Quantified Canine: Inferring Dog Personality From Wearables

Lakmal Meegahapola†
Idiap Research Institute & EPFL
Switzerland
lakmal.meegahapola@epfl.ch

Marios Constantinides
Nokia Bell Labs
United Kingdom
marios.constantinides@nokia.com

Zoran Radivojevic
Nokia Bell Labs
United Kingdom
zoran.radivojevic@nokia.com

Hongwei Li
Nokia Bell Labs
United Kingdom
hongwei.3.li@nokia.com

Daniele Quercia
Nokia Bell Labs
United Kingdom
dquercia@acm.org

Michael S. Eggleston
Nokia Bell Labs
USA
michael.eggleston@nokia.com

ABSTRACT

Being able to assess dog personality can be used to, for example, match shelter dogs with future owners, and personalize dog activities. Such an assessment typically relies on experts or psychological scales administered to dog owners, both of which are costly. To tackle that challenge, we built a device called "Patchkeeper" that can be strapped on the pet's chest and measures activity through an accelerometer and a gyroscope. In an in-the-wild deployment involving 12 healthy dogs, we collected 1300 hours of sensor activity data and dog personality test results from two validated questionnaires. By matching these two datasets, we trained ten machine learning classifiers that predicted dog personality from activity data, achieving AUCs in [0.63-0.90], suggesting the value of tracking psychological signals of pets using wearable technologies.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile devices**; **Computer supported cooperative work**; **Empirical studies in ubiquitous and mobile computing**; *User studies.*

KEYWORDS

dog personality, wearables, passive sensing, dog activity recognition, activity level, behavior modeling

ACM Reference Format:
Lakmal Meegahapola, Marios Constantinides, Zoran Radivojevic, Hongwei Li, Daniele Quercia, and Michael S. Eggleston. 2023. Quantified Canine: Inferring Dog Personality From Wearables. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3544548.3581088

†work done during the research internship at Nokia Bell Labs, Cambridge, UK

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581088

1 INTRODUCTION

When it comes to dog adoption, breed may not be the only important factor to consider [2, 8] as humans tend to favor, for example, a pet's looks (e.g., attractiveness [14] based on poses and facial areas [36, 42]) and perceived human-directed sociability [56]. However, according to a study from the Animal Farm Foundation [27], one in every four pets that are chosen based on breed (or looks) end up in shelters and rescues. By contrast, personality traits tend to offer a more comprehensive behavioral description of a dog, which is consistent over time and context [28]. Dog personality has been described in the literature using a variety of traits, including fearfulness, energy levels, aggression, excitability, motivation, and amicability [45, 46, 57].

Not only could dog personality assessment reduce the number of owner-dog mismatches, but it could also put an end to (unfortunate) cases in which dogs get flocked into shelters or destroyed by authorities when expelled from their homes [17]. In fact, a few dog agencies and shelters are already experimenting with the use of dog personality traits for matchmaking dogs with future owners [52, 85]. Further, like humans, dogs also need different levels and types of companionship, activities, and emotional connection depending on their inherent personality traits [10, 52]. It is therefore extremely important to identify a set of activities that 'work' for a dog, and to find the right companion dogs for socializing [6], not least because inadequate socialization may escalate the pet's fear levels and may lead to aggression [82].

Dog personality assessment is typically done through observational assessments by experts or psychological scales administered to the dog's owner [20, 45]. The former is expensive, time-consuming, and requires highly specialized facilities, and the latter is time-consuming, is prone to biases, and requires knowledge of someone who already knows the dog very well [45, 46, 77]. That is why we set out to computationally assess dog personality in everyday settings (compared to highly specialized facilities or laboratory settings) with wearables. In the wearable sensing literature, studies used devices for monitoring dog activity [11, 13, 54, 54, 98], detecting pruritic behaviors (i.e., scratching, head shaking) [35], and tracking breathing patterns [19]. This stream of research recently inspired the fast-growing market of pet wearables [102] with a number of consumer-grade platforms readily available such as FitBark[1] (location, activity, and sleep tracking), PetPace[2] (vital signs

[1]https://www.fitbark.com/
[2]https://petpace.com/

and behavior tracking), and PitPat[3] (activity tracking with gamified social elements).

Similar to how passive sensing of human personality drives the design of personalized apps [49], sensor-based modeling of dog behavior through activity trackers has the potential to benefit both dogs and owners [102]. It has been found to impact owners' motivation to increase their mutual physical activities with their dogs and increased human awareness to animals' needs [47]. However, computational personality assessment techniques for dogs are non-existent. Therefore, we set out to develop and test an automatic way of operationalizing dog personality through passively sensed data from wearables. In so doing, we made three sets of contributions:

- We developed a wearable device called "Patchkeeper", which can be easily strapped on a dog's chest (Section 4). The device is equipped with accelerometer and gyroscope sensors. Since its processing pipeline was initially developed for wearable data obtained from human subjects, we conducted a validation study of our device and the pipeline on dogs, together with a consumer-grade dog activity monitor. We found that our device is capable of determining four activity levels: moderate-vigorous activity with an accuracy of 92%; light and sedentary activity with an accuracy of 96%; and sleep with an accuracy of 98%.
- We launched a data collection campaign to recruit dog owners whose pets participated in a one-week study. The campaign was launched on four social media platforms (i.e., posts were made on Twitter, Facebook, Instagram, and NextDoor) and was also spread via word of mouth (Section 5), resulting in a total of 22 dogs being successfully recruited and monitored for one week (i.e., the entire period of study). Dog owners answered two validated questionnaires (the Dog Personality Questionnaire (DPQ) [45] and the Refined Monash Canine Personality Questionnaire (MCPQ-R) [58]), and provided self-reports about their dog's activities (e.g., images were taken when walking the dog). Using the passively sensed data, we developed a data processing pipeline and extracted two types of features: (a) activity-level features (e.g., % of sleep in the morning, % of sedentary activity in the afternoon) and (b) statistical features (e.g., acceleration histogram) (Section 6). We statistically analyzed the extracted features along with the self-reports from the two questionnaires and found that both types of features could discriminate dog personality traits (e.g., high or low fearfulness), with features capturing dog activity between 6am and 12pm (morning) being more informative for personality trait inferences than features capturing activity in the rest of the day. This is expected as most dogs will be the most active and full of energy in the mornings after a dedicated sleep, and that was reflected in the signal captured from our device's sensors.
- We set up an inference task to predict dog personality traits using both activity-level and statistical features (Section 7). Our models achieved AUC scores in the range of 0.63-0.90 with a time-window-based setup (i.e., using the same features computed at different times of the day) (Section 8). Interestingly, statistical features (e.g., acceleration histogram) were more informative than activity-level features (e.g., sedentary); yet, despite explaining more variance in personality traits, the former set of features is less interpretable than the latter one, opening up the need for Explainable AI in this kind of wearables too. When it comes to the usability of dog monitoring wearables, dog owners had split opinions about battery life (some found a day of battery life to be sufficient, while others expressed the opposite). For the development of future monitoring wearables, the majority of dog owners stressed their immense value, echoing a dog owner's statement: "as dogs cannot speak, a device that allows my dog to 'speak' and 'express her feelings' is worth everything".

[3]https://www.pitpat.com/

2 BACKGROUND AND RELATED WORK

Next, we surveyed various lines of research that our work draws upon, and grouped them into four main areas: *i)* dog personality research; *ii)* psychological scales for assessing dog personality; *iii)* monitoring dog activity with wearable sensing; and *iv)* activity levels and dog personality.

2.1 Dog Personality Research

Dogs have personality [45, 57, 58], which refers to a set of dog behaviors and traits that are consistent over time and context [28, 31]. These traits stem from the Five-Factor Model of personality, a.k.a. the Big-Five Traits [18]. As with personality, temperament is also being used in literature to describe both human and animal behavior. Researchers on animals and human infants tend to use the term temperament, while those studying human children and adults tend to use the term personality, with the two terms often being used interchangeably [60]. On the one hand, temperament has been defined as the inherited, early appearing tendencies that continue throughout life and serve as the foundation for personality [29, 46]; a definition that has not been widely adopted by animal researchers [30]. On the other hand, personality psychologists often study phenomena including temperament and character traits, attitudes, physical and bodily states, moods, and life stories [44]. Therefore, a broad definition includes characteristics of individuals that describe and account for consistent patterns of feeling, thinking, and behaving [73]. As the distinction between temperament and personality has not been maintained consistently in the literature, we echo the statement by Jones and Gosling [45, 46], that is, *the term "temperament" is used whenever possible while the term "personality" is more appropriate when, for example, referring to work that explicitly discusses personality research. Hence, we use the term personality throughout the paper.*

In the scientific literature, Elliott Humphrey first hinted at the idea of dogs having personality in 1934 [78]. He described German Shepherd dogs with the traits of jealousy, apport, wildness-tameness, affection, initiative, attentiveness, curiosity, alertness, fighting and protection instincts, willingness to bite humans, confidence, self-right, energy, willingness, and intelligence. Seventy years later, by reviewing more than 50 scientific articles on dog personality, Jones and Gosling [46] found several inconsistencies, and proposed the first five-factor dog personality instrument, covering the dimensions of reactivity, fearfulness, responsiveness to training, submissiveness, and aggression. Building on Jones and Gosling's seminal work, researchers have incrementally added other dimensions such as calmness, boldness, trainability, and sociability [53]; extraversion, neuroticism, self-assuredness (motivation), training focus, and amicability [58]; stranger-directed sociability, activity, aggressiveness, and trainability [65]; and playfulness, chase-proneness, curiosity/fearlessness, sociability, and aggressiveness [86]. Researchers, however, have split views when it comes to predictors of dog personality. Some studies found that different breeds have similar personalities [58, 80, 86], while others reported the lack of evidence for it [65, 84]. Two other attributes linked to personality traits are whether the dog is neutered or not (neutering is a surgical procedure to prevent a dog from reproducing) and its sex. Kubinyi et al. [53] found that not neutered dogs are more calm, while Lofgren et al. [59] found that neutered female dogs were less excitable and sought lower levels of attention. There is also evidence that older dogs are more calm [53] with lower amounts of fear [59] compared to their younger counterparts. Hence, as mentioned above, even though not conclusive, there is evidence that static attributes such as sex, age, and neutering could be associated with dog personality [53, 59].

2.2 Psychological Scales for Assessing Dog Personality

While there are many dog personality measurement questionnaires [74, 79], two widely established and validated psychological scales are: *a)* the Dog Personality Questionnaire (DPQ) [45], and *b)* the Refined Monash Canine Personality Questionnaire (MCPQ-R) [58]. Next, we explain each scale.

- **DPQ:** Building on the work of Jones and Gosling [46], the development of this scale aimed at reducing the time and resources (i.e., trained assessors, money, facilities) for dog personality assessment. Amanda Jones started from 1200 dog descriptors (i.e., statements describing dog behavior) identified in the literature and narrowed them down to 360 statements [45]. Then, in two studies with over 6000 participants, they narrowed these statements down to 75 items, grouped in five factors of *Fearfulness*, *Aggression towards People*, *Excitability*, *Responsiveness to Training*, and *Aggression towards Animals*. Scores for these traits can be derived using a list of statements marked by the dog owner on a Likert scale from 1 to 7 (1: disagree strongly; 7: agree strongly).
- **MCPQ-R:** This is the refined version of the original MCPQ questionnaire [57]. The original questionnaire was developed using an adjective-based technique similar to the Big-Five Model of personality [43]. Ley et al. [58] revised the original MCPQ in a study with more than 450 participants. This led to the development of MCPQ-R, which consists of five factors: *Extraversion* (perceived energy level of the dog), *Motivation* (perceived persistence in the face of distractions—e.g., begging for food, finding a particular toy), *Training Focus* (perceived trainability of the dog), *Amicability* (perceived tolerance of the dog while being around humans and animals), and *Neuroticism* (perceived nervous or cautious behavior of the dog). To assess these traits, dog owners rate 26 words (e.g., friendly, obedient, hyperactive) that describe their dog's personality by marking each word with the appropriate number from 1 to 6 (1 = really does not describe my dog; 6 = really describes my dog).

Even though the two scales come with different constructs, a fair amount of convergence has been observed [74] between neuroticism (MCPQ-R) and fearfulness (DPQ); excitability (DPQ) and extraversion (MCPQ-R); responsiveness to training (DPQ) and training focus (MCPQ-R). While other widely used questionnaires such as the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) [40] were developed, recent research suggested that it is not suitable for general research use because it was designed to identify specific dog behavioral problems [20]. Hence, in the current study, we focused on DPQ and MCPQ-R questionnaires that capture a total of ten personality traits (factors).

2.3 Dog Monitoring with Wearable Sensing

Dog tracking and activity detection have gained much popularity due to advancements in sensor technology [41], which led to a number of commercial dog monitoring products (e.g., FitBark, PetPace, PitPat). However, tying wearable sensing to behavioral tests (like dog personality in our case) is just starting to gain traction. In Animal-Computer Interaction research, prior studies focused on systems that facilitate better communication and interaction between dogs and owners [38] as well as among dogs [37]. Personality and dog behavior were also studied as part of certain games such as the spin-the-bottle [20], concluding that dogs' preferences for human involvement were likely attributed to subtle differences in personality traits or prior training experiences.

Brugarolas et al. [9] developed a non-invasive wearable sensor system for measuring dogs' vital signs using electrocardiogram (ECG), photoplethysmogram (PPG), and inertial measurement units (IMU). In a longitudinal study of monitoring puppies' cardiac changes, Foster et al. [26] developed machine learning models for predicting puppies' Behavior Checklist (BCL) scores (including changes in energy and smoothness of movement, vocalization, tongue flicking, use of coping strategies, body language, and changes in responsiveness to the handler), achieving up to 90% of accuracy. Weiss et al. [98] developed WagTag that infers three dog activity levels (i.e., walk, run, and minimal), and concluded that personal models for predicting activity levels are better than universal models. Ladha et al. [54] also demonstrated that 17 dog activities (e.g., barking, running, chewing, digging) can be inferred with an accuracy of 70% from a collar-worn wearable with accelerometers. More recently, Chambers et al. [13] used deep-learning models to infer dog activities with a collar-worn accelerometer, and showed that activities such as eating and drinking could be inferred with high accuracy, while behaviors such as licking, petting, rubbing, and sniffing were harder to identify. Beyond activity tracking, Griffies et al. [35] used wearables to detect pruritic behaviors (i.e., scratching, head shaking). In a laboratory study with over 360 dogs, they showed that algorithms could be trained to infer head shaking and scratching with sensitivities over 70% and specificities over 90%. Wearable devices have also been used to monitor dog breathing patterns with reasonable accuracies [19].

