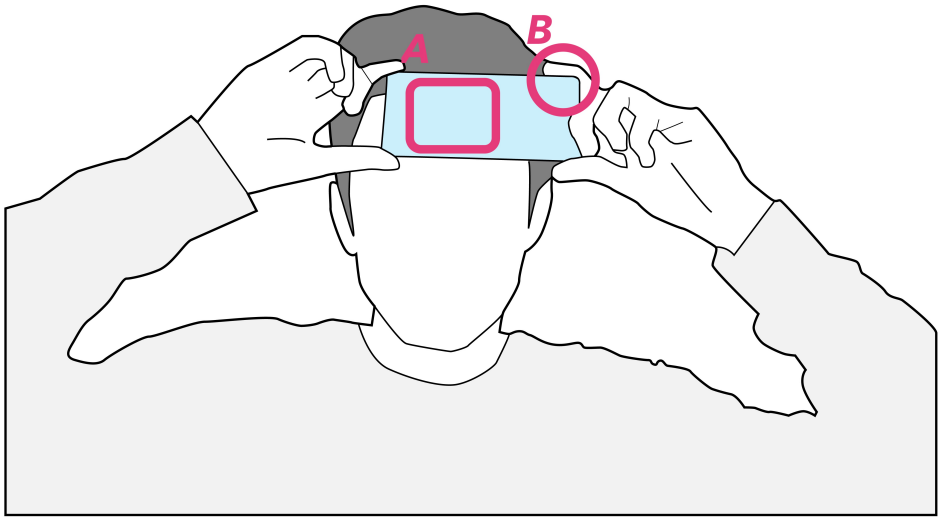


Fig. 2. Example posture of user interaction. Box A highlights the region of contact between the user's forehead and device screen closer to the top half of the device, and Box B highlights the "camera-grip" style where the four corners of the device are pinched between the user's fingers.

range is largely based on the thermistor's material[1], and using thermometers outside of these bounds can lead to inaccuracies [17, 37].

As with all other thermometers, FeverPhone requires that the user's body temperature is greater than the temperature of the device. We also require users to take a conservative break of 10 minutes between successive readings so that the thermal probe can reacclimate to its original temperature. We only test FeverPhone when the ambient air is between 18.3 °C and 23.6 °C, which is tighter than the operating ranges of clinical-grade thermometers like the Welch-Allyn SureTemp 690 thermometer[2] (10 °C–40 °C) and consumer-grade digital thermometers like those by Kerma Medical[3] (18 °C–28 °C). Establishing the operating range of FeverPhone and expanding it further would just be a matter of collecting data at more extreme temperatures since FeverPhone is data-driven, so we leave these questions to future work. In Section 4.3.4, we explore the impact of testing FeverPhone when the ambient air temperature is outside what is available during training.

## 3.2 Testing Procedure

*3.2.1 Testing Location and Smartphone Positioning.* There are various body locations that could be suitable for smartphone thermometry, namely the hands, ears, forehead, and armpits. These testing sites have unique advantages in terms of their ergonomics and affordances, but we have specifically designed FeverPhone with the forehead in mind for two reasons. First, the forehead is less susceptible to confounds like ambient air temperature [15, 49] and physiological thermoregulation [4, 14, 34]. For example, the hands can vary by as much as 15 °C when exposed to varying ambient air temperatures [15], whereas the forehead only varies by 3 °C across a similar range [49]. Second, the forehead has a sufficiently large and exposed surface to make contact with

[1] https://www.electronics-tutorials.ws/io/thermistors.html

[2] https://www.welchallyn.com/content/welchallyn/americas/en/products/categories/thermometry/oral-axillary-rectal-thermometers/suretemp-plus-690.html

[3] https://www.kermamedical.com/Products/Details/7

the device's screen. Although the armpit can result in even more heat transfer, the resulting signal can be less consistent due to variation in clothing and body curvature.

We define a standardized interaction to control the user's grip, the orientation of the device, and the region-of-contact between the user and the device as best as possible. By holding the phone in the "camera-grip" style illustrated in Fig. 2, the user minimizes that amount of heat that is transferred from their hand to the device. Holding the smartphone horizontally against the forehead ensures that the contact area spans most of the device's minor axis, and we found during our preliminary investigations that most smartphone thermistors are more sensitive to heat sources in contact with the middle or top-third of the smartphone. We therefore instruct users to prioritize making contact with this region of the device. Sensor data is recorded for a fixed duration that we investigate in our analyses, but the upper limit for the duration of heat transfer is 120 seconds.

*3.2.2 Preconditions of FeverPhone.* The working principle of thermometry is applicable to FeverPhone as long as the preconditions of thermometry are met and other heat-generating processes are either minimized or properly characterized. Prior work has measured the rate at which various software operations generate heat within smartphones (e.g., high CPU utilization, Wi-Fi or cellular activity) [12, 27]; in some cases, so much heat is generated that the smartphone becomes warmer than the user and the working principle of FeverPhone cannot be applied. In our own experiments, we found that unlocking the screen or plugging in the charger each increase the temperature by 0.5 °C to 1 °C when the device is otherwise idle and unplugged at room temperature (18.3 °C–22.2 °C). We also found that holding the device in hand also raised the temperature by a similar increment.

To minimize these confounds, we only collect data when the smartphone is in a trickle-charge state with minimal CPU utilization, which requires the device to be fully charged, plugged into an AC-portable charger, in airplane mode, and with the screen brightness fixed to 20%. Most of these constraints can be applied automatically via software and have been imposed in past work related to opportunistic temperature sensing with smartphones [11–13]. Given that FeverPhone calls for the user to place their smartphone against their forehead, we argue that these restrictions are even more reasonable since people would not be precluded from using their device any more than they would be from the active interaction required by FeverPhone in the first place. We also require the user to hold the smartphone in their hand for 60 seconds before taking a measurement since that is usually the amount of time it takes for the device to reach a steady-state temperature.

## 3.3 Sensor Data Collection and Modeling

*3.3.1 Thermistor Data.* Various thermistors with unique thermal response curves are distributed throughout different smartphone models. FeverPhone relies upon the thermistor that is most sensitive to heat conduction from an external source in contact with the phone's touchscreen. Fig. 3 illustrates the difference in thermal response curves for a selection of thermistors on a Google Pixel 6 while the device was in contact with a human forehead. We found that the strongest signals were acquired from thermistors that had lower resting temperatures. These thermistors exhibited greater dynamic ranges in temperature, and according to Newton's Law of Cooling, steeper rates of heat transfer. The fact that lower initial temperatures lead to better thermal response curves corroborates the preconditions for existing thermometers outlined in Section 3.1.2.

Given a sufficient number of temperature readings over a long period of time, Equation 1 could simply be fit to the thermistor data using an exponential model and $T_{env}$ would directly reveal the temperature of the user's forehead. However, getting reliable results using this approach requires longer contact periods (as much as 8 minutes in our experience) since exponential functions are highly sensitive to noise at smaller values of $t$. To make temperature estimation faster, we rely on the fact that the rate of change in thermistor temperature is relatively consistent during the initial few minutes of the heat transfer since there is a relatively small difference in the initial temperatures of the device and the user's forehead. This allows us to approximate Equation 1 as a

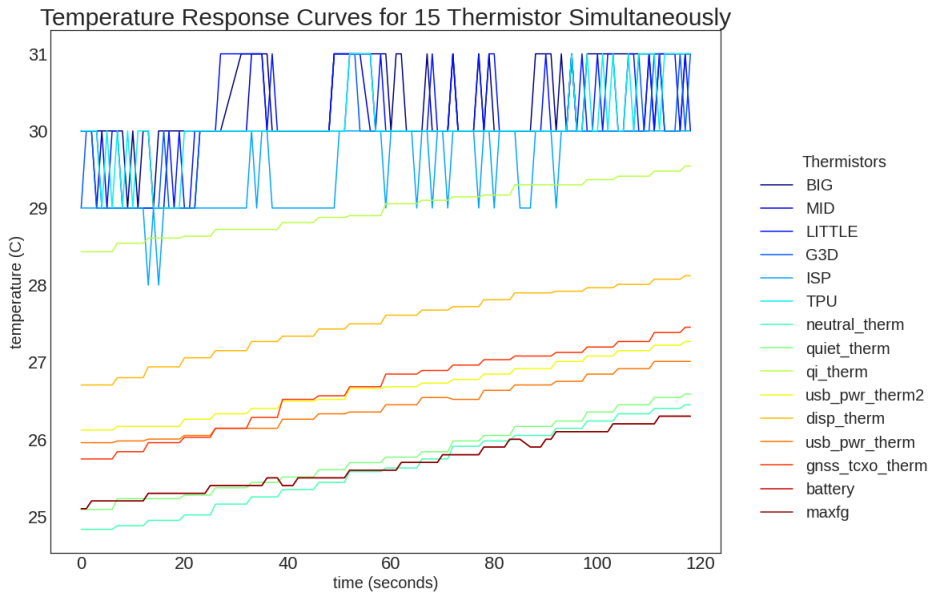


Fig. 3. An example of the differing thermal response curves for a selection of thermistors on the device during 120 seconds of contact with a human forehead.

linear function:

$$T(t) = mt + T_0 \tag{2}$$

where $m$ encodes similar characteristics to $k$ from Equation 1 along with some approximation of $(T_0 - T_{env})$ — the difference in device and body temperature during contact. We fit the temperature data from the thermistor to this model to calculate the rate-of-change $m$ and the initial temperature $T_0$ as features in our model.

