Quantifying Nocturnal Itch And Its Impact On Sleep Using Machine Learning And Radio Signals

by

Michail Ouroutzoglou

Diploma in Electrical and Computer Engineering,
Technical University of Crete (2019)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2022

Author . . . .
Department of Electrical Engineering and Computer Science
August 26, 2022

Certified by . . . .
Dina Katabi
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by . . . .
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Quantifying Nocturnal Itch And Its Impact On Sleep Using Machine Learning And Radio Signals

by

Michail Ouroutzoglou

Submitted to the Department of Electrical Engineering and Computer Science
on August 26, 2022, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

Today, chronic itch affects up to 15% of the population, and is associated with over $90 Billion in annual population-expenditures in the US. Despite all the interest around this area, there's still no solution for quantifying nocturnal scratching and its impact on patients' sleep quality in an objective, sensitive and privacy preserving way. In this work we collect large nocturnal scratching dataset, consisting of 370 nights of infrared footage, radio-frequency (RF) data, and human annotations of scratching. Using this data, we develop a neural network model that can detect occurrences of nocturnal scratching using only radio signals. The developed model can achieve very high accuracy in measuring meaningful scratching metrics, across a diverse population of patients. Additionally, by utilizing prior art on extracting sleep stages from radio signals, we can gain insights about the effect of itch on the sleep quality of a chronic itch patients, especially relative to healthy individuals.

Thesis Supervisor: Dina Katabi
Title: Professor of Electrical Engineering and Computer Science

Disclaimer

This dissertation is based on a soon to be submitted publication: Michail Ouroutzoglou, Mingmin Zhao, Hariharan Rahul, Asima Badic, Brian Kim, Dina Katabi. Machine Learning and Radio Waves to Measure and Understand Itch.

My contributions to this work include development of systems and algorithms, model implementation and evaluation, data collection, data annotation, result extraction, figure generation, and writing of paper.

Acknowledgments

I would like to express my deepest gratitude to my thesis supervisor, professor Dina Katabi. She gave me the opportunity to do research at this amazing institution, surrounded by ideas, innovation and brilliant people. Thank you Dina for your support and guidance!

I am grateful to Mingmin Zhao with whom I worked closely when I first joined Dina's group. He taught me many things.

I would like to thank the members of NETMIT and Emerald, as well as my friends, for all the insightful discussions and fun times we've had.

Finally, I would like to thank my family and my amazing girlfriend, Victoria Wu, for being there for me everyday with their unconditional love and support. I wouldn't have made it without you. Thank you for being there for me through thick and thin.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Itch, a basic sensation that can be evoked by a mosquito bite, becomes a torturous experience when chronic. Chronic itch is so brutal that, in Dante's Inferno, falsifiers were eternally punished by "the burning rage of fierce itching that nothing could relieve". Today, chronic itch affects up to 15% of the population, and is associated with over $90 Billion in annual population-expenditures in the US [1]. It is often as debilitating as chronic pain and has a profound negative impact on quality of life [2, 3, 4]. Despite all the recent advances resulting in a new FDA-approval for itch in the setting of chronic kidney disease and cholestasis [5], many chronic pruritic conditions such as atopic dermatitis (AD), prurigo nodularis (PN), and chronic pruritus of unknown origin (CPUO) do not have itch as a specific indication. A key challenge for both the management and development of new drugs is the lack of an objective measure for quantifying itch [6]. The current clinical standard for quantifying itch is the numerical rating scale (NRS), which is a score (from 0 to 10) patients assign to their itch over the past 24 hours [7]. This method has low sensitivity and generalizes poorly across populations [8, 9].

For two decades the research community has tried to address this problem eventually proposing nocturnal scratching as an objective measure for quantifying itch [10, 11, 12, 13, 14, 15]. One promising method for quantifying nocturnal scratching is through an infrared camera installed into the patient's bedroom. Although this method is considered the golden standard in terms of accurately measuring nocturnal