2.4 Dog Activity Levels and Personality Traits

Prior work in animal-computer interaction and canine behavior has highlighted certain relationships between personality traits and activity levels. For example, previous studies found that more extroverted dogs showed higher activity levels in the park [12], higher energy levels [32], with significantly greater proportions of time spent with other dogs. Amicable dogs showed frequent behaviors indicative of play (high activity level), while neurotic dogs showed higher frequencies of hunched posture (low activity level) [12]. Hence, extraversion, amicability, and neuroticism (traits that come from MCPQ-R) can be directly linked to activity levels. Further, even though not directly studied, prior studies linked psychological aspects such as fearfulness and aggression (corresponding to the three traits of DPQ— fearfulness, aggression towards people, and aggression towards animals) to activity levels. For example, in domestic dogs, it has been found that a higher degree of impulsivity correlates with high activity levels [92], poor attention span [88], and human-directed aggression [72, 75]. Further, previous studies linked activity levels to negative emotions and stress [4, 47, 75], which, in turn, can be seen as the roots of fearfulness and aggression [4]. Moreover, neuroticism has been directly linked to activity levels in some studies [12], but it has also been observed to be converging with fearfulness according to other studies [74], hence providing evidence on how activity-levels could be indirectly informative of fearfulness. Studies have also found that excessively high or low activity levels are predictive of successful dog training (i.e., trainability and certain levels of fearfulness [97]; traits captured from DPQ and MCPQ-R).

In summary, previous wearable sensing literature explored aspects such as monitoring dog activity, detecting pruritic behavior, and tracking breathing patterns. While previous literature explored a few aspects concerning the relationship between dog activity and personality traits, this relationship still represents an underexplored area. Our study aims to partly fill this gap by exploring the relationship between ten personality traits captured from two canine personality questionnaires and dog activity.

3 RESEARCH QUESTIONS

We set out to explore whether dog personality can be automatically inferred from wearable data in everyday settings by answering three questions:

$\mathbf{RQ_1}$: Which dog activity-level features and statistical ones can be extracted from wearable data?

$\mathbf{RQ_2}$: Which dog activity-level features and statistical ones are associated with dog personality?

$\mathbf{RQ_3}$: To what extent activity-level, statistical, and demographic features are predictive of dog personality?

4 PATCHKEEPER

Patchkeeper (Figure 1a) is a wearable device developed at Nokia Bell Labs for behavioral monitoring of both humans and animals. It contains a photoplethysmography (PPG) sensor, an electrocardiogram (ECG) sensor, an accelerometer, a gyroscope, and a microphone. In the current study, only the inertial measurement unit (IMU) sensors (i.e., accelerometer and gyroscope) were used, and PPG, ECG, and microphone were not used due to the dog hair and privacy concerns (more details in Section 5.1.1). The IMU sensor is a BMI160 from Bosch Sensortec[4]. It is a small, low-power, low-noise 16-bit chip designed for mobile applications. It provides highly accurate gyroscope and accelerometer data in real time. The IMU's sampling rate was set to 50 samples per second. This sampling rate allowed for striking the right balance between obtaining reasonably fine-grained data for our analysis and storage capacity requirements. The microcontroller unit (MCU) is an nRF52840 from Nordic Semiconductor[5], which contains a 64 MHz Cortex-M4 processor with floating point unit (FPU). All data was saved in a micro-SD card on the printed circuit board (PCB).

The device dimensions are 76x52x15mm with a weight of 56 grams. It contains a 400mAh lithium polymer battery, which can last more than 24 hours while continuously recording data. The battery takes around two hours to be fully charged and comes with a USB-C charging port for hassle-free charging with any commercially available charger. With a two-hour daily charge, the device runs continuously without any loss of data. The device has a switch with ON and OFF sides marked with red and green colors. For better user experience, we included different lights on the device (Figure 1a): *(i)* a green light flashing every 10 seconds indicates that the device is ON, it is working properly, and data is being recorded; *(ii)* a static red light indicates that the device is fully charged, and *(iii)* a flashing red light indicates an issue with the device or the memory card.

5 ANIMAL STUDY

Having developed our custom-made wearable device to collect dog activity data, we conducted a one-week in-the-wild study to understand the link between dog behavior and personality.

5.1 Materials and Apparatus

Each dog owner received a package, fitted in a medium-sized letter envelope, weighing approximately 500 grams. The package contained: a Patchkeeper device, a charging cable, three black elastic straps, a consent form, an information sheet, questionnaires (i.e., DPQ, MCPQ-R, and a post-study questionnaire), and a pre-paid return envelope. Upon completion of the study, the owner shipped back the package using the pre-paid return package.

5.1.1 Patchkeeper and Elastic Straps. As the device can be used on both human and animal subjects (Section 4), and given the requirement of continuous monitoring for one week, we decided to deactivate the ECG and PPG sensors, and the audio microphone. ECG and PPG were disabled for two reasons: first, they relied on skin conductance, which is made difficult by dog hair; and second, they required additional straps, which would place additional effort on the owners, making it more likely for them to drop out. Audio was also deactivated due to privacy reasons as the device would otherwise continuously capture audio throughout the day. It would be extremely awkward to listen to intimate moments or any audio conversation that creeps into the device due to the pet's movements. To ensure that the device would fit various dog sizes, we used an adjustable elastic band that can be strapped to the pet's chest (Figure 1b). These are off-the-shelf straps that can be found on Amazon and are comfortable to wear. The device can be simply attached to the strap using a sticky patch. We also considered alternative

[4]https://www.bosch-sensortec.com/products/motion-sensors/imus/bmi160/

[5]https://www.nordicsemi.com/products/nrf52840

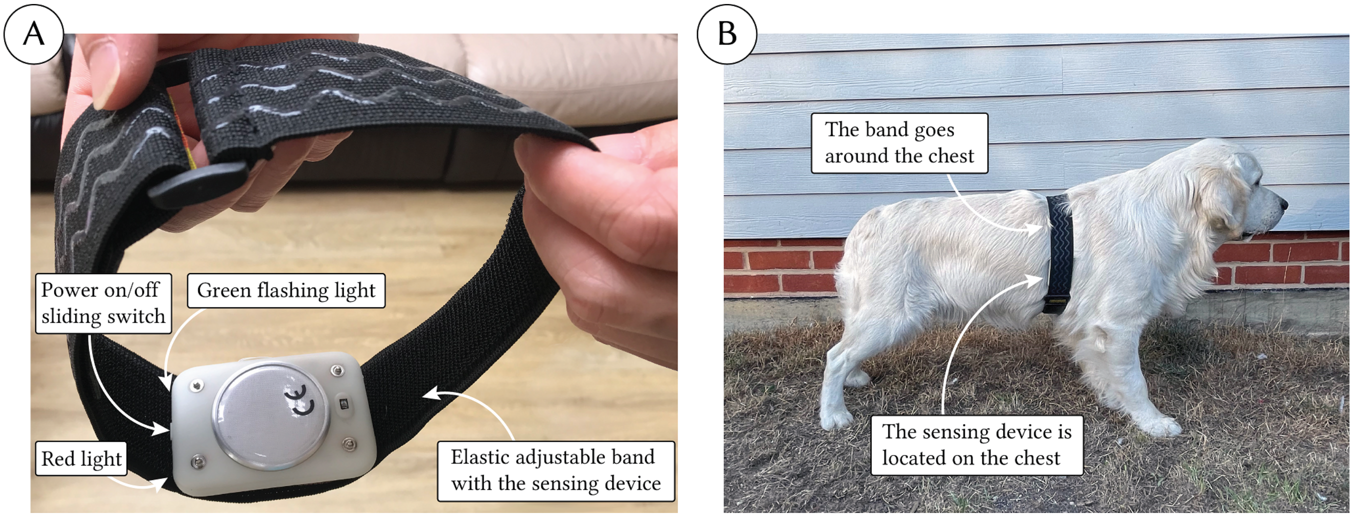


Figure 1: ***A:* The Patchkeeper device is attached to an elastic adjustable band. The device comes with two lights: a *green light* indicating whether the device is ON; and a *red light* indicating whether the device is in charging mode. *B:* The band can be strapped on the dog's chest.**

areas (e.g., neck) to place the device, weighing various aspects. First, some breeds have a more pronounced dewlap (loose, saggy skin around the neck/throat) than others, whilst the chest is generally not affected in such a way. Second, it has been found that the skin near the axilla (armpit) and ventral abdomen (lower chest/thorax, top of the belly) is significantly thinner than that in the dorsal (top of the dog) areas [87]. Third, double-coated breeds (e.g., the golden retriever, Samoyed, and German shepherd included in this study) have coarse guard hairs and dense undercoats, with this being particularly pronounced in the dorsal areas but less so near the axilla and ventral abdomen areas. Taking all these aspects into consideration, the chest area (behind the forelegs) has the benefits of thinner skin whilst removing breed-to-breed variation in our sample (previous work also favored the chest area [26]). To ensure that the position of the device did not impact the results, we intentionally used wider straps with 5cm wide rubbered features and strong tensioning force to make the device intimately connected to the fixed location at the pet's chest. During pilot studies in the design of the device, we estimated an approximate sensor dislocation of ±2cm, which was sufficient to guarantee fixed sensor location over a long period of use.

5.1.2 Questionnaires. Dog owners answered two types of questionnaires. A Pre-Study Questionnaire (Q1) and a Post-Study Questionnaire (Q2). Q1 was completed before the study and had two sections. The first section captured demographic information of the owner (i.e., age, sexual identity, occupation status, and ethnicity), followed up with the Personality Inventory (TIPI) [33], which is a 10-item measure of the Big Five (or Five-Factor Model) dimensions. The second section captured basic information about the dog (i.e., the dog's age, breed, sexual identity, weight, typical activity levels, disease conditions, and whether it is neutered or not). Q2 was completed after the study and had two sections as well. The first section captured user experience and dog owners' perceived utility of wearable platforms for dog monitoring. The user experience of the Patchkeeper was captured by a Likert scale of 1 to 7 (1 = very bad, 7 very good) and corresponding feedback. In a similar vein, we captured the perceived utility of commercial wearable pet monitoring devices in general by a Likert scale of 1-7 (1 = not very important; 7 = very important) and corresponding feedback. Additionally, we asked dog owners to rate on a scale of 1-7 (1 = strongly not preferred, 7 = strong preferred) their likelihood of adopting a mobile app that uses Patchkeeper's data for dog monitoring. We provided sample options including: monitoring activity types, identifying when dogs are not in a healthy state, finding a community of dogs with a similar personality, or monitoring the mood and stress of dogs. The second section of Q2 asked owners to complete the Dog Personality Questionnaire (DPQ) [45] and the Refined Monash Canine Personality Questionnaire (MCPQ-R) [58].

5.1.3 Information Sheet and Consent Form. The information sheet described the study protocol (Section 5.2). The consent form highlighted two aspects: confidentiality and voluntary participation. In terms of confidentiality, the form explained that all data would be kept confidential except in cases where the researchers were legally obligated to report specific incidents (e.g., dog abuse). The collected phone numbers and email addresses will not be used in any scientific output, and that confidentiality will be preserved by: *a)* assigning code numbers for dog owners in all research documents; and *b)* keeping notes, data, and any other dog owner identifiers in a password-protected hard drive, securely stored at the facilities of the Nokia Bell Labs. In terms of voluntary participation, the form explained that a signature was required to participate. Additionally, withdrawal from the study was allowed at any time and without giving a reason, even after signing the consent form. Upon withdrawal, all data will be deleted.

5.1.4 Pre-Paid Letter Cover and WhatsApp Hotline. To ease dog owners participation, we included in our package a pre-paid return

Table 1: **Overview of dog demographics.**

Dog ID	Breed	Sex	Weight	Neutered?	Birth Year
#1	Golden Retriever	Female	30 kg	Yes	2018
#2	Golden Retriever	Male	35 kg	No	2021
#3	Poodle (Toy)	Male	8 kg	No	2020
#4	Dalmadoodle - 75% Poodle, 25% Dalmatian	Female	13 kg	Yes	2020
#5	Golden Retriever	Male	40 kg	No	2021
#6	Working English Setter	Male	29 kg	Yes	2017
#7	Boxer	Female	25 kg	No	2020
#8	Samoyed	Male	25 kg	No	2018
#9	Cockapoo	Female	10 kg	Yes	2011
#10	Working English Setter	Female	31 kg	No	2016
#11	Mixed	Male	15 kg	Yes	2021
#12	Cavalier King Charles Spaniel	Female	8.5 kg	Yes	2019

package. Upon completion, they placed all materials received into the return package and posted it. To have effective communication with the dog owners throughout the study, we used a dedicated WhatsApp number as a hotline. This number was used by the first author to deal with matters related to the study (e.g., unable to place the strap or charge the device).

5.2 Study Protocol

The study protocol has three periods: pre-study, study period, and post-study (Figure 2).

5.2.1 Pre-Study. Once the dog owners received the package, they familiarized themselves with the device and answered the pre-study questionnaire. During that period, they were encouraged to ask questions via the hotline, and were instructed to fully charge the Patchkeeper and send a picture of the dog with the device turned on, after a full charge every day (this was a preemptive measure to ensure compliance, but, at the same time, to guarantee data quality).

5.2.2 Study Period. During the seven-day period, the device captured sensor data between 12am and 6pm (continuously for 18 hours), and it could be charged for two hours between 6pm and 12am. Enforcing the same charging schedule across all dog owners enabled us to obtain comparable data across dogs. Of course, this comes with the caveat that we might not have 2-3 hours of data between evening hours; a drawback that we were willing to accept to ensure high-quality data during other time slots. In summary, each evening, the dog owners would remove the strap from the dog, turn the device off, fully charge it, turn it back on, and put it back on the dog. Afterward, they would send a message with a picture of the dog wearing the device via the hotline. In the morning, they would again be asked to check that the device was working and was correctly positioned around the dog's chest. During all other time periods, no interaction was required from dog owners as the device would automatically capture all data. Furthermore, we encouraged the dog owners to voluntarily send us in-situ self-reports (in the form of images or short video clips) of various dog activities throughout the day.