*3.3.2 Capacitive Touchscreen Data.* One of the major factors that influences the thermistor signal observed by FeverPhone is the region of contact between the smartphone and the user's forehead. The closer the region is to the internal location of a given thermistor, the quicker the rate of heat transfer the thermistor experiences. This distance cannot be easily calculated since thermistor locations are typically not public knowledge, but we can account for it by extracting features that summarize where the user has placed their forehead along the smartphone according to the capacitive touchscreen.

Android offers a `TouchEvent` API for extracting the location of touchscreen interactions; however, this API assumes that all interactions are caused by fingers and can therefore be summarized as ellipses. FeverPhone instead uses a user-debug build of Android to read the raw capacitance values from the 2D array of capacitors inside the touchscreen every 5 seconds. We summarize this data by first averaging the capacitive values across all of the recorded matrices. We then mask out the cells with an average capacitance greater than 25% of the maximum capacitance value in the averaged matrix. This produces a binary mask that describes the region of contact created by the user.

To account for the distance between the region of contact and the thermistor, we calculate the vertical position of the region's weighted centroid relative to the phone's edge. To account for the amount of contact between the user's forehead and the device, we calculate the total area of the unmasked region. We explored other features like the maximum capacitance as a proxy for the amount of contact being applied against the smartphone; however,

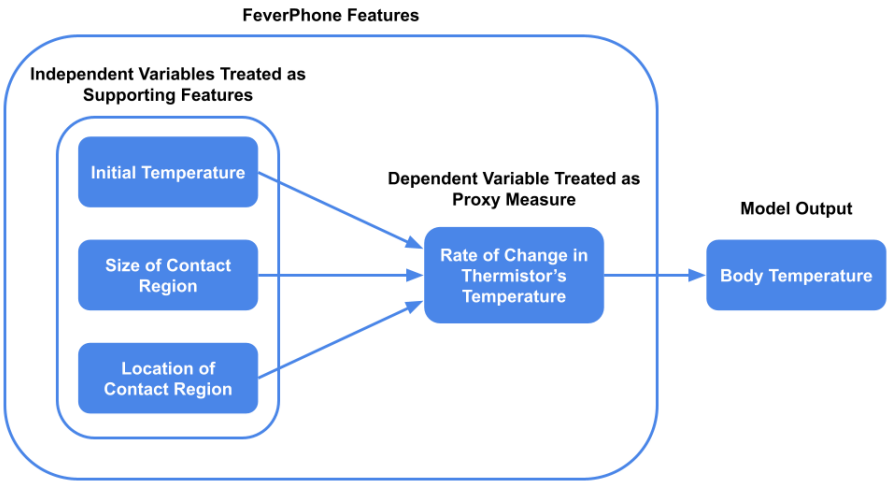


Fig. 4. A block diagram illustrating the relationship between the features used in FeverPhone. The rate of change in thermistor temperature is the primary feature in the model, while the other features provide context for that observation.

we did not find any correlation between this feature and rate of thermistor temperature change. We suspect that this is because as long as a sufficient amount of pressure is being applied to the smartphone, the rate of heat transfer from the heat source to the smartphone remains the same. However, this could also be a result of the max capacitance not being a perfect proxy of pressure since capacitance values have a limited range.

*3.3.3 Machine Learning Model.* We train a machine learning regressor to learn the complex interactions between our features as they relate to core body temperature. Our model uses four features for temperature estimation: (1) the initial temperature of the thermistor before contact with the user, (2) the rate of change of the thermistor's temperature over the data collection, (3) the vertical position of the contact surface's weighted centroid, and (4) the total area of the contact surface. As illustrated in Fig. 4, we conceptually split the features into two different groups. The rate of change in the thermistor's temperature is considered the primary feature since it is the only feature that varies with respect to the user's body temperature. The other features are considered to be supporting features, provide context to the thermal curve that is observed by the thermistor.

## 4 PRELIMINARY LAB INVESTIGATION

Before deploying FeverPhone to human participants, we first validated the working principle of the system in a lab setting using a simulated skin-like heat source: a sous-vide precision water heater. In this experiment we investigate the relationship between features. This validation methodology has been leveraged in prior work for body temperature monitoring [11, 28].

### 4.1 Experimental Procedure

We collected data on a Google Pixel 6 smartphone without any accessories (e.g., phone case, screen protector). We simulated a human forehead using a resealable plastic bag full of water and a PrimoEats sous-vide precision water heater[4]. The experimental setup for the experiment is illustrated in Fig. 5. The sous-vide was fixed to the same location inside the plastic bag throughout the experiment. The physical interaction of holding the smartphone

[4]https://www.amazon.com/Cooker-Immersion-Digital-Display-Stainless/dp/B08227CF83

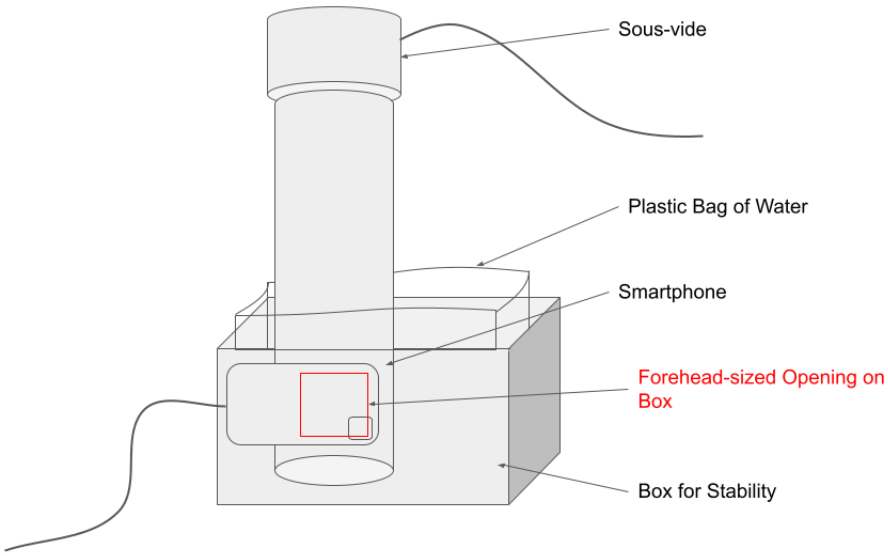


Fig. 5. An illustration of the experimental setup for the lab validation study. The sous-vide was placed inside a plastic bag filled with water and supported by a cardboard box. The phone was mounted against the cardboard box with a forehead-sized hole cut into the box to allow direct contact between the temperature controlled bag of water and the phone screen.

against the forehead was emulated by a cardboard clasp that applied enough pressure to the back of the device to make steady contact between the touchscreen and the bag of water. The bag was placed inside a cardboard box with a hole about one-third of the length of the device cut out to allow only a portion of the screen to maintain direct contact with the simulated heat source. This was done to control the region of the touchscreen that made contact with the plastic bag. The study was conducted in a laboratory room with an ambient temperature at roughly 23.6 °C.

We intentionally introduced a degree of variation during the data collection process to reduce the impact of some potential confounds. The smartphone's position was slightly adjusted between trials so that the device made contact with the bag of water at different points along the screen. We also varied the initial temperature of the smartphone by running compute-intensive tasks on the device so that it could heat up until it had reached a target temperature. The initial temperature of the smartphone was varied between 21.1 °C and 28.3 °C. Meanwhile, the sous-vide was used to sustain the water at temperatures between 36.4 °C and 39.2 °C in increments of approximately 0.25 °C[5]. This procedure resulted in the collection of 41 data samples across these dimensions.

## 4.2 Feature Relationships

As mentioned earlier, the change in the thermistor's temperature is the only feature in FeverPhone that shares a direct relationship with the temperature of the heat source being measured. Fig. 6 shows this relationship across all trials of our sous-vide experiment. The Pearson's correlation coefficient ($r$) between the two variables was 0.44 ($p < .01$), indicating that there is a weak linear correlation between them. To further explore the linear

[5]The sous-vide model used for this experiment only supported temperatures in Fahrenheit, so the exact temperature was varied in increments of 0.5 °F. The manufacturer of the sous-vide reports that the precision of their device is ±1%, although the baseline temperature from which this variance was reported is unclear.

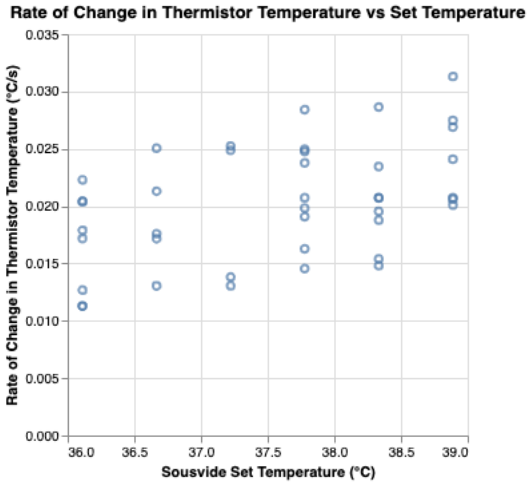


Fig. 6. A scatter plot showing the linear relationship between the rate of change of the thermistor's temperature and the sous-vide's temperature.

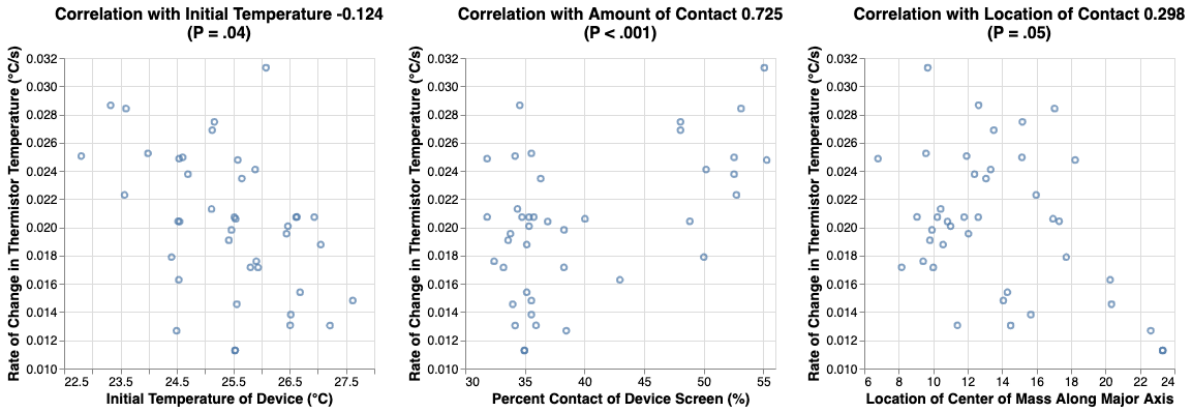


Fig. 7. Scatter plots showing the relationship between each of FeverPhone's supporting features — (from left to right) initial temperature of the smartphone, the amount of contact as measured by the touchscreen, and the location of the contact along the vertical axis — versus the rate of change of the thermistor's temperature.

relationship between the rate of change in the thermistor's temperature and the temperature of the sous-vide, we trained single-feature linear regression models using 5-fold cross-validation. This resulted in a mean absolute error (MAE) of 0.786 °C and a goodness of fit ($R^2$) of 0.194. Given that the temperature of the sous-vide ranged by approximately 3 °C, these results show that the rate of change in the thermistor's temperature provides some useful information for estimating the temperature of the heat source but not enough to be used on its own.

Although there is no meaningful relationship between the temperature of the heat source and either the initial temperature of the phone or the touch-based features, these features do share a direct relationship with the rate of heat transfer to the thermistor. We illustrate these relationships in Figure 7, which shows the correlation between each of the supporting features and the rate of change in the thermistor's temperature. The location of

contact led to the highest correlation out of the three supporting features, indicating that the heat transferred to the phone from the forehead is not transferred equally across the device and that the proximity of contact to the thermistor is very important in FeverPhone. The least correlated of the three supporting features was the initial temperature. Despite the software restrictions we impose on the smartphone during data collection (e.g., airplane mode, fully charged), the smartphone is constantly warming up or cooling down. Therefore, the initial temperature only captures a brief snapshot of the device's state before data collection. We also trained linear regression models that estimate the change in the thermistor's temperature as a function of three supporting features to further support their utility in providing context to the observed rate of heat transfer. As before, the models were trained using 5-fold cross-validation, but we express the error using mean absolute percent error since the relative changes in the output are more meaningful for FeverPhone's final temperature estimate. This experiment resulted in a mean absolute percent error of 9%, a Pearson's $r$ of 0.853, and an $R^2$ of 0.850. We then trained linear regressions models to estimate the sous-vide's temperature as a function of the primary feature and the three supporting features. After training these models using 5-fold cross validation, we achieved an MAE of 0.495 °C, a Pearson's $r$ of 0.864, and an $R^2$ of 0.715. Adding the supporting features resulted in a 43% decrease in error from using the primary feature alone, showing that the supporting features add important contextual information to the rate of change in thermistor's temperature.

## 4.3 Potential Confounds

There are a number of potential confounders that could influence FeverPhone's generalizability to real-world scenarios. To measure the impact of each one in isolation, we repeated the data collection protocol described in Section 4.1 while varying one of the following parameters: the model and instance of the smartphone, the accessories on the smartphone, the pressure applied to the smartphone by the sous-vide bag, or the ambient air temperature. After collecting 4 samples at 3 sous-vide temperature (36.7 °C, 37.8 °C, and 38.9 °C), we evaluated FeverPhone's sensitivity to these changes in two ways. The first method, which we consider to represent FeverPhone's *uncalibrated* performance, involved training a single model on the original dataset and then evaluating its performance on the new dataset with the confound introduced. The second method, which we consider to represent FeverPhone's *calibrated* performance, involved merging 10-15% (6-10 samples) of the data with confounds present with the original dataset before conducting 5-fold cross-validation; the characteristics of the datasets that were merged varied slightly depending on the nature of the confound being tested, so we clarify these details below before describing the corresponding results. We aim to show that only a small amount of data affected by confounds is necessary for the *calibrated* error to outperform both the *uncalibrated* error and dummy regression without significantly impacting the *uncalibrated* baseline error. Since this evaluation method produced predictions for all data including the baseline dataset, the reported error metrics were stratified according to confound state. The results of these experiments are presented in Table 1.

*4.3.1 Phone Model & Instance.* To test how well FeverPhone generalizes to new devices beyond the Google Pixel 6 used for baseline training, we collected additional data using a second Google Pixel 6, a Google Pixel 3, and a Huawei P20. In the uncalibrated approach, the MAE increased to 0.709 °C when the baseline model was applied on the other Pixel 6. Although the increase in error may indicate that there were subtle differences in the underlying hardware components that were sourced for these two instances of the same device, the moderate change in error indicates that the model was still able to generalize across instances with reasonable accuracy. As expected, the MAE rose significantly when the model trained on the Pixel 6 was deployed on other phone models with significantly different hardware composition (e.g., thermistor location, case material, heat capacity), to the point where they both exceeded the MAE of the dummy regression. Given the heterogeneity across devices, our calibration approach entailed merging the baseline dataset with only the dataset of the new phone being introduced rather than merging all phones into a single dataset before cross-validation. This procedure

Table 1. A table of regression metrics for the experiments used to evaluate the impact of various confounds. The baseline model was trained on data collected in a 23.6 °C room using a Google Pixel 6 without any accessories that was being pressed against the sous-vide bag with moderate pressure. Each row indicates the only variable that was modified for each given experiment.