5.2.3 Post-Study. During that period, dog owners answered the post-study questionnaire. They placed all materials and apparatus in the pre-paid package and shipped it back to the return address. Upon successful completion of the study, dog owners received a $25 Amazon gift voucher and a report summarizing their dog activity profile over the seven days of the study.

5.3 Recruitment

Recruitment for in-the-wild human studies is typically difficult [24], and so it is for animal studies. We employed two techniques that were proven (un)successful to varying degrees.

- Social media and local communities (Twitter, Facebook, Instagram, and NextDoor): Twitter and Facebook are used to advertise scientific studies [83, 100]—both channels were not very successful in this study. Instead, Instagram posts on profiles dedicated to dogs with 1000s of followers were successful to some extent. Finally, Nextdoor[6], a social media site for local communities, was the most successful recruitment strategy (40% of the dogs were recruited through it). A banner of the study was also shared within the communities of Cambridge Dog Meetup.
- Word of Mouth: One researcher from Nokia Bell Labs, who is not part of conducting the study, participated in the study with his dog. He spoke to his neighbors about the study, who also signed up. Shortly after, this created a snowball effect (30% of the dogs were recruited through word of mouth).

Having a variety of recruitment techniques, we were able to reach out to 31 dog owners in Cambridge, United Kingdom. Of these, 22 signed up for the study and received the package. 10 of them withdrew during the study for various reasons: high temperature, including a heatwave, making it difficult for the dog to wear the strap continuously (2/10), owners going away for summer holidays (2/10), strap not holding to the body of the dog due to its curvy shape (1/10), dogs not being in healthy conditions (i.e., leg injury after a run, wound on the neck, bug bites) during the time of the experiment (3/10), and dogs not appearing to feel happy about wearing the strap (2/10). This left us with 12 healthy dogs that successfully completed the study. Note that these 12 dogs were all healthy (as reported by their owners), and every morning the first author checked with the owners whether any of the dogs displayed peculiar behavior (e.g., snagging on objects, appearing

[6]https://nextdoor.com/

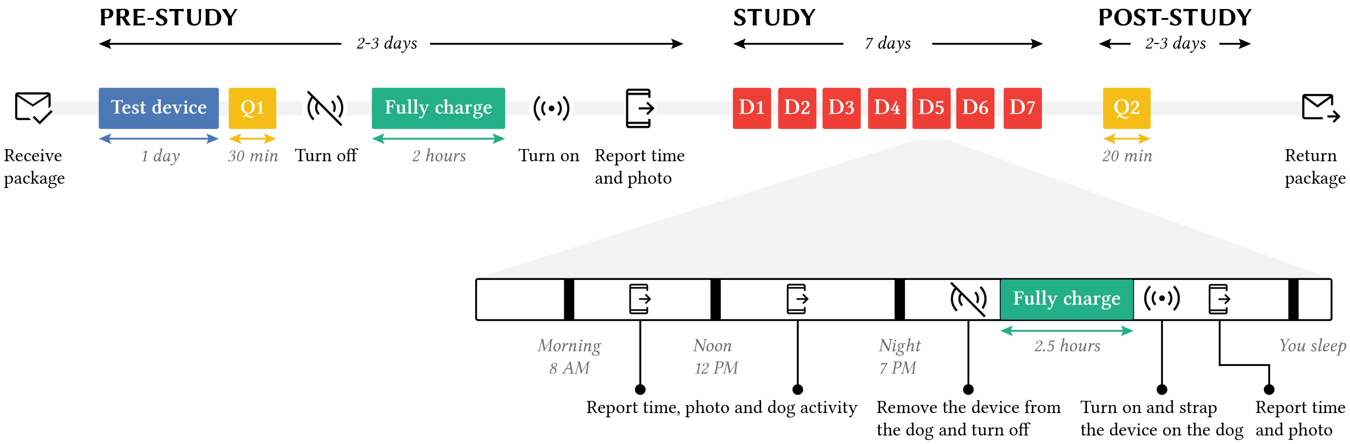


Figure 2: **Our study protocol has three phases. In the *pre-study* phase, dog owners received the study package, including the PatchKeeper device and questionnaires; in the *study* phase, data collection for seven days took place; and, in the *post-study* phase, dog owners returned the package and answered a follow-up survey about their experience using the device.**

to feel uncomfortable) due to the wearable. No such incident was reported. However, we had an incident wherein a dog jumped into a body of water, destroying the device. This dog continued the study later with a replacement device. The recruitment took place during the summer period, with starting dates ranging from July to August. The study was approved by Nokia Bell Labs, and the study protocol stated that the collected data will be analyzed for research purposes only. In accordance to GDPR, no researcher involved in the study could have tracked the identities of the dog owners after the end of the study, and all responses were analyzed after anonymization at an aggregated level.

6 DATASET

Having successfully deployed Patchkeeper in an in-the-wild study and collected more than 1300 hours of accelerometer and gyroscope data, we then applied a processing pipeline to that data.

6.1 Data Processing Pipeline

6.1.1 Activity-Level features. This set of features describes dog behaviors derived from accelerometer data, and is interpretable. To extract these features, we used a state-of-the-art data processing pipeline to convert the triaxial data to acceleration [22, 25, 101]. The processing included four steps: *a)* ten-second samples from static sections (no movement) of accelerometer data were obtained to optimize the gain and offset for each of the X, Y, and Z axes to fit a unit gravity sphere [101]; *b)* data were re-sampled at 100Hz using linear interpolation, and acceleration was calculated using the euclidean norm of X, Y, and Z axis values; *c)* a fourth-order butterworth filter was used to remove noise, and *d)* one gravity (1G) unit was removed from the data, and the remaining negative values were truncated at zero.

Next, using non-overlapping time windows of 60 seconds, 126 time and frequency domain features such as mean, standard deviation, median, minimum, maximum, 25^{th} and 75^{th} percentiles of vector magnitude, kurtosis, and skewness were generated [23, 90, 101]. Using these features, we used a pre-trained model based on Hidden Markov Models and Balanced Random Forests [91] to classify acceleration into four different activity levels: *sleep*, *sedentary*, *light*, and *moderate-vigorous.* These activity levels are in line with prior studies on dog activity levels [66, 70, 98].

As the data processing pipeline was initially developed for wrist wearables worn by humans, we conducted a validation step to ensure transferability to animals. To do that, we used a consumer-grade dog activity monitor called PitPat[7] on two dogs (dog#3 and dog#8) for three days. These two dogs also took part in the larger in-the-wild study. In total, we collected over 120 hours of sensor data from both devices, and a total of 83 self-reports (e.g., the dog is sleeping, running) from dog owners. A comparison of our 24-hour data processing pipeline and PitPat's output is shown in Figure 3. In terms of ground truth obtained from PitPat (in total, we analyzed 100 data points), our model performed with an accuracy of: 98% in detecting sleeping (sections where PitPat showed no activity); 92% in detecting high intense moderate-vigorous activities (sections where PitPat showed a peak in activity levels), and 96% in detecting sedentary or light activity levels (sections where PitPat showed a medium level of activities). In terms of self-reported ground truth (including pictures), our model was 91% accurate in determining the 83 activity levels provided by dog owners. This answered our **RQ_1**, allowing us to conclude that activity levels can be extracted with accuracies over 90%.

Having established the reliability of our data processing pipeline, we first obtained how long a particular dog had been engaging in activities at different levels (i.e., percentage of time spent in sleep, sedentary, light, and moderate-vigorous activity levels), resulting in four features. We then used the acceleration time series to extract statistical features such as its minimum, maximum, mean, median, and standard deviation, resulting in five features. For simplicity, we call these nine features *activity-level* features throughout the paper.

6.1.2 Statistical Features. This set of features was derived from complex associations in the time series of both accelerometer (x,y,z)

[7] https://www.pitpat.com/

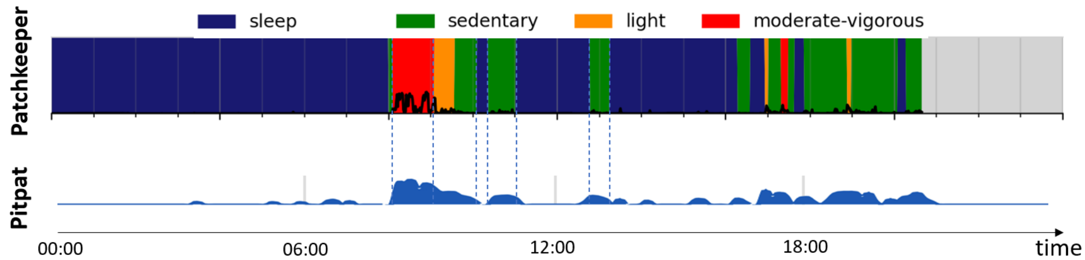


Figure 3: **Comparison of dog#3 is monitored by activity levels generated from Patchkeeper (top) and Pitpat (bottom) for 24 hours.**

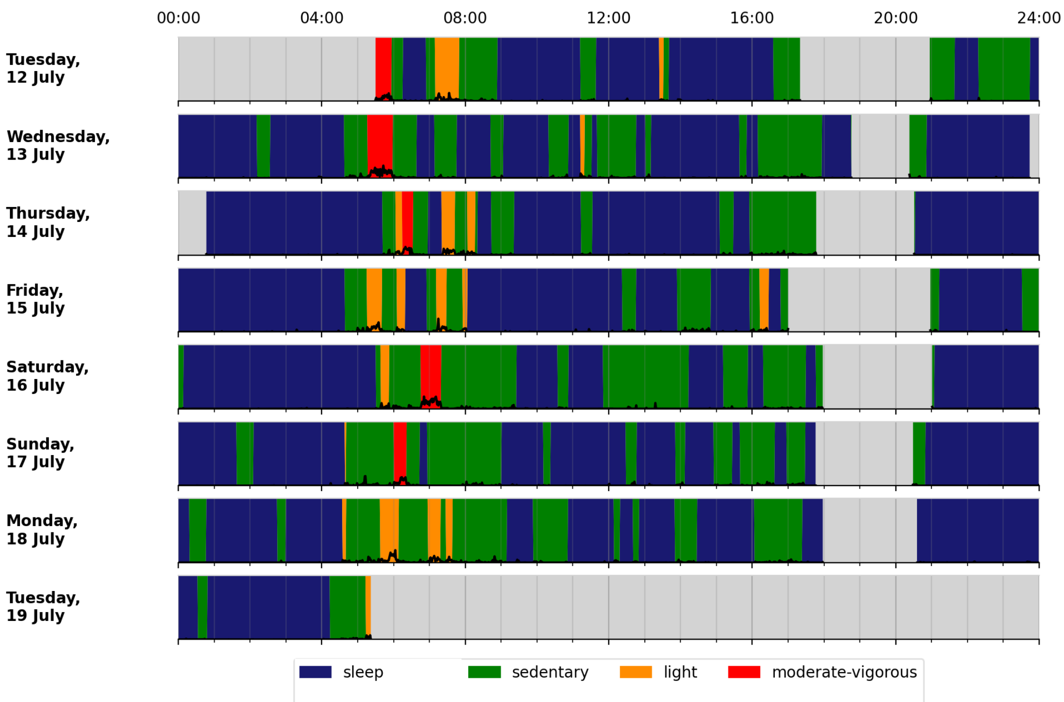
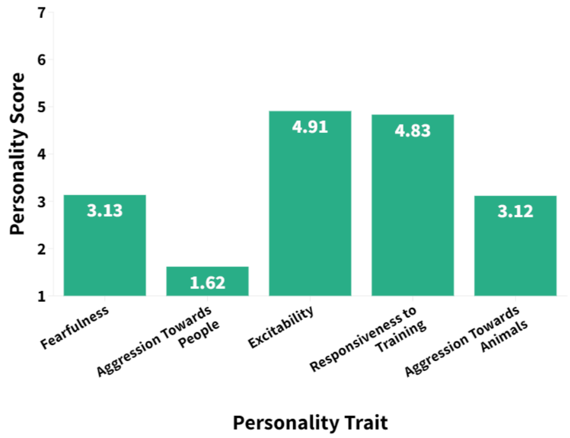


Figure 4: **Example of four types of activity levels of dog#1 (i.e., sleep, sedentary, light, and moderate-vigorous) generated from Patchkeeper.**

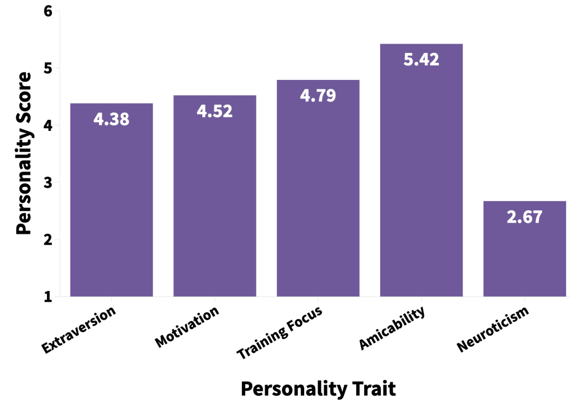
and gyroscope (x,y,z) and, as such, is less interpretable compared to the activity-level features but computationally less expensive to obtain. To extract these statistical features, we used the tsfel library [3]. The library allowed us to extract 56 features (e.g., min, max, std, mean, median, kurtosis, skewness, absolute energy, zero crossing rate, histogram, and empirical cumulative distribution function) that describe temporal and statistical aspects of the time series nature of the data[8].

6.1.3 Unit of Analysis. A typical way of capturing temporal dynamics in HCI and UbiComp studies is to use time windows at different times of day when calculating features [16, 67, 69, 93–95]. A large time window of eight hours, dividing the day into three periods, has been previously used in dog studies, and it has been found, for example, that studying night sleep separately from day sleep provided more meaningful insights about sleeping patterns [5, 81] than studying sleeping during the whole day. Drawing from this prior line of work, we resorted to three time windows for our analysis: *a)* night (N): time from 12am to 5.59am; *b)* morning (M): time from 6am to 11.59am.; and *c)* afternoon (A): time from 12pm to 5.59pm. For example, at night, a dog could be showing activity levels as 60% sleeping, 20% sedentary, 15% light, and 5% moderate-vigorous. Hence, for each time window, we extracted a total of 65 features, including the nine activity-level and the 56 statistical features.