| | | Uncalibrated | | Calibrated | | | |
|---|---|---|---|---|---|---|---|
| **Confound** | **Variation** | **Test MAE** | **Test Pearson's *r*** | **Test MAE** | **Test Pearson's *r*** | **Baseline MAE** | **Baseline Pearson's *r*** |
| Baseline | Dummy regression | 1.072 °C | -1.000 | — | — | — | — |
| Baseline | Baseline dataset | 0.495 °C | 0.864 | — | — | — | — |
| Phone Model | Different Google Pixel 6 | 0.709 °C | 0.491 | 0.790 °C | 0.701 | 0.680 °C | 0.720 |
| Phone Model | Google Pixel 3 | 8.76 °C | 0.375 | 0.825 °C | 0.312 | 0.852 °C | 0.448 |
| Phone Model | Huawei P20 | 4.12 °C | 0.410 | 0.649 °C | 0.599 | 0.728 °C | 0.632 |
| Accessory | Screen protector | 1.094 °C | 0.823 | 0.634 °C | 0.691 | 0.664 °C | 0.754 |
| Accessory | Phone case | 1.285 °C | 0.751 | 0.692 °C | 0.678 | 0.664 °C | 0.754 |
| Accessory | Phone case and screen protector | 0.813 °C | 0.775 | 0.695 °C | 0.797 | 0.664 °C | 0.754 |
| Pressure | Gentle | 0.845 °C | 0.878 | 0.545 °C | 0.831 | 0.595 °C | 0.804 |
| Pressure | Forceful | 1.606 °C | 0.553 | 0.874 °C | 0.758 | 0.595 °C | 0.804 |
| Ambient Temperature | 21.3 °C | 0.610 °C | 0.745 | 0.584 °C | 0.766 | 0.545 °C | 0.786 |
| Ambient Temperature | 16.9 °C | 1.180 °C | 0.952 | 0.629 °C | 0.745 | 0.545 °C | 0.786 |

significantly improved the correlation for the other Google Pixel 6 at the cost of increasing the MAE slightly. Meanwhile, this procedure significantly decreased the MAE for the Huawei P20 and the Google Pixel 3 to be under 1 °C. After calibrating the models for each phone model, the performance those models had on the baseline dataset slightly decreased. These results indicate that FeverPhone can generalize fairly well across different devices with some amount of calibration data, but the benefit is far more profound when the devices vary significantly in hardware.

*4.3.2 Phone Accessories.* To test FeverPhone's robustness when a person places a phone case or screen protector on their device, we collected additional data with these phone accessories independently and together. In the uncalibrated approach, we found that both accessories increased the MAE near that of the dummy regression but with significantly higher correlation. The error was higher with the phone case than it was with the screen protector, which is likely due to its higher heat capacity. To our surprise, using the two accessories in tandem actually led to better estimates in terms of MAE, although the MAE of that model was still worse than the MAE on the baseline dataset. Since these accessories vary the heat capacity in a continuous manner, our calibration approach entailed merging the baseline dataset with the data for all three variations before cross-validation. This procedure improved the MAE of all three configurations while decreasing their correlation in two cases. The calibrated model also had similar performance on the baseline dataset. These findings indicate that larger or more insulating phone accessories will result in a greater impact on model performance after calibration. Although it

is possible to account for these errors with calibration, temporarily removing these accessories may be a more robust way of improving FeverPhone estimates.

*4.3.3 Pressure of Device Application.* To test how well FeverPhone generalizes to different amounts of pressure being applied by the user, we collect additional data with notably "gentle" and "forceful" pressure. We note that more rigorous definitions of pressure are difficult to specify since the water bag is not rigid and applies a counter-pressure of its own; a load cell would have also absorbed some heat on its own, thereby impacting our measurements. In the uncalibrated approach, the MAE for the gentle and forceful pressure scenarios was 0.845 °C and 1.606 °C respectively. The temperature estimates for the forceful pressure scenario tended to be higher than expected since more heat was being transferred from the sous-vide bag to the thermistor, and the converse was true for gentle pressure scenario. The significantly worse accuracy in the forceful pressure scenario may be partially attributed to deformation of the sous-vide bag, which may have exposed the smartphone to heat in different ways than anticipated. Since pressure is a continuous quantity, our calibration approach entailed merging the baseline dataset with the data for both of the other pressure applications. The MAE across both of the unseen pressure levels improved after calibration, but the error in the forceful pressure scenario was still relatively high. These findings indicate that FeverPhone is reasonably robust to pressure of application when trained with measurements across different amounts of pressure. Furthermore, forceful pressure may have more of a deleterious impact than gentle pressure.

*4.3.4 Ambient Air Temperature.* To test how well FeverPhone generalizes to different ambient air temperatures beyond the 23.6 °C room that was used for baseline training, we collected additional data in a 21.3 °C room and a 16.9 °C room. In the uncalibrated approach, the MAE degraded as the ambient air temperature deviated further from the temperature in which the model was trained. This result can be attributed to the fact that the baseline model was never trained on data samples where the initial thermistor temperature reached these colder temperatures and was thus unable to generalize. Since ambient temperature is a continuous quantity, our calibration approach entailed merging the baseline dataset with the data for both of the other room temperatures before cross-validation. Although the MAE at the original room temperature increased by nearly 0.05 °C, the MAE across the colder room temperatures improved to within 0.1 °C of the calibrated MAE on the baseline dataset. These findings indicate that FeverPhone can reasonably generalize to new ambient air temperatures as long as the training dataset includes variety in the initial temperature of the thermistor.

*4.3.5 Summary.* We observed that all the examined confounds negatively impacted the performance of FeverPhone when the uncalibrated approach was used for testing. Attempting to generalize FeverPhone across completely different phone models, using a phone case, or applying forceful pressure increased the error to worse than the dummy regression, albeit with significantly better Pearson's $r$. For most of the other confound variations, the MAE remained close to within 1 °C. By incorporating some additional training data across a broader set of conditions (e.g., ambient temperatures, amounts of pressure), we were able to calibrate the model to accommodate these scenarios. The MAE fell within 1 °C for all confound variations without significantly hampering the performance on the baseline dataset.

## 5 CLINICAL VALIDATION STUDY

Having supported the validity of FeverPhone's design through our lab validation study, we deployed our app to a local emergency department to assess its efficacy with human participants as both a thermometer and a low-grade fever detector. We answer the following questions related to the design of FeverPhone along the way:

**RQ.1** How does the selection of the machine learning model impact the performance of FeverPhone?
**RQ.2** How does the duration of data collection impact the performance of FeverPhone?
**RQ.3** How does the inclusion or exclusion of features impact performance of FeverPhone?

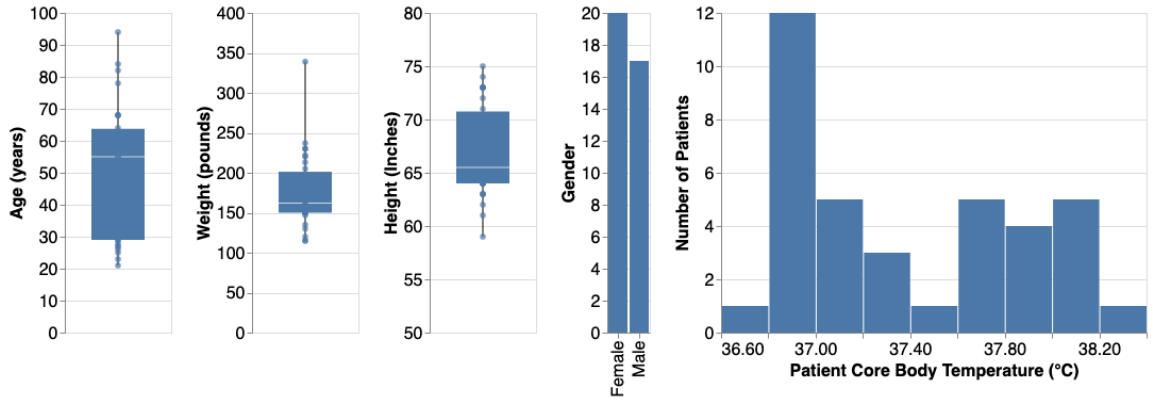


Fig. 8. The demographic distribution of participants in the clinical study.

**RQ.4** How can FeverPhone detect fevers in the absence of labeled febrile training samples?

## 5.1 Participants

Individuals who arrived at the University of Washington's Emergency Department were given the opportunity to participate in our study regardless of their reason for visiting the hospital. The only inclusion criteria related to the study were that the participants had to be 18 years or older, able to provide informed consent, and in a stable medical condition. Participants were excluded from our protocol if they had their blood drawn immediately before recruitment or had taken an antipyretic drug like Tylenol since both of these actions can rapidly impact core body temperature. A summarization of participant demographics can be found in Fig 8.

Prior to enrollment, eligible participants had their temperature measured using Welch-Allyn SureTemp Plus 690 oral thermometer. This thermometer also provided our ground-truth measurements since it had a reported accuracy of ±0.1 °C and was able to deliver readings noninvasively within 10 seconds; the convenience of the thermometer was particularly important for minimizing interference with patient care. Participants with nominal core body temperatures (~37 °C) were often patients who were visiting the hospital for ailments like physical trauma that do not elevate core body temperature. Participants with at least a low-grade fever (≥37.5 °C) were verified by clinicians to have known cause of fever, often a viral or bacterial infection. Participants rarely presented with a high-grade fever above 38.6 °C, and in these cases, their skin was so sweaty that it was difficult to gather a reliable measurement with FeverPhone. This is a known limitation not only with FeverPhone, but also other temporal and peripheral thermometers; temporal thermometers and other skin contact-based thermometers suffer from reduced accuracy when the skin is sweaty or even clammy [5, 29, 40]. Since it is easy to identify fevers in such individuals without thermometry, we excluded these participants from our analyses. In total, we recruited 37 participants (17 male, 20 female), 16 of whom presented with at least a low-grade fever. The average age of our participants was 51 ± 20 years.

## 5.2 Experimental Procedure

After obtaining informed consent and a ground-truth measurement of the participant's temperature, a research team member measured the participant's temperature using FeverPhone implemented on a Google Pixel 6 smartphone. To avoid interfering with patient care, all trials were conducted opportunistically. Data collection occurred in different parts of the emergency department. Patients in the waiting room and hallway were sitting

Table 2. A table of regression and classification metrics for various model architectures. All of the models were trained using all four features based on the first 90 seconds of sensor data.

| Model Type | MAE | Pearson's $r$ | Limit of Agreement | Sensitivity | Specificity |
|---|---|---|---|---|---|
| Dummy regression | 0.452 °C | -0.884 | 1.017 °C | 0.000 | 1.000 |
| Linear regression | 0.320 °C | 0.621 | 0.754 °C | 0.750 | 0.857 |
| Quadratic regression | 0.367 °C | 0.557 | 0.897 °C | 0.625 | 0.857 |
| Random forest | **0.229 °C** | **0.763** | **0.731 °C** | **0.813** | **0.904** |

upright while patients in private rooms were laying down in beds. Since some individuals were too weak to hold the smartphone on their own, the researcher held the smartphone against the participant's forehead for them.

Participants were asked to have their temperature recorded by FeverPhone up to three times, but this was not always possible without impacting patient care. Participants who were willing and able to contribute repeated trials waited 10 minutes in between to allow both their forehead and the smartphone to acclimate back to their steady-state temperatures. The phone was wiped down using a sanitary alcohol wipe between trials, which simultaneously minimized the spread of germs while cooling the device back down between trials. New ground-truth measurements were also collected immediately before each repeated trial.

## 5.3 Analysis

We trained our regression models using leave-one-out validation such that one model was trained for each participant using all of the other participants' data. We report the average error across all folds according to the MAE, limit of agreement, and Pearson's correlation coefficient ($r$). We further contextualize the results by reporting the accuracy with which FeverPhone is able to discriminate febrile participants. Rather than directly training a classification model that could have different behavior from the regression model, we instead report classification accuracy by binarizing the ground-truth temperatures and FeverPhone's estimates according to the threshold for low-grade fever (≥37.5 °C) and then comparing their labels. To demonstrate the utility of the various FeverPhone model configurations, we also include the results of a dummy regression model that simply guesses the median of the dataset.

## 5.4 Results

*5.4.1 RQ.1: Model Selection.* We tested three traditional regression models suited to handle low-dimensional data with varying degrees of nonlinearity: linear regression, quadratic regression, and random forest regression. The models were trained using all four features based on sensor data that was collected over 90 seconds. The results of this analysis are presented in Table 2.

Random forest regression outperformed the other options with respect to both regression and classification. Although there were linear relationships between the variables explored during our lab simulation study, the idealized conditions did not replicate the nonlinear nature of heat transfer that is inherent to the human body. This is one potential explanation for why random forest regression outperformed the other models. The ability of the random forest regression to estimate core body temperature is further illustrated in Fig. 9. The model achieved an MAE of 0.229 °C, a Pearson's $r$ of 0.763 ($p < .001$), and a limit of agreement of 0.731 °C. Framing the performance of the random forest regression in terms of fever classification, the model achieved a sensitivity of 0.813 and a specificity of 0.904. These results are comparable and occasionally superior to some consumer-grade peripheral thermometers for both temperature regression and fever detection according to findings by Niven et al. [41] and Farnell et al. [18].

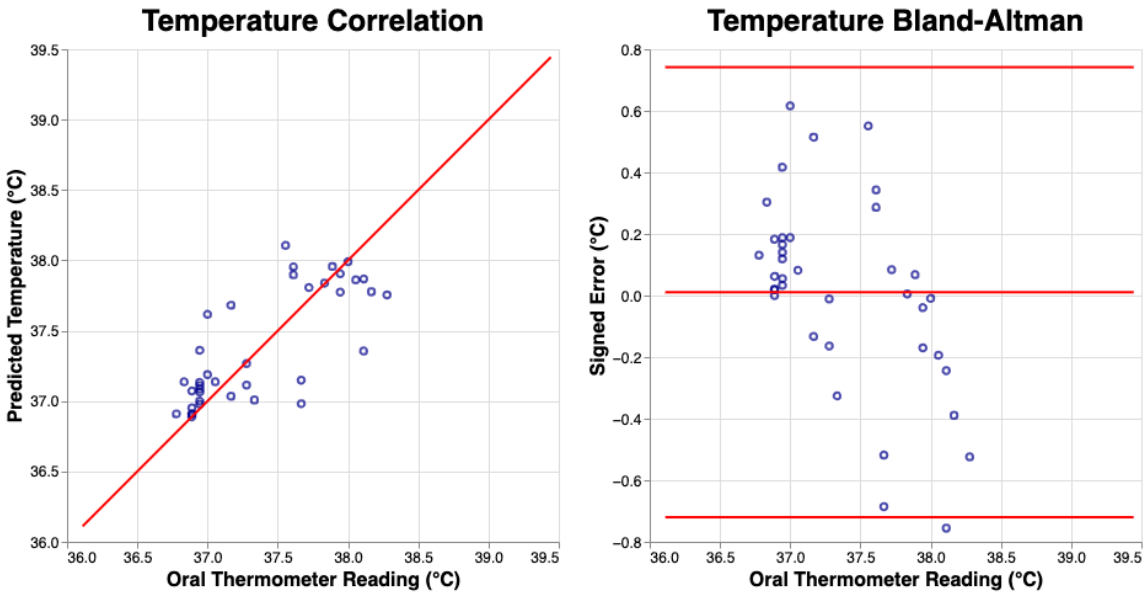


Fig. 9. The (left) correlation and (right) Bland-Altman plot for the temperature estimates made by the random forest regression model. The models were trained using all four features based on sensor data that was collected over 90 seconds. The red lines on the Bland-Altman plot indicate the 95% limit of agreement.

*5.4.2 RQ.2: Duration of Data Collection.* To investigate the optimal duration of contact between the user and the smartphone, we simulated different durations by restricting feature extraction to varying amounts of sensor data. The models were trained using random forest regression and all four features. The results of this analysis are presented in Table 3.

Increasing the duration of data collection generally improved the accuracy of FeverPhone and decreased the limit of agreement, which follows the intuition that more sensor data should lead to a more accurate representation of the thermistor's temperature curve. However, the performance started to degrade when the duration of data collection approached 120 seconds. This is likely due to the fact that the linear approximation we used for Newton's Law of Cooling was no longer applicable at these longer durations. We found that restricting feature extraction to the first 90 seconds of contact resulted the optimal results. This duration is shorter than what is required for mercury-in-glass thermometers and non-predictive digital oral thermometers, which can take between 30 seconds and 10 minutes to produce reliable estimates [3]. Although infrared temporal thermometers and consumer-grade digital oral thermometers with built-in prediction algorithms can make estimates faster, it is important to note that these devices can still suffer from high error [41].