[8]https://tsfel.readthedocs.io/en/latest/descriptions/feature_list.html



(a) DPQ



(b) MCPQ-R

Figure 5: Average personality trait scores for five personality factors measured through DPQ (in a Likert scale 1-7) and MCPQ-R (in a Likert scale 1-6), respectively. Scores for MCPQ-R were higher (except Neuroticism), whereas scores for DPQ were more spread out.

6.2 Descriptive Statistics of Personality Traits and Activity Levels

The distributions of personality traits are shown in Figure 5, and a summary of statistics of the recruited dogs is in Table 1. Recruited dogs were over one year old, with a mean age of three years and ten months, 65% of them were female dogs and were all medium to small dogs. In terms of dog personalities, the DPQ factors: *Fearfulness*, *Aggression Towards People*, and *Aggression Towards Animals* had average scores below 3.2, whereas the *Excitability* and *Responsiveness to Training* had high average scores above 4.8. For MCPQ-R, *Neuroticism* had an average score of 2.67, while the other four factors had average scores on or above 4.38. Overall, the mean percentages across the personality dimensions were comparable to previous studies [12, 58]. Further, Figure 6 shows the average activity level across all dogs as a percentage of total time (y-axis) for different time periods of the day (x-axis). The night was predominantly spent sleeping (63.3%), whereas morning was predominantly spent at other types of activity such as sedentary (36.5%), light (13.5%), or moderate-vigorous (9.3%).

7 METHODOLOGY

Using the extracted features and self-reported personality (Section 6), we set out to understand which features are associated with dog personality (**RQ$_2$**), and, to what extent these features are predictive of personality (**RQ$_3$**). In so doing, we defined our dependent variables, conducted a series of statistical analyses, and developed machine learning classifiers to predict dog personality, which we describe next.

7.1 Dependent Variables

The ten personality traits of both DPQ and MCPQ-R (five each) served as our dependent variables for both statistical analyses and classification tasks. We binarized each personality trait (i.e., whether a dog scored high or low in a given trait —for example, high or low in fearfulness) using the median value across all dogs in our dataset. In other words, we computed the median value across all dogs for each trait, and binarized each trait to be either high or low. The choice of binary traits was reinforced by previous literature on inferring human personality from mobile data [49, 89].

7.2 Statistical Analyses

These analyses allowed us to identify statistically significant features that help discriminate between high and low personality scores for each trait (Table 2). We report the top five features for each personality trait with: *a)* the highest t-statistic[9] [51], and *b)* the highest Cohen's-d[10] [76]. As a rule of thumb, a Cohen's-d of 0.2 illustrates a small effect size, 0.5 a medium effect size, while 0.8 a large effect size [50]. Results are presented in Section 8.1.

7.3 Classification and Cross-Validation Methods

In the next set of experiments, we used Python with Keras [15] and scikit-learn [71] frameworks. For dimensionality reduction, we used principle component analysis (PCA) [1] and retained features with a variance of 95% (leaving us with 3-5 features to train models, depending on the set of features used—we discuss these features in a subsequent section). All experiments were done with the leave-k-dogs-out strategy (in a similar manner to the leave-one/K-out setting, which is typical cross-validation for human subjects [64]) in which data in training and testing splits do not come from the same dog. Hence, this is subject-independent [61]. We conducted all experiments with five iterations and K = 4 — that is, in each experiment, four dogs were left out for the testing set such that this set contains instances from both high and low scores of each personality trait. This allowed us to measure the mean and standard deviation of the models' performance across iterations. Given the small sample size, the choice of four dogs was a reasonable

[9] p-values [34] are marked with an asterisk (*) after bonferroni correction [99].
[10] the 95% confidence intervals [55] overlapping with zero are marked with an asterisk (*).

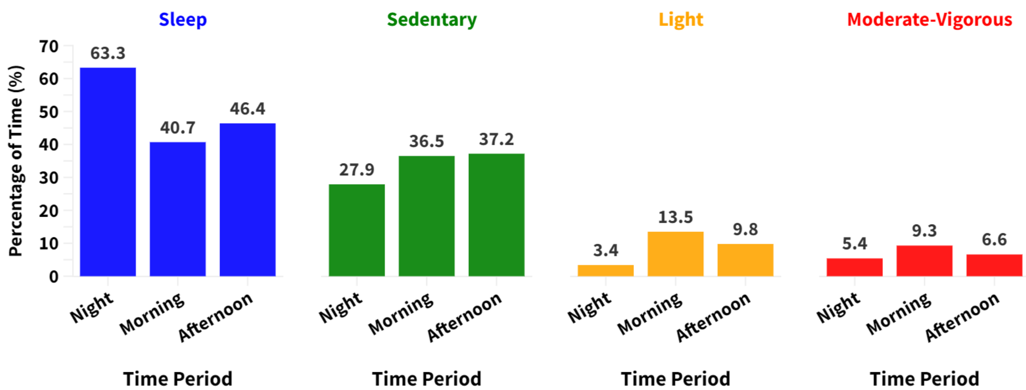


Figure 6: **Percentage of time spent by all dogs on average in each of the four activity levels for sleep, sedentary, light, moderate-vigorous at three different times of day (Night, Morning, Afternoon). Dogs slept more at night, engaged in sedentary activity in the morning and the afternoon, and engaged in physical activity (light or moderate-vigorous) during the morning.**

one. As for performance metric, we chose the area under the receiver operating characteristic curve (AUC), which is a holistic measure assessing how well a model performs for both classes (i.e., distinguishing high and low traits) [7]. In total, we set up three experiments and tested four types of models (S1) using a combination of features (S2 and S3).

- S1—Model Types: Given the small dataset size (typical in animal studies [38, 54, 98]), we used four types of classifiers: *(a)* Support Vector Machines (SVM) [68], *(b)* Light Gradient Boosting Machine (L-GBM) [48], *(c)* Naive Bayes (NB) [96], and *(d)* Random Forest (RF) [21].
- S2—Feature Types: As previously mentioned (Section 6.1), we generated two main types of features from the inertial data: activity-level features (ACT) and statistical features (STAT). The former set of features is typically more interpretable but costly to obtain due to the power processing needed to generate the features, while the latter set of features is not computationally expensive but it is less interpretable. In addition, we used demographic attributes (DEM) such as sex of the dog, weight, age, neutered, and training rating as input to the model because prior work suggested connections between these attributes and dog personality (Section 2). As the dog owner's personality has been previously linked to the dog's personality and well-being [39, 53], we also used the dog owner's information (O-INFO) including their sex and personality traits captured from the Personality Inventory [33].
- S3—Time of Day: These features were computed for three time periods of the day (i.e., night, morning, and afternoon). In S1 and S2, we used all the features. As mentioned in Section 7, the same feature captured at different times of day may have differences in statistical significance values. For example, sleeping time in the morning (M) could be informative to discriminate high and low levels of Extraversion, while sleeping time in the afternoon (A) might not be. Hence, in this set of experiments, we incorporated the time period of the day, and sought to understand whether developing separate models for different time periods yield better performance. For example, if a model trained with data from only morning features performs better, it would mean that we only need six hours of data from a dog to perform the inference.

8 RESULTS

8.1 Activity Level and Statistical Features Discriminating Dog Personality

The statistical features showed high t-statistic, low p-values, and high Cohen's-d values for a majority of dog personality traits on both questionnaires (Table 2). Most of these features were derived from gyroscope, and were captured during the morning. The activity-level features also showed high Cohen's-d values, but not as high as the statistical ones.

- DPQ: For *Fearfulness*, the percentage of time spent in sedentary activity in the afternoon had a Cohen's-d of 0.88, and the amount time of time spent doing light activity at night had a Cohen's-d of 0.78; both of which a large effect size. This translates into saying that a dog's sedentary activity in the afternoon or light activity at night is both informative of high vs. low levels of fearfulness. These were, in fact, the highest Cohen's-d obtained for any activity-level features. In general, the lowest Cohen's-d values came for *Aggression Towards People*. For that trait, light activity percentage at night, morning, and afternoon had Cohen's-d of 0.17, 0.43, and 0.32, respectively, while moderate-vigorous activity level during the night had a Cohen's-d of 0.27. These features had above small effect size with low reliability (because the 95% confidence interval was crossing zero in many cases); thus, no association could have been drawn.
- MCPQ-R: Across all traits, the statistical features had comparably similar values for t-statistic and Cohen's-d to those previously obtained by DPQ. However, activity-level features had Cohen's-d for MCPQ-R higher compared to those for DPQ, showing a better discriminative capability of high class vs. low class for MCPQ-R traits compared to DPQ traits. In particular, statistical features capturing acceleration (i.e., dog movements throughout

Table 2: t-statistic (p-value=<0.05:*; =<0.5:**; >0.5: ***) and Cohen's-d (95% confidence intervals do not overlap with zero: *) of activity-level and statistical features for personality traits in DPQ and MCPQ-R: Night (12am-6am), Morning (6am-12pm), and Afternoon (12pm-6pm) are denoted by N, M, and A, respectively; t-statistic and Cohen's-d values are sorted in descending order, with the highest value of each trait in boldface. Results for activity-level features (left) and statistical features (right) are shown separately. Standard notation min, max, mean, median, and std were used for the minimum, maximum, mean, median, and standard deviation of the signal, respectively. Other notations include: acceleration – acceleration value was calculated in Section 6.1, acc – accelerometer, gyro – gyroscope, x,y,z – axes of the accelerometer and gyroscope, % – the percentage of time spent doing a particular activity, ecdf – empirical cumulative distribution function.

	Activity-Level				**Statistical**			
				DPQ				
		t-statistic		*Cohen's-d*		*t-statistic*		*Cohen's-d*
Fearfulness	**sedentary % (A)**	**(-) 3.55***	**sedentary % (A)**	**0.88***	**gyro z histogram_5 (N)**	**(+) 17.92***	**gyro z histogram_5 (N)**	**4.46***
	light % (N)	(-) 3.12*	light % (N)	0.78*	gyro z histogram_5 (M)	(+) 13.07*	gyro z histogram_5 (M)	3.25*
	acceleration std (M)	(+) 2.46*	acceleration std (M)	0.61*	gyro z histogram_6 (N)	(+) 12.86*	gyro z histogram_6 (N)	3.21*
	sedentary % (M)	(-) 2.43*	sedentary % (M)	0.60*	gyro z zero_crossing_rate (N)	(+) 9.94*	gyro z zero_crossing_rate (N)	2.47*
	sleep % (A)	(+) 2.37*	acceleration std (A)	0.59*	gyro z histogram_6 (M)	(+) 9.56*	gyro z histogram_6 (M)	2.38*
Aggression Towards People	**light % (M)**	**(+) 1.74****	**light % (M)**	**0.43**	**gyro y histogram_8 (N)**	**(-) 4.44***	**gyro y histogram_8 (N)**	**1.11***
	light % (A)	(+) 1.31**	light % (A)	0.32	gyro y histogram_4 (N)	(+) 4.40*	gyro y histogram_4 (N)	1.09*
	sleep % (M)	(-) 1.16***	sleep % (M)	0.29	acc x histogram_5 (M)	(-) 4.38*	gyro y histogram_8 (M)	1.09*
	moderate-vigorous % (N)	(-) 1.08***	moderate-vigorous % (N)	0.27	gyro y histogram_8 (M)	(-) 4.35*	acc x histogram_5 (M)	1.09*
	light % (N)	(+) 0.72***	light % (N)	0.17*	gyro y histogram_7 (M)	(-) 4.33*	gyro y histogram_7 (M)	1.08*
Excitability	**sedentary % (A)**	**(+) 2.05***	**sedentary % (A)**	**0.51***	**gyro z histogram_5 (M)**	**(-)v5.86***	**gyro z histogram_5 (M)**	**1.45***
	acceleration std (A)	(-) 1.50**	acceleration std (A)	0.37	gyro z histogram_5 (N)	(-) 5.85*	gyro z histogram_5 (N)	1.45*
	light % (M)	(-) 1.44**	light % (M)	0.36	gyro z zero_crossing_rate (M)	(-) 5.67*	gyro z zero_crossing_rate (M)	1.41*
	sedentary % (M)	(+) 1.32**	sedentary % (M)	0.33	gyro z zero_crossing_rate (N)	(-) 5.63*	gyro z zero_crossing_rate (N)	1.40*
	acceleration max (A)	(-) 1.27*	acceleration max (A)	0.32	gyro y zero_crossing_rate (N)	(-) 5.54*	gyro y zero_crossing_rate (N)	1.37*
Responsiveness to Training	**light % (M)**	**(-) 2.17***	**light % (M)**	**0.52***	**acc x mean (M)**	**(+) 4.06***	**acc x mean (M)**	**0.99***
	light % (A)	(-) 1.76**	light % (A)	0.42	gyro x ecdf__perc_0 (M)	(+) 3.87*	gyro x histogram_8 (A)	0.97*
	sleep % (M)	(+) 1.52**	sleep % (M)	0.38	gyro z ecdf__perc_0 (M)	(+) 3.81*	gyro x histogram_7 (A)	0.95*
	light % (N)	(-) 1.08**	acceleration std (A)	0.27	gyro z median (M)	(+) 3.73*	gyro z histogram_8 (M)	0.95*
	acceleration std (A)	(+) 1.06**	light % (N)	0.26	gyro z median (N)	(+) 3.71**	gyro x ecdf__perc_0 (N)	0.94*
Aggression Towards Animals	**sedentary % (M)**	**(+) 2.80***	**sedentary % (M)**	**0.69***	**acc y neighbourhood_peaks (M)**	**(+) 5.43***	**acc y neighbourhood_peaks (M)**	**1.35***
	sleep % (M)	(-) 2.61*	sleep % (M)	0.65*	acc z neighbourhood_peaks (M)	(+) 5.20*	acc z neighbourhood_peaks (M)	1.29*
	acceleration mean (M)	(+) 2.33*	acceleration mean (M)	0.58*	gyro y zero_crossing_rate (M)	(+) 5.13*	gyro y zero_crossing_rate (M)	1.27*
	acceleration median (M)	(+) 2.32*	acceleration median (M)	0.58*	gyro x neighbourhood_peaks (M)	(+) 5.06**	gyro x neighbourhood_peaks (M)	1.25*
	acceleration median (N)	(+) 2.32*	acceleration median (N)	0.58*	gyro x ecdf_0 (M)	(-) 4.99*	gyro z ecdf_percentile_count_0 (M)	1.24*
				MCPQ-R				
		t-statistic		*Cohen's-d*		*t-statistic*		*Cohen's-d*
Extraversion	**acceleration min (M)**	**(+) 2.34***	**acceleration min (M)**	**0.60***	**gyro x histogram_4 (N)**	**(-) 4.86***	**gyro x histogram_4 (N)**	**1.24***
	acceleration max (A)	(+) 2.19*	acceleration max (A)	0.56*	gyro x zero_crossing_rate (N)	(-) 4.23*	gyro x zero_crossing_rate (N)	1.08*
	acceleration min (A)	(+) 2.09*	acceleration min (A)	0.54*	acc y histogram_8 (M)	(-) 3.71*	acc y histogram_1 (A)	0.94*
	acceleration min (N)	(+) 1.96**	acceleration min (N)	0.50*	acc y histogram_1 (A)	(+) 3.68*	acc y histogram_8 (M)	0.92*
	acceleration mean (N)	(+) 1.95**	acceleration mean (N)	0.50*	gyro y auc (M)	(-) 3.45*	gyro y auc (M) (N)	0.86*
Motivation	**sedentary % (M)**	**(+) 3.22***	**sedentary % (M)**	**0.80***	**gyro z histogram_7 (M)**	**(-) 4.45***	**gyro z histogram_7 (M)**	**1.10***
	acceleration std (M)	(-) 3.11*	acceleration std (M)	0.76*	gyro z histogram_6 (M)	(-) 4.25**	gyro z histogram_6 (M)	1.06*
	sleep % (A)	(-) 2.58*	acceleration min (M)	0.65*	gyro z histogram_9 (N)	(-) 4.18**	gyro z median (M)	1.05*
	acceleration min (M)	(+) 2.57*	sleep % (A)	0.64*	gyro z median (M)	(-) 4.18*	gyro z median (N)	1.05*
	sleep % (M)	(-) 2.40*	sleep % (M)	0.60*	gyro z histogram_7 (A)	(-) 4.18*	gyro z median (A)	1.04*
Training Focus	**light % (A)**	**(-) 1.88****	**light % (A)**	**0.46**	**gyro x histogram_4 (M)**	**(+) 4.74***	**gyro x histogram_4 (M)**	**1.17***
	acceleration std (M)	(-) 1.80**	acceleration std (M)	0.44	acc x histogram_5 (M)	(+) 4.71*	acc x histogram_5 (M)	1.17*
	moderate-vigorous % (M)	(-) 1.69**	moderate-vigorous % (M)	0.41	acc x mean (M)	(+) 4.56*	acc x mean (M)	1.12*
	sedentary % (A)	(-) 1.59**	sedentary % (A)	0.40	acc x median (M)	(+) 3.98*	acc x median (M)	0.98*
	light % (M)	(-) 1.59**	light % (M)	0.39	gyro y negative_turning_points (M)	(+) 3.88*	gyro y negative_turning_points (M)	0.96*
Amicability	**sleep % (N)**	**(-) 3.80***	**sleep % (N)**	**0.97***	**gyro y zero_crossing_rate (N)**	**(-) 9.79***	**gyro y zero_crossing_rate (N)**	**2.42***
	sleep % (M)	(-) 3.44*	sleep % (M)	0.87*	gyro y zero_crossing_rate (M)	(-) 9.15*	gyro y zero_crossing_rate (M)	2.30*
	sedentary % (M)	(+) 3.29*	light % (M)	0.84*	gyro y histogram_4 (M)	(-) 9.03*	gyro y histogram_4 (M)	2.22*
	light % (M)	(+) 3.20*	sedentary % (M)	0.83*	gyro y histogram_4 (N)	(-) 8.89*	gyro y histogram_4 (N)	2.21*
	accleration min (M)	(+) 3.17*	accleration min (M)	0.83*	gyro y histogram_5 (M)	(-) 7.82*	gyro y histogram_5 (M)	1.94*
Neuroticism	**acceleration std (M)**	**(+) 3.02***	**acceleration std (M)**	**0.75***	**gyro z histogram_8 (N)**	**(+) 6.19***	**gyro z histogram_8 (N)**	**1.53***
	moderate-vigorous % (M)	(+) 2.72*	acceleration min (M)	0.67*	gyro z histogram_7 (N)	(+) 5.80*	gyro z histogram_7 (N)	1.44*
	acceleration min (M)	(-) 2.69*	moderate-vigorous % (M)	0.67*	gyro z histogram_9 (N)	(+) 5.78*	gyro z histogram_9 (N)	1.43*
	acceleration max (A)	(-) 2.49*	acceleration max (A)	0.62*	gyro z histogram_9 (M)	(+) 5.53*	gyro z histogram_9 (M)	1.37*
	acceleration min (A)	(-) 2.44*	acceleration min % (A)	0.61*	gyro z ecdf_percentile_1 (M)	(+) 5.43*	gyro z ecdf_percentile_1 (M)	1.35*

the day) had Cohen's-d values above 0.5, suggesting a link between Extraversion and high levels of activity; a finding in line with previous work [12].

These results suggest that both activity levels and the statistical features have discriminative power to various degrees, allowing us to draw conclusions on which features are associated with dog personality (**RQ$_2$**).

Table 3: Inference results for each personality trait in DPQ and MCPQ-R for different types of models: Mean ($\bar{S}$) and Standard Deviation (S_σ) area under the receiver operating characteristic curve (AUC) computed from five iterations. Results are presented as $\bar{S}(S_\sigma)$, where S is the AUC score, and the highest performing model is marked in bold text. SVM: Support Vector Machines; L-GBM: Light Gradient Boosting Machine; NB: Naive Bayes; RF: Random Forest.

	DPQ				
	Fearfulness	**Aggression Towards People**	**Excitability**	**Responsiveness to Training**	**Aggression Towards Animals**
Baseline	.50 (.00)	.50 (.00)	.50 (.00)	.50 (.00)	.50 (.00)
SVM	.71 (.02)	.66 (.05)	.59 (.03)	**.72 (.06)**	.63 (.09)
L-GBM	.76 (.09)	**.68 (.06)**	.61 (.08)	.66 (.03)	.59 (.10)
NB	.71 (.05)	.63 (.08)	**.62 (.02)**	.65 (.03)	.59 (.04)
RF	**.78 (.07)**	.65 (.08)	.62 (.08)	.70 (.09f)	**.68 (.06)**
	MCPQ-R				
	Extraversion	**Motivation**	**Training Focus**	**Amicability**	**Neuroticism**
Baseline	.50 (.00)	.50 (.00)	.50 (.00)	.50 (.00)	.50 (.00)
SVM	**.64 (.01)**	.70 (.02)	.73 (.06)	.67 (.03)	.62 (.09)
L-GBM	.62 (.07)	.69 (.11)	.70 (.08)	.71 (.05)	.70 (.08)
NB	.64 (.11)	.72 (.05)	.81 (.02)	.70 (.08)	.70 (.02)
RF	.59 (.07)	**.76 (.07)**	**.89 (.13)**	**.74 (.08)**	**.73 (.11)**

Table 4: Random Forest inference results for each personality trait in DPQ and MCPQ-R for different types of features: Mean ($\bar{S}$) and Standard Deviation (S_σ) Area Under the receiver operating characteristic Curve (AUC) computed from five iterations. Results are presented as $\bar{S}(S_\sigma)$, where S is AUC, and the highest performing model is marked in bold text. ACT: activity level features; STAT: statistical features; DEM: dog demographic attributes including its sex, weight, age, training rating, and whether neutered; O-INFO: dog owner's sex and personality traits.

	DPQ				
	Fearfulness	**Aggression Towards People**	**Excitability**	**Responsiveness to Training**	**Aggression Towards Animals**
B1	.50 (.00)	.50 (.00)	.50 (.00)	.50 (.00)	.50 (.00)
B2: O-INFO	.54 (.10)	.48 (.15)	.40 (.18)	.52 (.11)	.49 (.09)
B3: DEM	.57 (.08)	.49 (.06)	.48 (.08)	.55 (.07)	.50 (.12)
G1: ACT	.67 (.05)	.47 (.02)	.60 (.06)	.53 (.08)	.56 (.05)
G2: STAT	.55 (.05)	.47 (.13)	.25 (.09)	.58 (.07)	.53 (.04)
G3: ACT+DEM	**.80 (.11)**	.63 (.13)	.47 (.12)	.61 (.15)	.67 (.05)
G4: STAT+DEM	.63 (.08)	.57 (.11)	.33 (.11)	**.70 (.04)**	.59 (.04)
G5: ACT+STAT	.78 (.07)	.65 (.08)	**.62 (.08)**	.70 (.09)	**.68 (.06)**
G6: ACT+STAT+DEM	.61 (.03)	**.65 (.04)**	.51 (.09)	.61 (.09)	.51 (.06)
	MCPQ-R				
	Extraversion	**Motivation**	**Training Focus**	**Amicability**	**Neuroticism**
B1	.50 (.00)	.50 (.00)	.50 (.00)	.50 (.00)	.50 (.00)
B2: O-INFO	.43 (.02)	.52 (.04)	.56 (.03)	.47 (.10)	.51 (.11)
B3: DEM	.44 (.05)	.54 (.08)	.56 (.04)	.49 (.09)	.52 (.10)
G1: ACT	.41 (.03)	.63 (.06)	.47 (.06)	.41 (.04)	.57 (.02)
G2: STAT	.64 (.06)	.36 (.11)	.34 (.04)	.71 (.05)	.69 (.09)
G3: ACT+DEM	.40 (.09)	**.77 (.09)**	**.89 (.08)**	.38 (.07)	.47 (.02)
G4: STAT+DEM	**.63 (.04)**	.33 (.08)	.43 (.06)	.70 (.08)	.67 (.09)
G5: ACT+STAT	.59 (.07)	.76 (.07)	.89 (.13)	**.74 (.08)**	**.73 (.11)**
G6: ACT+STAT+DEM	.62 (.04)	.33 (.08)	.43 (.07)	.63 (.06)	.70 (.09)

8.2 Predicting Dog Personality

Table 3 shows the classification results for four model types predicting dog personality from both activity levels and statistical features. The columns show the performance for each personality trait, while the rows show model types. The baseline for all experiments is 0.5 as the testing sets were balanced [61]. For DPQ, the best AUCs were obtained with Random Forest (in the range of 0.62-0.80), so for MCPQ-R (in the range of 0.63-0.89). AUC was highest for *Training Focus* with a score of 0.89 using Random Forest and lowest for *Excitability* with a score of 0.59 using SVM. In general, the AUC scores of DPQ questionnaire were lower compared to those of MCPQ-R, meaning that our sensed features are better at classifying MCPQ-R traits (in line with the results in Section 8.1). Given that Random Forest models showed the best performance for the majority of inferences, for brevity, we only present that model's results in the remainder of this section.

Table 4 shows the performance for various feature type combinations and two additional baseline models using: *a)* dog owner's sex and personality (B2: O-INFO), and *b)* dog demographics (i.e.,

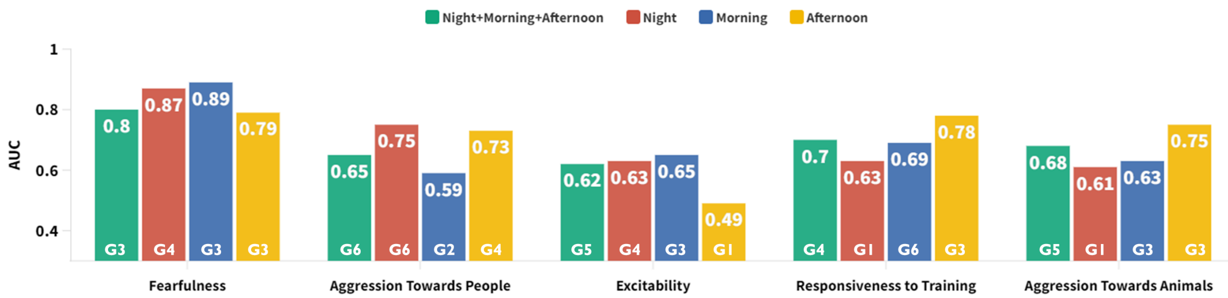


Figure 7: **Area Under the receiver operating characteristic Curve (AUC) score comparison for DPQ traits with models that used features from: night (N); morning (M); afternoon (A); and all time periods (N+M+A). Feature type combinations (G1-G6 from Table 4) that provided the AUC score is marked in white color at the bottom of each bar.**

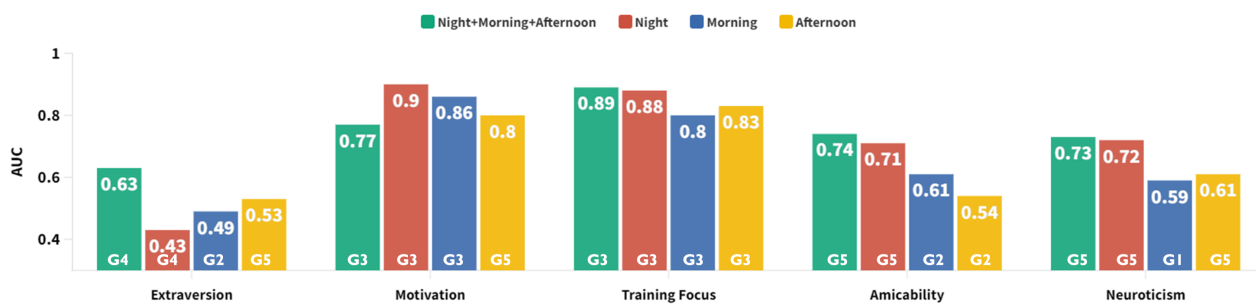


Figure 8: **Area Under the receiver operating characteristic Curve (AUC) score comparison for MCPQ-R traits with models that used features from: night (N); morning (M); afternoon (A); and all time periods (N+M+A). Feature type combinations (=G1-G6 from Table 4) that provided the AUC score is marked in white color at the bottom of each bar.**

sex, weight, age, training rating, and whether neutered) (B3: DEM). Models that used activity-level features alone performed with AUC scores in the range of 0.47-0.67 for DPQ and 0.41-0.63 for MCPQ-R. Statistical features performed worse than activity-level features, with AUC scores in the range of 0.25-0.58 for DPQ and 0.34-0.71 for MCPQ-R. This suggests that, while being more interpretable, activity-level features offer higher predictive accuracies. When adding dog demographics to the models, their performance increased by considerable margins. The best performance for most dog personality traits was obtained when either one or both sensed feature types (activity-level and statistical features) were combined with demographic features. Overall, DPQ traits had AUC scores in the range of 0.62-0.80, with two traits above 0.70, while MCPQ-R traits had scores in the range of 0.63-0.89, with four traits having scores above 0.70. These results show that a combination of sensed features is predictive of dog personality traits with reasonable AUC scores above 0.70.