*5.4.3 RQ.3: Impact of Features.* We explore the importance of various subsets of features. We first examine the performance of FeverPhone based on the primary feature — the rate of change in the thermistor's temperature — on its own. We then examine how adding supporting features from the thermistor or the touchscreen impact the model's performance. The models were trained using random forest regression based on sensor data that was collected over 90 seconds. The results of this analysis are shown in Table 4.

The model with all of the features led to the best performance as expected, but the baseline model with just the primary feature was not necessarily the worst performing configuration. Adding the supporting features from

Table 3. A table of regression and classification metrics for models trained with features extracted from varying amounts of sensor data. All of the models were trained using random forest regression and all four features.

| Duration | MAE | Pearson's *r* | Limit of Agreement | Sensitivity | Specificity |
|---|---|---|---|---|---|
| Dummy regression | 0.452 °C | -0.884 | 1.017 °C | 0.000 | 1.000 |
| 30 sec | 0.417 °C | 0.224 | 0.951 °C | 0.375 | 0.714 |
| 60 sec | 0.324 °C | 0.533 | 0.822 °C | 0.625 | 0.809 |
| 70 sec | 0.250 °C | 0.717 | 0.764 °C | **0.813** | 0.857 |
| 80 sec | 0.236 °C | 0.747 | 0.724 °C | **0.813** | 0.857 |
| 90 sec | **0.229 °C** | **0.763** | 0.731 °C | **0.813** | **0.904** |
| 120 sec | 0.240 °C | 0.743 | **0.691 °C** | **0.813** | **0.904** |

Table 4. A table of regression and classification metrics for different feature sets. All of the models were trained using random forest regression with features based on the first 90 seconds of sensor data.

| Feature Subset | MAE | Pearson's *r* | Limit of Agreement | Sensitivity | Specificity |
|---|---|---|---|---|---|
| Dummy regression | 0.452 °C | -0.884 | 1.017 °C | 0.000 | 1.000 |
| ROT + IT + CA + CL | **0.229 °C** | **0.763** | **0.731 °C** | **0.813** | 0.904 |
| ROT + CA + CL | 0.251 °C | 0.736 | 0.746 °C | **0.813** | 0.904 |
| ROT + IT | 0.311 °C | 0.610 | 1.028 °C | **0.813** | **1.000** |
| ROT | 0.309 °C | 0.582 | 0.979 °C | **0.813** | 0.904 |

ROT = rate of change in the thermistor's temperature; IT = initial thermistor temperature; CA = contact area; CL = contact location.

the touchscreen data improved the model, which was expected since the proximity to the thermistor and the amount of contact both control the amount of heat transferred from the user to the thermistor. However, adding the initial thermistor temperature as the only supporting feature degraded model performance with respect to metrics like MAE and AUC. This indicates that the touchscreen-based features provide important context to the thermistor readings and that there is an interaction between the supporting features that improves performance only if they are both included.

Although the best performing model in this experiment included all of the features, we anecdotally explored various combinations of features in other configurations. Most notably, we found that while the performance of the model including both thermistor and touchscreen-based features did not improve with interaction durations approaching 120 seconds, the model trained on just thermistor-based features did. This result indicates that the addition of touchscreen-based features allowed the model to be robust over shorter durations, whereas in the absence of these features, the model benefited from longer interactions.

*5.4.4 RQ.4: Fever Detection without Labeled Samples.* In Section 4.3.1, we found that FeverPhone requires significant calibration to work on new devices. Although it is best to recollect a new dataset for each phone model, it can be difficult to acquire new training samples from febrile patients. Anomaly detection has been used in past work to distinguish healthy data samples from unhealthy ones in domains ranging from healthcare to

mechanical engineering [9, 33], so we explored the potential of using anomaly detection for unsupervised fever classification. For FeverPhone, we assume that data samples from afebrile patients are more readily available during the training of a new phone model. Rather than fitting all four of FeverPhone's features to core body temperature using this dataset, we suggest the training of a model that maps the 3 supporting features to the rate of heat transfer — the primary feature that is correlated with core body temperature. This procedure generates a regression model that estimates the rate of heat transfer for healthy individuals. When this model is used for testing, the predicted rate of heat transfer is compared to the rate that is actually measured by the thermistor. If the difference between the two exceeds a predefined threshold, then we surmise that the rate of heat transfer has been significantly mischaracterized and that the user likely has a fever provided that there is sufficient coverage in the training dataset.

We evaluated this approach using the FeverPhone configuration selected for our previous research questions, only this time by training a model to estimate the rate of heat transfer from the three supporting features. This model was trained using 5-fold cross-validation on all the data from patients without a low-grade fever. Using a threshold of -0.001 to classify samples as anomalies and testing this model on the febrile patients, we found that this unsupervised approach achieved a sensitivity of 1.00 and specificity of 0.713.

*5.4.5 Summary.* To summarize our analyses, the optimal configuration of FeverPhone uses a random forest regression model trained on features from the thermistor and the capacitive touchscreen to describe the rate of heat transfer from the forehead to the smartphone. These features are optimally calculated over a 90-second period, which provides a suitable amount of time to accurately characterize the rate of heat transfer while still obeying assumptions related to linearity. In this configuration, FeverPhone achieved an MAE of 0.229 °C, a limit of agreement of ±0.731 °C, and a Pearson's $r$ of 0.763. When those readings are used for fever classification, FeverPhone achieved a sensitivity and specificity of 0.813 and 0.904 respectively. We also found that it is possible to detect low-grade fevers without training a model on such instances. Using unsupervised anomaly detection, we were able to identify febrile patients with sensitivity and specificity of 1.00 and 0.713 respectively.

## 6 DISCUSSION

This work serves as a foundation for core body temperature measurement using unmodified smartphones. Our goal is not to replace existing thermometers but rather to develop an accessible alternative for cases when traditional thermometers are not readily available, namely telehealth consultations and resource-constrained environments. Nevertheless, we demonstrated that FeverPhone achieves accuracy levels that are comparable to some consumer-grade peripheral, tympanic, and auxiliary thermometers. In the discussion that follows, we elaborate upon the impact of our findings, considerations for future work, and the limitations of our research.

### 6.1 Key Findings

The design of FeverPhone is grounded in Newton's Law of Cooling, the same working principle that underlies most clinical-grade thermometers. However, using this equation alone was not enough to make FeverPhone feasible, which led to two innovations. First, we found that properly characterizing the exponential rate of heat transfer requires as long as 8 minutes of data collection, which would be unsuitable for human interaction. Our results demonstrate that approximating Newton's Law of Cooling using a linear function over a shorter period can lead to satisfactory results. Second, we observed that the region of contact between the user and the smartphone's screen can vary according to how they position the phone against themselves and the curvature of the forehead itself. Since the contact surface influences the rate of heat transfer from the user to the thermistor, our feature ablation study shows that incorporating features representative of it improves the accuracy of FeverPhone. These two innovations should be useful to any future systems that would rely on direct contact between the phone and another object for temperature estimation.

### 6.2 Supporting Practical Scalability of FeverPhone

One of FeverPhone's supporting features relates to the weighted centroid of the region of contact between the user and the smartphone as measured by the touchscreen. Conceptually, this feature was designed to encode information related to the distance between that region and the thermistor within the phone. When the thermistor's position is fixed and known, the distance can be calculated using a uniform offset. However, the same offset cannot be applied to other phones since their thermistors are likely in different locations, making this feature in particular difficult to generalize across smartphone models. The distance could be directly calculated if smartphone manufacturers were to expose the location of their thermistors to model developers, which we believe would dramatically improve the generalizability of FeverPhone. It would also be beneficial for FeverPhone if smartphone operating systems were to provide an API for throttling temperature-sensitive tasks like network utilization and momentarily increasing the sensor resolution. Although we found the optimal duration of data collection to be 90 seconds, we believe that a shorter duration could be feasible given a more precise and higher resolution thermistor. Smartphone operating systems could also provide an API for directly reading raw touchscreen capacitance and thermistor values to make FeverPhone and similar applications more ubiquitous. Finally, manufacturers could provide open-access data on characteristic thermal curves for the thermistors on their devices, which would allow for faster model generation. This data could be generated by logging thermistor data during factory stress tests. With enough data across phone models and thermistors, a more general model could be created based on thermistor location as well as the phone's material, size, and mass. Having such a model would significantly improve the ability to generalize FeverPhone across smartphone models.