When the same features were computed at different times of day, they contributed differently to the predictive power. Figure 7 (DPQ) and Figure 8 (MCPQ-R) show varying performances for different time period-specific models (night, morning, or afternoon) compared to generic models that used the features computed throughout the whole day. For DPQ, models that used morning features were the best for predicting *Fearfulness* and *Excitability*, afternoon features were the best for *Responsiveness to Training* and *Aggressiveness Towards Animals*, and finally, night features were the best for *Aggressiveness Towards People*. For MCPQ-R, models that used period-specific features did not yield better results—with the exception of *Motivation* that yielded an AUC of 0.90 with night features. These results suggest that it would be better to use period-specific models for DPQ and generic models for MCPQ-R.

Overall, these results suggest that a specific combination of features works best (i.e., using activity levels, statistical features, and demographics together), and that the same feature computed at different times of day contributes differently to prediction power, yielding higher accuracy when predicting DPQ traits (**RQ_3**).

8.3 Follow-up Survey with Dog Owners

Dog owners answered a series of open-ended questions (in a free-text form) about the usability of Patchkeeper and a series of Likert-scale questions about the desirability of features in future dog monitoring wearables (Figure 9). We structured their responses into two sections: *Usability of PatchKeeper* and *Future dog monitoring wearables.* As a convention, for dog owners' quotes, we use the letter **P** (indicating a human participant) followed by the dog ID from Table 1 (e.g., **P1** is the owner of Dog#1).

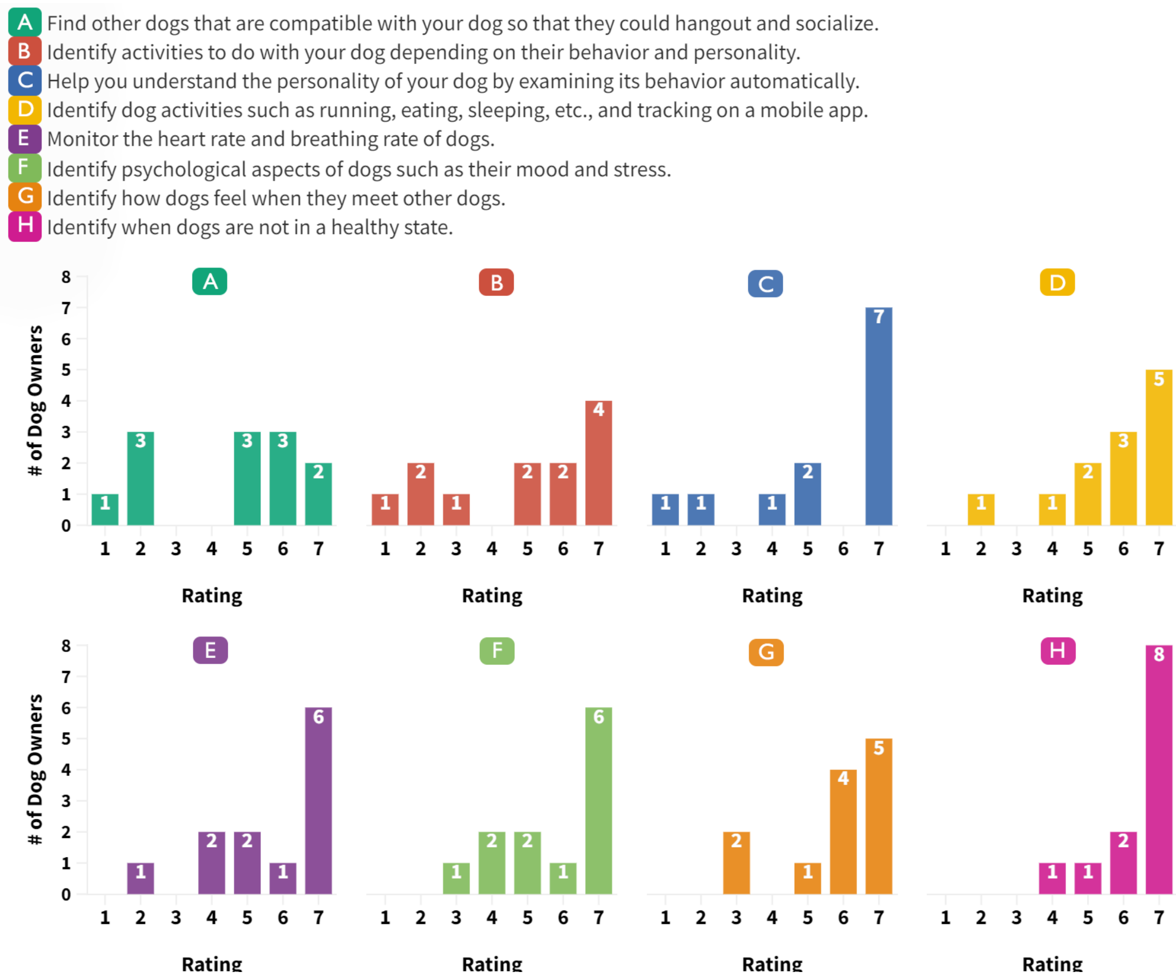


Figure 9: **Response distribution from dog owners about features they would like to see in future dog wearables. Most of them mentioned measurements of aspects that are hard to quantify (e.g., psychological and behavioral states as proxies for a dog's health, feelings, mood, stress, and personality).**

Usability of the PatchKeeper device. A majority of dog owners had a good experience using the device. **P5** mentioned that *" it [PatchKeeper] was easy to use and charge"*, and **P7** stated that *"dog didn't mind wearing it"*. While the device was overall easy to use, owners had mixed feelings about its battery life. One said that *"battery life is too short [...] charging it every day was a hassle"* (**P9**), while another stated that *"battery lasted a full day without an issue"* (**P2**). The split opinions about battery life, however, are very subjective and might be driven by the charging patterns of the owner's other devices. For example, for a person who uses a smartwatch (e.g., Apple or Samsung Galaxy watches) that needs to be charged daily, the Patchkeeper's battery life might seem fair. However, for people who are accustomed to devices with longer battery lives (e.g., AmazFit, Fitbit, or Garmin), PatchKeeper's battery life would be perceived to be much shorter. In terms of developing tools for dog monitoring, owners provided a range of answers, including health monitoring and understanding their pet better. For health monitoring, **P12** mentioned that *"such devices can be useful to catch dog problems early [...] like maybe breathing too fast constantly, or cardiac problems"*. For understanding their pet better, **P3** put it nicely that *"as dogs cannot speak [...] a device that allows my dog to 'speak' and 'express her feelings' is worth everything"*.

Future dog monitoring wearables. Dog owners expressed their opinion about a set of features for future dog wearable devices. They provided answers on a Likert scale of 1 to 7 (1 – strongly not preferred; 3 – not preferred; 5 – preferred; 7 – strongly preferred), and their responses are summarized in Figure 9. Most dog owners (66.67%) strongly preferred features that would allow them to monitor their dog's health. The second most desirable feature was to understand how dogs feel when they meet other dogs. Moreover, even though it was sixth in terms of the mean score, 58.33% of dog owners (7 out of 12) gave a rating of seven out of seven for a feature that measured the personality of their dogs. This shows that dog owners were polarized regarding knowing dog personality. Another feature that received a very high rating was the ability of a wearable to track the mood and stress of dogs. In fact, most of the preferred features were about measures that are hard to quantify (e.g., psychological and behavioral states as proxies for a dog's health, feelings, mood, stress, and personality).

9 DISCUSSION

9.1 Summary of the Results

In summary, our results showed that features captured from the inertial measurement unit along with dog demographic features are predictive of dog personality traits with reasonable AUC scores in the range of 0.62-0.89 (Table 4) in leave-K-dogs-out setting, with K = 4 (around 33% of data was used for testing). In addition, as shown in Table 3, we found that random forest classifiers performed best for the majority of inferences. However, results in Figure 7 and Figure 8 showed that, by using separate models for a particular time of day (i.e., night, morning, or afternoon), and by using different feature groups for each model (i.e., different combinations of activity level, statistical, and demographic features), led to increased performance. In particular, we observed an increased performance across all five personality traits under DPQ (AUC scores of 0.89 for *Fearfulness*; 0.75 for *Aggression towards People*; 0.65 for *Excitability*; 0.78 for *Responsiveness to Training*; and 0.75 for *Aggression towards Animals*), while the performance gains for MCPQ-R (AUC scores of 0.63 for *Extraversion*; 0.90 for *Motivation*; 0.89 for *Training Focus*; 0.74 for *Amicability*; and 0.73 for *Neuroticism*) was only visible for one personality trait, that is, Motivation. This highlights that capturing data from a certain period of the day provides better predictive power for certain personality traits, while for other traits, using all available features was a better option.

9.2 Implications

Our work has both theoretical and practical implications. From a theoretical standpoint, our work adds empirical evidence to the growing body of research on animal personality [31]. We corroborated the previous findings suggesting that more extraverted dogs are associated with higher levels of activity (measured through our device's accelerometer sensor) [12], while amicable dogs engage in light activity (as previous work found [12]). When it comes to aggression towards other animals, we found a moderate association with light activity. While previous work associated aggression with higher activity levels, they did so by studying Siberian Husky dogs [92]; a breed not well represented in our sample. Our models also showed high performance for inferring fearfulness. Even though activity levels and fearfulness were not directly linked in previous literature, a possible explanation could be attributed to the relationship between activity levels and negative emotions (or stress), and negative emotions and stress have been associated with fearfulness [4, 47, 75]. Overall, these results corroborate previous findings in the literature. At the same time, our study provides a fine-grained empirical analysis of the relationship between personality traits and activity levels, extending our theoretical understanding. Beyond animal personality research, our work contributes to the Animal-Computer Interaction literature, including recruitment techniques for in-the-wild studies and the use of pre-trained models for animal activity monitoring. In this work, we argue that dog monitoring could go beyond the typical activity level recognition to capture hard-to-quantify psychological aspects such as dog personality. Turning into our data processing pipeline, we showed that state-of-the-art wearable processing pipelines tailored to humans transferred, to a great extent, to our animal study. In a way, this might seem obvious because accelerometer data capture motion. However, only when data from dogs were properly scaled and processed (using a 1G scale) the pipeline and the inferred activities started to work (§6.1.1). This finding holds great promise for future research in dog activity recognition. As for recruitment techniques, it was evident that dog recruitment required the element of trust. Word of mouth and recruiting in physical proximity through NextDoor turned out to be the best technique. Traditional techniques such as mailing lists, distributing leaflets, and posting on social media (Twitter, Facebook, Instagram) worked to a lesser extent. Another challenge was to retain participants. Given the battery life of 24 hours (even though it is similar to consumer-grade wearables such as Apple Watches[11]), dog owners found it time-consuming and cumbersome to charge the device daily.

From a practical perspective, our findings speak to both dog owners and shelters. A practical application would be a 'dog health app' that tracks a dog's behavior patterns over time and detects its personality or psychological aspects, such as the pet's valence, arousal, and stress levels. Such an app could increase dog owners' awareness of their pet's health and allow them to take proactive actions (e.g., walk the pet to reduce its stress levels). In addition to dog health monitoring, our work can be used for dog socializing. Future platforms could offer owners the ability to receive personalized recommendations for their pets. For example, an owner could subscribe to a service wherein tailored pet social activities are recommended, or their pet is matched with another 'like-minded' pet. Similar to how dating apps allow like-minded individuals to match, such a platform could offer the same experiences for a dog-dog social matching. Finally, dog shelters could benefit by developing platforms for matching dogs with prospective owners. Currently, matching dogs to owners based on personality is difficult for both small and large shelters because of the sheer amount of effort needed to characterize dog personality. On the one hand, using experts to do personality assessments needs specialized facilities and money. On the other hand, psychological scales are time-consuming, prone to biases, and require someone who knows the dog very well. In contrast, letting dogs wear a device like Patchkeeper for a week (or a few days) and obtaining a data-driven personality assessment could be immensely useful.

9.3 Limitations and Future Work

Our work has several limitations that call for future research. First, we modeled dog personality by considering time-level (i.e., features were extracted by considering the signal for three time periods: night, morning, afternoon) and day-level (i.e., features were extracted by combining the signal of the three time periods) data. This allowed us to obtain several data points of the same dog on different days. In this way, we ensured robustness but also increased the relatively small (N=12 dogs) dataset size for inferences (up to 72 dog days). Second, while we obtained reasonable results with a small sample size, future studies could replicate our methodology

[11] around 18 hours of battery: https://www.apple.com/uk/apple-watch-series-7/

with larger sample sizes. However, it is also worth noting that recruiting pet dogs for an in-the-wild study that runs for several days is challenging, and previous work had to resort to similar/lower sample sizes [38, 54, 98]. Additionally, while we enforced the same time schedule for the data collection to obtain comparable results, we acknowledge that different dogs might have different routines. Dogs with similar psychological readings (i.e., personality), but different routines might end up with different physiological readings (i.e., activity levels). Thus future studies could account for dog routines.

Our findings are based on the assumption that dog activity levels serve as a good proxy for activity types. Given that the range of activity types in dogs is narrower (e.g., walking, eating, sleeping, running [41]) compared to that of humans, that was a reasonable assumption. While we prompted dog owners to share activity-type labels with supporting pictures and videos with us (e.g., eating, drinking, running, playing), not all owners were compliant, preventing any further analysis. Future studies could attempt to disentangle dog activity types from dog activity levels by building upon our results. Third, while we resorted to previous literature to control for factors that might have influenced our results (e.g., dog demographics such as sex, age, and neutering [53, 59]), future studies may well incorporate additional factors such as the size of the dog's living environment, or even the presence of other pets in that environment. Fourth, all the dogs in the study, are from the same city in the United Kingdom. Thus, whether these results replicate in other cities or countries remains a subject of future work, specially given that the generalization of mobile sensing-based models across countries is an important topic of interest [62, 63]. Fifth, data collection occurred during the summer period and might not be generalizable. Therefore, future studies could explore whether our findings generalize to other seasons (i.e., winter, fall, spring), when dog behaviors vary (especially in countries closer to hemispheres, where weather drastically changes in different seasons[12]). Sixth, capturing personality could be done in many ways, and, in this study, we only focused on two commonly used personality measurement questionnaires that are filled by dog owners, which were specifically designed for shelter rehoming. Future work should evaluate these scales in the context of other facilities in the USA, and could also explore other dog personality measurement techniques (i.e., test batteries, experts).

10 CONCLUSION

We built a device called "Patchkeeper" that can be strapped on a dog's chest and that measures its activity through an accelerometer and a gyroscope. We experimented with the device on 12 dogs and collected sensor activity data for a week, along with dog personality test results. By matching these two datasets, we trained machine learning classifiers that predicted dog personality from activity data. We found that a combination of activity-level features (describing the activity as sleeping, sedentary, light, or moderate-vigorous) and statistical features (describing temporal and statistical aspects of the time-series accelerometer and gyroscope data) extracted from sensor data, together with dog demographics worked the best. We also found that the same feature computed at different times of day contributed differently to prediction power, with morning features being predictive of fearfulness and excitability, afternoon features being predictive of responsiveness to training and aggressiveness towards animals, and night features being predictive of aggressiveness towards people and motivation.

[12] https://www.pdsa.org.uk/pet-help-and-advice/pet-health-hub/conditions/seasons-in-dogs

ACKNOWLEDGMENTS

We thank Ryo Koblitz, who supported the recruitment, and the owners and dogs who took part in the study. We also thank Edyta Paulina Bogucka for creating the diagrams.

REFERENCES

[1] Herve Abdi and Lynne J Williams. 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2, 4 (2010), 433–459. https://doi.org/10.1002/wics.101
[2] Michelson Found Animals. 2019. *How to Select a Shelter Pet: Choose Personality Over Breed.* Retrieved November 2022 from https://www.foundanimals.org/select-shelter-pet-choose-personality-breed/
[3] Marília Barandas, Duarte Folgado, Letícia Fernandes, Sara Santos, Mariana Abreu, Patrícia Bota, Hui Liu, Tanja Schultz, and Hugo Gamboa. 2020. TSFEL: Time Series Feature Extraction Library. *SoftwareX* 11 (2020), 100456. https://doi.org/10.1016/j.softx.2020.100456
[4] R Beaudet, A Chalifoux, and A Dallaire. 1994. Predictive value of activity level and behavioral evaluation on future dominance in puppies. *Applied Animal Behaviour Science* 40, 3-4 (1994), 273–284.
[5] Róbert Bódizs, Anna Kis, Márta Gácsi, and József Topál. 2020. Sleep in the dog: comparative, behavioral and translational relevance. *Current Opinion in Behavioral Sciences* 33 (2020), 25–33.
[6] CAROLINE BRACKIN. 2022. *Do Dogs Need to Socialize With Other Dogs? (Why Socialization Is Key).* Retrieved August 19, 2022 from https://pethelpful.com/dogs/Why-Socializing-With-Other-Dogs-Is-More-Important-Than-You-Think
[7] Andrew P Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition* 30, 7 (1997), 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2
[8] Yolonda F Bradshaw. 2021. *The Impact of Breed Identification, Potential Adopter Perceptions and Demographics, and Dog Behavior on Shelter Dog Adoptability.* Ph.D. Dissertation. The Ohio State University.
[9] Rita Brugarolas, Tahmid Latif, James Dieffenderfer, Katherine Walker, Sherrie Yuschak, Barbara L Sherman, David L Roberts, and Alper Bozkurt. 2015. Wearable heart rate sensor systems for wireless canine health monitoring. *IEEE Sensors Journal* 16, 10 (2015), 3454–3464.
[10] burgesspetcare. 2021. *Tailor your training to your dog's personality.* Retrieved August 19, 2022 from https://www.burgesspetcare.com/blog/dogs/tailor-your-training-to-your-dogs-personality/
[11] Ceara Byrne, Jay Zuerndorfer, Larry Freil, Xiaochuang Han, Andrew Sirolly, Scott Cilliland, Thad Starner, and Melody Jackson. 2018. Predicting the Suitability of Service Animals Using Instrumented Dog Toys. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 127 (jan 2018), 20 pages. https://doi.org/10.1145/3161184
[12] Lydia Ottenheimer Carrier, Amanda Cyr, Rita E Anderson, and Carolyn J Walsh. 2013. Exploring the dog park: Relationships between social behaviours, personality and cortisol in companion dogs. *Applied Animal Behaviour Science* 146, 1-4 (2013), 96–106.
[13] Robert D. Chambers, Nathanael C. Yoder, Aletha B. Carson, Christian Junge, David E. Allen, Laura M. Prescott, Sophie Bradley, Garrett Wymore, Kevin Lloyd, and Scott Lyle. 2021. Deep Learning Classification of Canine Behavior Using a Single Collar-Mounted Accelerometer: Real-World Validation. *Animals* 11, 6 (2021). https://doi.org/10.3390/ani11061549
[14] Nadine Chersini, Nathan J Hall, and Clive DL Wynne. 2018. Dog pups' attractiveness to humans peaks at weaning age. *Anthrozoös* 31, 3 (2018), 309–318.
[15] François Chollet et al. 2015. keras.
[16] Marios Constantinides, Jonas Busk, Aleksandar Matic, Maria Faurholt-Jepsen, Lars Vedel Kessing, and Jakob E Bardram. 2018. Personalized versus generic mood prediction models in bipolar disorder. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers.* 1700–1707. https://doi.org/10.1145/3267305.3267536
[17] Sara Corsetti, Luisa Pimpolari, and Eugenia Natoli. 2021. How Different Personalities Affect the Reaction to Adoption of Dogs Adopted from a Shelter. *Animals* 11, 6 (2021), 1816. https://doi.org/10.3390/ani11061816
[18] Paul T Costa Jr and Robert R McCrae. 1992. Four ways five factors are basic. *Personality and individual differences* 13, 6 (1992), 653–665. https://doi.org/10.

1016/0191-8869(92)90236-I

[19] Yasin Cotur, Selin Olenik, Tarek Asfour, Michael Bruyns-Haylett, Michael Kasimatis, Ugur Tanriverdi, Laura Gonzalez-Macia, Hong Seok Lee, Andrei S Kozlov, and Firat Güder. 2022. Bioinspired Stretchable Transducer for Wearable Continuous Monitoring of Respiratory Patterns in Humans and Animals. *Advanced Materials* (2022), 2203310. https://doi.org/10.1002/adma.202203310 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.202203310

[20] Elizabeth Cox, Clara Mancini, and Luisa Ruge. 2020. Understanding Dogs' Engagement with Interactive Games: Interaction Style, Behaviour and Personality. In *Proceedings of the Seventh International Conference on Animal-Computer Interaction*. 1–12. https://doi.org/10.1145/3446002.3446122

[21] Adele Cutler, David Cutler, and John Stevens. 2011. *Random Forests*. Vol. 45. 157–176. https://doi.org/10.1007/978-1-4419-9326-7_5

[22] Aiden Doherty, Dan Jackson, Nils Hammerla, Thomas Plötz, Patrick Olivier, Malcolm H Granat, Tom White, Vincent T Van Hees, Michael I Trenell, Christoper G Owen, et al. 2017. Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study. *PloS one* 12, 2 (2017), e0169649. https://doi.org/10.1371/journal.pone.0169649

[23] Aiden Doherty, Karl Smith-Byrne, Teresa Ferreira, Michael V Holmes, Chris Holmes, Sara L Pulit, and Cecilia M Lindgren. 2018. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nature communications* 9, 1 (2018), 1–8. https://doi.org/10.1038/s41467-018-07743-4

[24] Amy Ellard-Gray, Nicole K Jeffrey, Melisa Choubak, and Sara E Crann. 2015. Finding the hidden participant: Solutions for recruiting hidden, hard-to-reach, and vulnerable populations. *International Journal of Qualitative Methods* 14, 5 (2015), 1609406915621420. https://doi.org/10.1177/1609406915621420 arXiv:https://doi.org/10.1177/1609406915621420

[25] Aiden Doherty et al. 2023. *Biobank Accelerometer Analysis*. Retrieved January 31, 2023 from https://biobankaccanalysis.readthedocs.io/

[26] Marc Foster, Rita Brugarolas, Katherine Walker, Sean Mealin, Zach Cleghern, Sherrie Yuschak, Julia Condit Clark, Darcy Adin, Jane Russenberger, Margaret Gruen, et al. 2020. Preliminary evaluation of a wearable sensor system for heart rate assessment in guide dog puppies. *IEEE Sensors Journal* 20, 16 (2020), 9449–9459.

[27] Animal Farm Foundation. 2020. *Does breed really matter when choosing a pet dog?* Retrieved November 2022 from https://www.animalfarmfoundation.org/dont-make-this-mistake-when-adopting-a-dog/

[28] Marieke Cassia Gartner. 2015. Pet personality: A review. *Personality and Individual Differences* 75 (2015), 102–113. https://doi.org/10.1016/j.paid.2014.10.042

[29] H Hill Goldsmith, Arnold H Buss, Robert Plomin, Mary Klevjord Rothbart, Alexander Thomas, Stella Chess, Robert A Hinde, and Robert B McCall. 1987. Roundtable: What is temperament? Four approaches. *Child development* (1987), 505–529.

[30] Samuel D Gosling. 2001. From mice to men: what can we learn about personality from animal research? *Psychological bulletin* 127, 1 (2001), 45.

[31] Samuel D Gosling. 2008. Personality in non-human animals. *Social and Personality Psychology Compass* 2, 2 (2008), 985–1001. https://doi.org/10.1111/j.1751-9004.2008.00087.x arXiv:https://compass.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-9004.2008.00087.x

[32] Samuel D Gosling, Virginia SY Kwan, and Oliver P John. 2003. A dog's got personality: a cross-species comparative approach to personality judgments in dogs and humans. *Journal of personality and social psychology* 85, 6 (2003), 1161.

[33] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

[34] Greenland Sander, Senn Stephen J., Rothman Kenneth J., Carlin John B., Poole Charles, Goodman Steven N., and Altman Douglas G. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology* 31, 4 (2016), 337–350. https://doi.org/10.1007/s10654-016-0149-3

[35] Joel Griffies, Jason Zutty, Marcel Sarzen, and Stuart Soorholtz. 2018. Wearable sensor shown to specifically quantify pruritic behaviors in dogs. *BMC Veterinary Research* 14 (04 2018), 124. https://doi.org/10.1186/s12917-018-1428-x

[36] Julie Hecht and Alexandra Horowitz. 2015. Seeing dogs: Human preferences for dog physical attributes. *Anthrozoös* 28, 1 (2015), 153–163.

[37] Ilyena Hirskyj-Douglas and Andrés Lucero. 2019. On the Internet, Nobody Knows You're a Dog... Unless You're Another Dog. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3290605.3300347

[38] Ilyena Hirskyj-Douglas, Roosa Piitulainen, Andrés Lucero, et al. 2021. Forming the Dog Internet: Prototyping a Dog-to-Human Video Call Device. *Proc. ACM Hum. Comput. Interact.* 5, ISS (2021), 1–20. https://doi.org/10.1145/3488539

[39] Amanda Höglin, Enya Van Poucke, Rebecca Katajamaa, Per Jensen, Elvar Theodorsson, and Lina SV Roth. 2021. Long-term stress in dogs is related to the human–dog relationship and personality traits. *Scientific Reports* 11, 1 (2021), 1–9.

[40] Yuying Hsu and James A Serpell. 2003. Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *Journal of the American Veterinary Medical Association* 223, 9 (2003), 1293–1300. https://doi.org/10.2460/javma.2003.223.1293

[41] Ali Hussain, Sikandar Ali, Hee-Cheol Kim, et al. 2022. Activity Detection for the Wellbeing of Dogs Using Wearable Sensors Based on Deep Learning. *IEEE Access* 10 (2022), 53153–53163.

[42] Sara Isgate and Justin J Couchman. 2018. What makes a dog adoptable? An eye-tracking Investigation. *Journal of applied animal welfare science* 21, 1 (2018), 69–81.

[43] Oliver P John. 1990. The" Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. *Handbook of personality: Theory and research* (1990).

[44] Oliver P John and Samuel D Gosling. 2000. Personality traits. (2000).

[45] Amanda Claire Jones. 2008. *Development and* validation of a Dog Personality Questionnaire*. The University of Texas at Austin.

[46] Amanda C Jones and Samuel D Gosling. 2005. Temperament and personality in dogs (Canis familiaris): A review and evaluation of past research. *Applied Animal Behaviour Science* 95, 1-2 (2005), 1–53. https://doi.org/10.1016/j.applanim.2005.04.008

[47] Sarah Jones, Seana Dowling-Guyer, Gary J Patronek, Amy R Marder, Sheila Segurson D'Arpino, and Emily McCobb. 2014. Use of accelerometers to measure stress levels in shelter dogs. *Journal of Applied Animal Welfare Science* 17, 1 (2014), 18–28.

[48] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30 (2017). https://doi.org/10.5555/3294996.3295074

[49] Mohammed Khwaja, Sumer S Vaid, Sara Zannone, Gabriella M Harari, A Aldo Faisal, and Aleksandar Matic. 2019. Modeling personality vs. modeling personalidad: In-the-wild mobile data analysis in five countries suggests cultural impact on personality models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24. https://doi.org/10.1145/3351246

[50] Hae-Young Kim. 2015. Statistical notes for clinical researchers: effect size. *Restorative dentistry & endodontics* 40, 4 (2015), 328–331. https://doi.org/10.5395/rde.2015.40.4.328

[51] Tae Kim. 2015. T test as a parametric statistic. *Korean Journal of Anesthesiology* 68 (11 2015), 540. https://doi.org/10.4097/kjae.2015.68.6.540

[52] Jo Knowsley. 2021. *Meet the dog charity that matches up pooches to new owners by personality*. Retrieved August 19, 2022 from https://metro.co.uk/2021/06/15/meet-the-dog-charity-that-matches-up-pets-to-new-owners-by-personality-14768607/

[53] Enikő Kubinyi, Borbála Turcsán, and Ádám Miklósi. 2009. Dog and owner demographic characteristics and dog personality trait associations. *Behavioural processes* 81, 3 (2009), 392–401. https://doi.org/10.1016/j.beproc.2009.04.004

[54] Cassim Ladha, Nils Hammerla, Emma Hughes, Patrick Olivier, and Thomas Ploetz. 2013. Dog's Life: Wearable Activity Recognition for Dogs. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 415–418. https://doi.org/10.1145/2493432.2493519

[55] Daniel Lakens. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology* 4 (2013). https://doi.org/10.3389/fpsyg.2013.00863

[56] Fiona Lamb, Allison Andrukonis, and Alexandra Protopopova. 2021. The role of artificial photo backgrounds of shelter dogs on pet profile clicking and the perception of sociability. *PloS one* 16, 12 (2021), e0255551.