### 6.3 Estimating Core Body Temperature from Multiple Locations

The mapping between external and core body temperature could be facilitated by applying the working principle of FeverPhone to other smart devices such as smartwatches or other body-mounted wearables. Although we selected the forehead for ergonomic and stability reasons, future investigations may find other testing sites to have their own advantages. Leveraging multiple testing sites across the body could also yield performance benefits. Beyond providing redundant measurements that can be compared and averaged, thermoregulation may vary across body sites in a way that reveals new information about the user's core body temperature. Combining measurements across the body could also serve as an indicator of ambient air temperature or user activity as the surface temperature of different body sites have been shown to be impacted by those factors at different rates [15, 16]. Despite these opportunities, idiosyncrasies in people's physiology may require user-specific modeling to properly reconcile measurements across the body, so we leave this opportunity to future work.

### 6.4 Limitations

Although we explored a number of confounds in both our lab validation and clinical studies, there were other confounds in that latter that were not characterized. For example, participants used FeverPhone in varied postures due to the nature of their medical ailments. Some participants were required to be upright while others had to be reclined as part of their treatment, which may have impacted their thermoregulation during our trials. Trials were also conducted in varying locations of the emergency department (waiting room, hallway, patient room). We explored the impact of ambient air temperature in our lab validation study, but subtle changes across the emergency department may have contributed to the variance in our clinical study.

Another limitation to our study was the range of core body temperatures included in our study. We focused specifically on FeverPhone's performance on people who had nominal (36.6 °C–37.2 °C), abnormal (37.2 °C–37.5 °C), low-grade fever (37.5 °C–38 °C), or borderline fever (~38 °C). People with core body temperature within these ranges are the most likely to be unsure of their condition and could benefit from a fever screening technology. We omitted participants with extremely severe fevers (>38.6 °C) since the need for fever screening is less important

for these individuals and commercial thermometers are known to have reduced accuracy when people have sweaty or clammy skin [5, 29, 40]. Nevertheless, future work may want to consider additional features to identify these potential confounds so that FeverPhone can better support people with severe fevers.

Lastly, FeverPhone uses root-level access for reading a broad set of thermistors and a custom user-debug Android kernel for accessing raw touchscreen capacitance data. These requirements limit FeverPhone to phones with fine-grained sensor access. Temperature data from the battery thermistor is more readily accessible but did not yield comparable performance relative to temperature data from other thermistors. Still, future work could explore a more accessible version of FeverPhone with specific considerations paid to the battery thermistor. Similarly, the `TouchEventAPI` is publicly accessible at the cost of coarse-grained information about touch location. The use of this API in place of the raw capacitance data could be explored in future work.

## 7 CONCLUSION

We present FeverPhone, a system for estimating core body temperature and identifying low-grade fevers using a commodity off-the-shelf smartphone. FeverPhone adheres to the same constraints as existing contact-based thermometers and has comparable accuracy to some consumer-grade temporal, tympanic, and other peripheral thermometers. However, FeverPhone only relies on the smartphone's built-in thermistors and touch screen to serve this functionality. In a clinical trial with human participants of nominal and elevated core body temperature, we found FeverPhone to have an MAE of 0.229 °C, a limit of agreement of ±0.731 °C, and a Pearson's correlation coefficient of 0.763. When those readings are used for fever classification, FeverPhone achieved a sensitivity and specificity of 0.813 and 0.904 respectively. We believe that, if deployed at scale, FeverPhone has the potential to improve remote care and even epidemiology by providing an accessible method of temperature screening to smartphone users.

## ACKNOWLEDGMENTS

The Research was approved by the University of Washington Institutional Review Board (IRB) under protocol ID STUDY00008912 and was supported by the University of Washington Gift Fund. We would like to thank the participants and staff at the University of Washington Emergency Department for making this research possible.

## REFERENCES

[1] [n. d.]. Fever. https://www.cedars-sinai.org/health-library/diseases-and-conditions/f/fever.html. Accessed: 2022-06-30.
[2] [n. d.]. Huawei's clever new smartphone can take your temperature. https://www.fastcompany.com/90513189/huaweis-new-smartphone-can-take-your-temperature. Accessed: 2021-02-18.
[3] [n. d.]. Temperature: Digital and Glass Thermometers. https://www.nationwidechildrens.org/family-resources-education/health-wellness-and-safety-resources/helping-hands/temperature-digital-and-glass-thermometers. Accessed: 2022-08-15.
[4] [n. d.]. Temperature Measurement for Patients with Fever. https://www.uspharmacist.com/article/temperature-measurement-for-patients-with-fever. Accessed: 2022-10-25.
[5] [n. d.]. Thermometers: Understand the options. https://www.mayoclinic.org/diseases-conditions/fever/in-depth/thermometers/art-20046737. Accessed: 2022-07-31.
[6] Anwar Ahmed and Birhanu Sintayehu. 2022. Implementation of Covid-19 protection protocols and its implication on learning & teaching in public schools. *Heliyon* 8, 5 (2022), e09362.
[7] Forsad Al Hossain, Andrew A Lover, George A Corey, Nicholas G Reich, and Tauhidur Rahman. 2020. FluSense: a contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–28.
[8] Diego Barrettino, Thomas Gisler, Christoph Zumbühl, Christian Di Battista, Raphael Kummer, and Markus Thalmann. 2020. Wearable Medical Device for Remote Monitoring the Health of Elderly People at Home. In *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 1–6.
[9] Nivedita Bijlani, Ramin Nilforooshan, Samaneh Kouchaki, et al. 2022. An Unsupervised Data-Driven Anomaly Detection Approach for Adverse Health Conditions in People Living With Dementia: Cohort Study. *JMIR aging* 5, 3 (2022), e38211.