[57] Jacqueline Ley, Pauleen Bennett, and Grahame Coleman. 2008. Personality dimensions that emerge in companion canines. *Applied Animal Behaviour Science* 110, 3-4 (2008), 305–317. https://doi.org/10.1016/j.applanim.2007.04.016

[58] Jacqui M Ley, Pauleen C Bennett, and Grahame J Coleman. 2009. A refinement and validation of the Monash Canine Personality Questionnaire (MCPQ). *Applied Animal Behaviour Science* 116, 2-4 (2009), 220–227. https://doi.org/10.1016/j.applanim.2008.09.009

[59] Sarah E. Lofgren, Pamela Wiener, Sarah C. Blott, Enrique Sanchez-Molano, John A. Woolliams, Dylan N. Clements, and Marie J. Haskell. 2014. Management and personality in Labrador Retriever dogs. *Applied Animal Behaviour Science* 156 (2014), 44–53. https://doi.org/10.1016/j.applanim.2014.04.006

[60] Robert R McCrae, Paul T Costa Jr, Fritz Ostendorf, Alois Angleitner, Martina Hřebíčková, Maria D Avia, Jesús Sanz, Maria L Sanchez-Bernardos, M Ersin Kusdil, Ruth Woodfield, et al. 2000. Nature over nurture: temperament, personality, and life span development. *Journal of personality and social psychology* 78, 1 (2000), 173.

[61] Lakmal Meegahapola, Wageesha Bangamuarachchi, Anju Chamantha, Salvador Ruiz-Correa, Indika Perera, and Daniel Gatica-Perez. 2022. Sensing Eating Events in Context: A Smartphone-Only Approach. *IEEE Access* 10, ARTICLE

(2022). https://doi.org/10.1109/ACCESS.2022.3179702
[62] Lakmal Meegahapola, William Droz, Peter Kun, Amalia de Götzen, Chaitanya Nutakki, Shyam Diwakar, Salvador Ruiz Correa, Donglei Song, Hao Xu, Miriam Bidoglia, George Gaskell, Altangerel Chagnaa, Amarsanaa Ganbold, Tsolmon Zundui, Carlo Caprini, Daniele Miorandi, Alethia Hume, Jose Luis Zarza, Luca Cernuzzi, Ivano Bison, Marcelo Rodas Britez, Matteo Busso, Ronald Chenu-Abente, Can Günel, Fausto Giunchiglia, Laura Schelenz, and Daniel Gatica-Perez. 2023. Generalization and Personalization of Mobile Sensing-Based Mood Inference Models: An Analysis of College Students in Eight Countries. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 176 (jan 2023), 32 pages. https://doi.org/10.1145/3569483
[63] Lakmal Meegahapola and Daniel Gatica-Perez. 2021. Smartphone Sensing for the Well-Being of Young Adults: A Review. *IEEE Access* 9 (2021), 3374–3399. https://doi.org/10.1109/ACCESS.2020.3045935
[64] Lakmal Meegahapola, Salvador Ruiz-Correa, Viridiana del Carmen Robledo-Valero, Emilio Ernesto Hernandez-Huerfano, Leonardo Alvarez-Rivera, Ronald Chenu-Abente, and Daniel Gatica-Perez. 2021. One More Bite? Inferring Food Consumption Level of College Students Using Smartphone Sensing and Self-Reports. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 26 (mar 2021), 28 pages. https://doi.org/10.1145/3448120
[65] Erika Mirkó, Enikő Kubinyi, Márta Gácsi, and Ádám Miklósi. 2012. Preliminary analysis of an adjective-based dog personality questionnaire developed to measure some aspects of personality in the domestic dog (Canis familiaris). *Applied Animal Behaviour Science* 138, 1-2 (2012), 88–98. https://doi.org/10.1016/j.applanim.2012.02.016
[66] R Morrison, V Penpraze, A Beber, JJ Reilly, and PS Yam. 2013. Associations between obesity and physical activity in dogs: a preliminary investigation. *Journal of Small Animal Practice* 54, 11 (2013), 570–574.
[67] Subigya Nepal, Shayan Mirjafari, Gonzalo J Martinez, Pino Audia, Aaron Striegel, and Andrew T Campbell. 2020. Detecting job promotion in information workers using mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–28. https://doi.org/10.1145/3414118
[68] William S Noble. 2006. What is a support vector machine? *Nature biotechnology* 24, 12 (2006), 1565–1567. https://doi.org/10.1038/nbt1206-1565
[69] Mikio Obuchi, Jeremy F Huckins, Weichen Wang, Alex daSilva, Courtney Rogers, Eilis Murphy, Elin Hedlund, Paul Holtzheimer, Shayan Mirjafari, and Andrew Campbell. 2020. Predicting brain functional connectivity using mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–22. https://doi.org/10.1145/3381001
[70] Heidi K Ortmeyer, Lynda Robey, and Tara McDonald. 2018. Combining actigraph link and PetPace collar data to measure activity, proximity, and physiological responses in freely moving dogs in a natural environment. *Animals* 8, 12 (2018), 230.
[71] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-Learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830. http://arxiv.org/abs/1201.0490
[72] Kathelijne Peremans, Kurt Audenaert, Frank Coopman, Peter Blanckaert, Filip Jacobs, Andreas Otte, Francis Verschooten, Henri van Bree, Kees van Heeringen, John Mertens, et al. 2003. Estimates of regional cerebral blood flow and 5-HT2A receptor density in impulsive, aggressive dogs with 99mTc-ECD and 123I-5-I-R91150. *European journal of nuclear medicine and molecular imaging* 30, 11 (2003), 1538–1546.
[73] Lawrence A Pervin and Oliver P. John. 1997. *Personality: Theory and research, 7th ed.* John Wiley & Sons.
[74] Julie A Posluns, Rita E Anderson, and Carolyn J Walsh. 2017. Comparing two canine personality assessments: Convergence of the MCPQ-R and DPQ and consensus between dog owners and dog walkers. *Applied Animal Behaviour Science* 188 (2017), 68–76. https://doi.org/10.1016/j.applanim.2016.12.013
[75] Diana J Rayment, Bert De Groef, Richard A Peters, and Linda C Marston. 2015. Applied personality assessment in domestic dogs: Limitations and caveats. *Applied Animal Behaviour Science* 163 (2015), 1–18.
[76] Marnie Rice and Grant Harris. 2005. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d. *Law and human behavior* 29 (11 2005), 615–20. https://doi.org/10.1007/s10979-005-6832-7
[77] Andrew N Rowan. 1992. Shelters and pet overpopulation: A statistical black hole. , 140–143 pages. https://doi.org/10.2752/089279392787011430
[78] Humphrey. Elliott S. 1934. "Mental tests" for shepherd dogs: An attempted classification and evaluation of the various traits that go to make up "temperament" in the german shepherd dog. *Journal of Heredity* 25, 4 (1934), 129–136. https://doi.org/10.1093/oxfordjournals.jhered.a103899 arXiv:https://academic.oup.com/jhered/article-pdf/25/4/129/2591195/25-4-129.pdf
[79] Milla Salonen, Salla Mikkola, Emma Hakanen, Sini Sulkama, Jenni Puurunen, and Hannes Lohi. 2021. Reliability and Validity of a Dog Personality and Unwanted Behavior Survey. *Animals* 11, 5 (2021). https://doi.org/10.3390/ani11051234
[80] Luke A Schneider, Paul H Delfabbro, and Nicholas R Burns. 2013. Temperament and lateralization in the domestic dog (Canis familiaris). *Journal of Veterinary Behavior* 8, 3 (2013), 124–134. https://doi.org/10.1016/j.jveb.2012.06.004
[81] Ivana Gabriela Schork, Isabele Aparecida Manzo, Marcos Roberto Beiral De Oliveira, Fernanda Vieira da Costa, Robert John Young, and Cristiano Schetini de Azevedo. 2022. The cyclic interaction between daytime behavior and the sleep behavior of laboratory dogs. *Scientific reports* 12, 1 (2022), 1–9.
[82] shibashake. 2017. *Do Dogs Need Other Dogs?* Retrieved August 19, 2022 from https://shibashake.com/dog/do-dogs-need-other-dogs/#:~:text=Dogs%20do%20not%20need%20to,companion%2C%20so%20much%20the%20better.
[83] Christopher Sibona and Steven Walczak. 2012. Purposive sampling on Twitter: A case study. In *2012 45th Hawaii international conference on system sciences*. IEEE, 3510–3519. https://doi.org/10.1109/HICSS.2012.493
[84] David L Sinn, Samuel D Gosling, and Stewart Hilliard. 2010. Personality and performance in military working dogs: Reliability and predictive validity of behavioral tests. *Applied Animal Behaviour Science* 127, 1-2 (2010), 51–65. https://doi.org/10.1016/j.applanim.2010.08.007
[85] NH Humane Society. 2022. *Canine Personalities*. Retrieved August 19, 2022 from https://nhhumane.org/adopt/meet-your-match/dogs
[86] Kenth Svartberg. 2006. Breed-typical behaviour in dogs—historical remnants or recent constructs? *Applied Animal Behaviour Science* 96, 3-4 (2006), 293–313. https://doi.org/10.1016/j.applanim.2005.06.014
[87] Sirin Theerawatanasirikul, Gunnaporn Suriyaphol, Roongroje Thanawongnuwech, and Achariya Sailasuta. 2012. Histologic morphology and involucrin, filaggrin, and keratin expression in normal canine skin from dogs of different breeds and coat types. *Journal of Veterinary Science* 13, 2 (2012), 163–170.
[88] Judit Vas, József Topál, Eva Péch, and Adám Miklósi. 2007. Measuring attention deficit and activity in dogs: a new application and validation of a human ADHD questionnaire. *Applied Animal Behaviour Science* 103, 1-2 (2007), 105–117.
[89] Alessandro Vinciarelli and Gelareh Mohammadi. 2014. A survey of personality computing. *IEEE Transactions on Affective Computing* 5, 3 (2014), 273–291. https://doi.org/10.1109/TAFFC.2014.2330816
[90] Rosemary Walmsley, Shing Chan, Karl Smith-Byrne, Rema Ramakrishnan, Mark Woodward, Kazem Rahimi, Terence Dwyer, Derrick Bennett, and Aiden Doherty. 2021. Reallocation of time between device-measured movement behaviours and risk of incident cardiovascular disease. *British journal of sports medicine* (2021). https://doi.org/10.1136/bjsports-2021-104050
[91] Rosemary Walmsley, Shing Chan, Karl Smith-Byrne, Rema Ramakrishnan, Mark Woodward, Kazem Rahimi, Terence Dwyer, Derrick Bennett, and Aiden Doherty. 2022. Reallocation of time between device-measured movement behaviours and risk of incident cardiovascular disease. *British journal of sports medicine* 56, 18 (2022), 1008–1017.
[92] Michele Wan, Krisztina Hejjas, Zsolt Ronai, Zsuzsanna Elek, Maria Sasvari-Szekely, Frances A Champagne, Ádám Miklósi, and Enikő Kubinyi. 2013. DRD 4 and TH gene polymorphisms are associated with activity, impulsivity and inattention in Siberian Husky dogs. *Animal genetics* 44, 6 (2013), 717–727.
[93] Rui Wang, Weichen Wang, Min SH Aung, Dror Ben-Zeev, Rachel Brian, Andrew T Campbell, Tanzeem Choudhury, Marta Hauser, John Kane, Emily A Scherer, et al. 2017. Predicting symptom trajectories of schizophrenia using mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–24. https://doi.org/10.1145/3130976
[94] Rui Wang, Weichen Wang, Mikio Obuchi, Emily Scherer, Rachel Brian, Dror Ben-Zeev, Tanzeem Choudhury, John Kane, Martar Hauser, Megan Walsh, et al. 2020. On predicting relapse in schizophrenia using mobile sensing in a randomized control trial. In *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–8. https://doi.org/10.1109/PerCom45495.2020.9127365
[95] Weichen Wang, Subigya Nepal, Jeremy F Huckins, Lessley Hernandez, Vlado Vojdanovski, Dante Mack, Jane Plomp, Arvind Pillai, Mikio Obuchi, Alex daSilva, et al. 2022. First-Gen Lens: Assessing Mental Health of First-Generation Students across Their First Year at College Using Mobile Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–32. https://doi.org/10.1145/3543194
[96] Geoffrey I Webb, Eamonn Keogh, and Risto Miikkulainen. 2010. Naïve Bayes. *Encyclopedia of machine learning* 15 (2010), 713–714. https://doi.org/10.1007/978-1-4899-7502-7_581-1
[97] Emily Weiss. 2002. Selecting shelter dogs for service dog training. *Journal of Applied Animal Welfare Science* 5, 1 (2002), 43–62.
[98] Gary M. Weiss, Ashwin Nathan, J.B. Kropp, and Jeffrey W. Lockhart. 2013. WagTag: A Dog Collar Accessory for Monitoring Canine Activity Levels. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication* (Zurich, Switzerland) *(UbiComp '13 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 405–414. https://doi.org/10.1145/2494091.2495972

[99] Eric W Weisstein. 2004. Bonferroni correction. *https://mathworld. wolfram. com/* (2004).

[100] Christopher Whitaker, Sharon Stevelink, Nicola Fear, et al. 2017. The use of Facebook in recruiting participants for health research purposes: a systematic review. *Journal of medical Internet research* 19, 8 (2017), e7071. https://doi.org/10.2196/jmir.7071

[101] Matthew Willetts, Sven Hollowell, Louis Aslett, Chris Holmes, and Aiden Doherty. 2018. Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. *Scientific reports* 8, 1 (2018), 1–10. https://doi.org/10.1038/s41598-018-26174-1

[102] Anna Zamansky, Dirk Van Der Linden, Irit Hadar, and Stephane Bleuer-Elsner. 2019. Log my dog: perceived impact of dog activity tracking. *Computer* 52, 9 (2019), 35–43. https://doi.org/10.1109/MC.2018.2889637