[10] Guy Boivin, Isabelle Hardy, Guy Tellier, and Jean Maziade. 2000. Predicting influenza infections during epidemics with use of a clinical case definition. *Clinical infectious diseases* 31, 5 (2000), 1166–1169.
[11] Joseph Breda and Shwetak Patel. 2021. Intuitive and Ubiquitous Fever Monitoring Using Smartphones and Smartwatches. *arXiv preprint arXiv:2106.11855* (2021).
[12] Joseph Breda, Amee Trivedi, Chulabhaya Wijesundara, Phuthipong Bovornkeeratiroj, David Irwin, Prashant Shenoy, and Jay Taneja. 2019. Hot or Not: Leveraging Mobile Devices for Ubiquitous Temperature Sensing. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 41–50.
[13] Nguyen Hai Chau. 2019. Estimation of air temperature using smartphones in different contexts. *Journal of Information and Telecommunication* 3, 4 (2019), 494–507.
[14] Hsuan-Yu Chen, Andrew Chen, and Chiachung Chen. 2020. Investigation of the impact of infrared sensors on core body temperature monitoring by comparing measurement sites. *Sensors* 20, 10 (2020), 2885.
[15] Charmaine Childs. 2018. Body temperature and clinical thermometry. *Handbook of clinical neurology* 157 (2018), 467–482.
[16] Joon-Ho Choi and Vivian Loftness. 2012. Investigation of human body skin temperatures as a bio-signal to indicate overall thermal sensations. *Building and Environment* 58 (2012), 258–269.
[17] Frances Doyle, W John Zehner, and Thomas E Terndrup. 1992. The effect of ambient temperature extremes on tympanic and oral temperatures. *The American journal of emergency medicine* 10, 4 (1992), 285–289.
[18] Sarah Farnell, Lorraine Maxwell, Seok Tan, Andrew Rhodes, and Barbara Philips. 2005. Temperature measurement: Comparison of non-invasive methods used in adult critical care. *Journal of Clinical Nursing* 14, 5 (2005), 632–639.
[19] Chen Ge, Xie Jiarong, Dai Guangli, Peijun Zheng, Hu Xiaqing, Lu Hongpeng, Xu Lei, Chen Xueqin, and Chen Xiaomin. 2020. Validity of the use of wrist and forehead temperatures in screening the general population for covid-19: A prospective real-world study. *Iranian Journal of Public Health* 49, Suppl 1 (2020), 57.
[20] ThM Govaert, GJ Dinant, K Aretz, and JA Knottnerus. 1998. The predictive value of influenza symptomatology in elderly people. *Family practice* 15, 1 (1998), 16–22.
[21] Kyra H Grantz, Hannah R Meredith, Derek AT Cummings, C Jessica E Metcalf, Bryan T Grenfell, John R Giles, Shruti Mehta, Sunil Solomon, Alain Labrique, Nishant Kishore, et al. 2020. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature communications* 11, 1 (2020), 1–8.
[22] Amos Grünebaum, Frank A Chervenak, Laurence B McCullough, Joachim W Dudenhausen, Eran Bornstein, and Philip A Mackowiak. 2021. How fever is defined in COVID-19 publications: a disturbing lack of precision. *Journal of Perinatal Medicine* 49, 3 (2021), 255–261.
[23] Wei-jie Guan, Zheng-yi Ni, Yu Hu, Wen-hua Liang, Chun-quan Ou, Jian-xing He, Lei Liu, Hong Shan, Chun-liang Lei, David SC Hui, et al. 2020. Clinical characteristics of coronavirus disease 2019 in China. *New England journal of medicine* 382, 18 (2020), 1708–1720.
[24] Liang He, Youngmoon Lee, and Kang G Shin. 2020. Mobile device batteries as thermometers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–21.
[25] Robert P Hirten, Matteo Danieletto, Lewis Tomalin, Katie Hyewon Choi, Micol Zweig, Eddye Golden, Sparshdeep Kaur, Drew Helmus, Anthony Biello, Renata Pyzik, et al. 2021. Use of physiological data from a wearable device to identify SARS-CoV-2 infection and symptoms and predict COVID-19 diagnosis: observational study. *Journal of medical Internet research* 23, 2 (2021), e26107.
[26] Minh Long Hoang, Marco Carratù, Vincenzo Paciello, and Antonio Pietrosanto. 2021. Body temperature—indoor condition monitor and activity recognition by MEMS accelerometer based on IoT-alert system for people in quarantine due to COVID-19. *Sensors* 21, 7 (2021), 2313.
[27] Yoshitaka Inui, Satoshi Hirayama, and Tadashi Tanaka. 2019. Detailed estimation method of heat generation during charge/discharge in lithium-ion battery using equivalent circuit. *Electronics and Communications in Japan* 102, 12 (2019), 3–14.
[28] Sanghoon Jun, Kilho Lee, and Jinkyu Lee. 2022. TherMobile: Measuring Body Temperature Using a Mobile Device. *IEEE Sensors Journal* (2022).
[29] JA Kistemaker, EA Den Hartog, and HAM Daanen. 2006. Reliability of an infrared forehead skin thermometer for core temperature measurements. *Journal of medical engineering & technology* 30, 4 (2006), 252–261.
[30] S Kohno, T Otsubo, K Hara, Y Tomii, and J Seki. 1995. Program and abstracts of the 35th Interscience Conference on Antimicrobial Agents and Chemotherapy. (1995).
[31] Fan Lai, Xin Li, Qi Wang, Yingjuan Luo, Xin Wang, Xiuhua Huang, Jiajia Zhang, Jieru Peng, Qin Wang, Li Fan, et al. 2022. Reliability of Non-Contact Infrared Thermometers for Fever Screening Under COVID-19. *Risk management and healthcare policy* 15 (2022), 447.
[32] Lari Lawson, Elizabeth J Bridges, Isabelle Ballou, Ruthe Eraker, Sheryl Greco, Janie Shively, and Vanessa Sochulak. 2007. Accuracy and precision of noninvasive temperature measurement in adult intensive care patients. *American journal of critical care* 16, 5 (2007), 485–496.
[33] Yaguo Lei, Naipeng Li, Liang Guo, Ningbo Li, Tao Yan, and Jing Lin. 2018. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical systems and signal processing* 104 (2018), 799–834.
[34] Rainer Lenhardt and Daniel I Sessler. 2006. Estimation of mean body temperature from mean skin and core temperature. *The Journal of the American Society of Anesthesiologists* 105, 6 (2006), 1117–1121.

[35] Christine E Long, Caroline B Hall, Coleen K Cunningham, Leonard B Weiner, Kathleen P Alger, Maria Gouveia, Carol B Colella, Kenneth C Schnabel, and William H Barker. 1997. Influenza surveillance in community-dwelling elderly compared with children. *Archives of family medicine* 6, 5 (1997), 459.

[36] Alex Mariakakis, Edward Wang, Shwetak Patel, and Mayank Goel. 2019. Challenges in Realizing Smartphone-Based Health Sensing. *IEEE Pervasive Computing* 18, 2 (apr 2019), 76–84. https://doi.org/10.1109/MPRV.2019.2907007

[37] Michela Masè, Andreas Werner, Gabriel Putzer, Giovanni Avancini, Marika Falla, Hermann Brugger, Alessandro Micarelli, and Giacomo Strapazzon. 2022. Low Ambient Temperature Exposition Impairs the Accuracy of a Non-invasive Heat-Flux Thermometer. *Frontiers in Physiology* (2022), 283.

[38] Mike A Merrill and Tim Althoff. 2022. Self-supervised Pretraining and Transfer Learning Enable Flu and COVID-19 Predictions in Small Mobile Sensing Datasets. *arXiv preprint arXiv:2205.13607* (2022).

[39] Aaron C Miller, Inder Singh, Erin Koehler, and Philip M Polgreen. 2018. A smartphone-driven thermometer application for real-time population-and individual-level influenza surveillance. *Clinical Infectious Diseases* 67, 3 (2018), 388–397.

[40] Ricardo Morán-Navarro, Javier Courel-Ibáñez, Alejandro Martínez-Cava, Elena Conesa-Ros, Alejandro Sánchez-Pay, Ricardo Mora-Rodriguez, and Jesús G Pallarés. 2019. Validity of skin, oral and tympanic temperatures during exercise in the heat: effects of wind and sweat. *Annals of Biomedical Engineering* 47, 1 (2019), 317–331.

[41] Daniel J Niven, Jonathan E Gaudet, Kevin B Laupland, Kelly J Mrklas, Derek J Roberts, and Henry Thomas Stelfox. 2015. Accuracy of peripheral thermometers for estimating temperature: a systematic review and meta-analysis. *Annals of internal medicine* 163, 10 (2015), 768–777.

[42] Aart Overeem, JC R. Robinson, Hidde Leijnse, Gert-Jan Steeneveld, BK P. Horn, and Remko Uijlenhoet. 2013. Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters* 40, 15 (2013), 4081–4085.

[43] RC Parmar, DR Sahu, SB Bavdekar, et al. 2001. Knowledge, attitude and practices of parents of children with febrile convulsion. *Journal of postgraduate medicine* 47, 1 (2001), 19.

[44] Valentina Pecoraro, Davide Petri, Giorgio Costantino, Alessandro Squizzato, Lorenzo Moja, Gianni Virgili, and Ersilia Lucenteforte. 2021. The diagnostic accuracy of digital, infrared and mercury-in-glass thermometers in measuring body temperature: a systematic review and network meta-analysis. *Internal and emergency medicine* 16, 4 (2021), 1071–1083.

[45] Sujit Shinde, Swapna Agarwal, Dibyanshu Jaiswal, Avik Ghose, Sanjay Kimbahune, and Pravin Pillai. 2020. ThermoTrak: smartphone based real-time fever screening: demo abstract. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 639–640.

[46] K Taylor and L Siliver. 2019. In Emerging Economies, Smartphone Adoption Has Grown More Quickly among Younger Generations. *Pew Research Cetner, Feb* 5 (2019).

[47] Amee Trivedi, Phuthipong Bovornkeeratiroj, Joseph Breda, Prashant Shenoy, Jay Taneja, and David Irwin. [n. d.]. Phone-based ambient temperature sensing using opportunistic crowdsensing and machine learning. *Sustainable Computing: Informatics and Systems* 29 ([n. d.]), 100479.

[48] Srinivasan Venkatramanan, Adam Sadilek, Arindam Fadikar, Christopher L Barrett, Matthew Biggerstaff, Jiangzhuo Chen, Xerxes Dotiwalla, Paul Eastham, Bryant Gipson, Dave Higdon, et al. 2021. Forecasting influenza activity using machine-learned mobility map. *Nature communications* 12, 1 (2021), 1–12.

[49] Paul Webb. 1992. Temperatures of skin, subcutaneous tissue, muscle and core in resting men in cold, comfortable and hot conditions. *European journal of applied physiology and occupational physiology* 64, 5 (1992), 471–476.

[50] Peter Wei, Chenye Yang, and Xiaofan Jiang. 2020. Low-cost multi-person continuous skin temperature sensing system for fever detection. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 705–706.