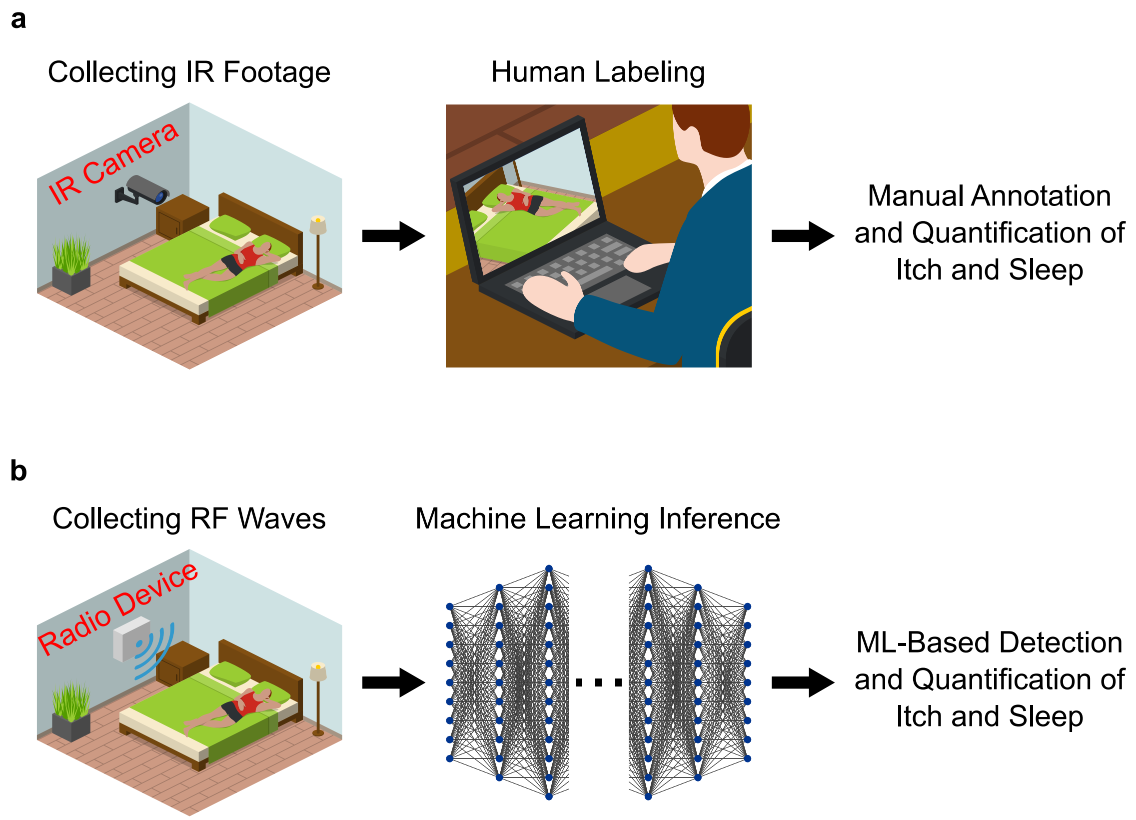

Figure 1-1: **Illustration of video-based and radio-based assessment of scratching and sleep.** (**a**) Assessment of nocturnal scratching through human annotation of infrared videos of a patient in their bed. This approach is privacy intrusive, has high overhead, and requires a lot of manual human labor. (**b**) Assessment of nocturnal scratching by analyzing the radio waves that bounce off a patient in their bed, using a machine learning model. This approach is privacy preserving, has low overhead, and inference can be performed automatically.

scratching, it comes with some important drawbacks, such as privacy intrusion, and high overhead in time and cost [13, 14, 15]. Finally, some prior work has proposed solutions based on wearables (e.g., wrist actigraphy watches). However, this too has limitations, such as low accuracy and compliance issues.

Thus, while nocturnal scratching is a promising metric for quantifying itch, there remains a strong need for an accurate, low overhead, privacy-preserving method for measuring scratching.

Furthermore, chronic itch leads to disturbed sleep and diminished quality of life [16, 17, 18], hence an ideal solution would concurrently measure scratching and its impact on sleep. Poor sleep is associated with major co-morbidities including behavioral and neurocognitive deficits in children with chronic eczema [16], and cardiovascular death in patients suffering uremic pruritus [19]. Sleep studies in itch patients have mostly relied on patients' self-assessment via sleep-related questionnaires such as the Pittsburgh Sleep Quality Index (PSQI) and the Medical Outcomes Study (MOS) [20, 21]. Like self-assessment of itch, sleep self-assessment is subjective, insensitive, and difficult to conduct in children and cognitively impaired patients. One option for characterizing the impact of itch on sleep is to conduct polysomnography (PSG) sleep studies [22, 23], in which patients sleep in a sleep lab instrumented with a plethora of sensors and electrodes on the head and body. Such option however imposes a high burden on patients and is impractical for continuous daily sleep measurements.

This work develops an objective, sensitive, and easy-to-use solution that measures both nocturnal scratching and its impact on sleep. We build on recent results in wireless sensing, which have demonstrated the feasibility and accuracy of monitoring physiological signals (e.g., breathing, heart rate, sleep stages, and gait) by analyzing the radio signals that bounce off people's bodies, without physical contact or wearable devices [24, 25, 26, 27, 28, 29, 30, 31, 32].

Our design is based on a non-obtrusive wireless device that sits in the background at home, much like a Wi-Fi router (see Fig. 1-2). It transmits a low power radio signal (1000 times lower power than home Wi-Fi) and captures its reflections. Since the human body is primarily water, it reflects the radio signal and modulates it with the subject's movements. The device analyses the reflected radio signals using machine learning (ML) to infer scratching and sleep. To infer sleep, we leverage past work on sleep characterization using radio signals [29, 32]. To infer scratching instances, we have developed a novel neural network model. The network first identifies repetitive motion patterns that characterize scratching, then processes those patters to output a time series labeled with the occurrence of scratching instances. The approach is passive and does not require patients to wear sensors or report the severity of their symptoms. Figure 1-1 (**b**) illustrates the operation of our solution.

We have conducted an observational clinical trial to evaluate our radio-based

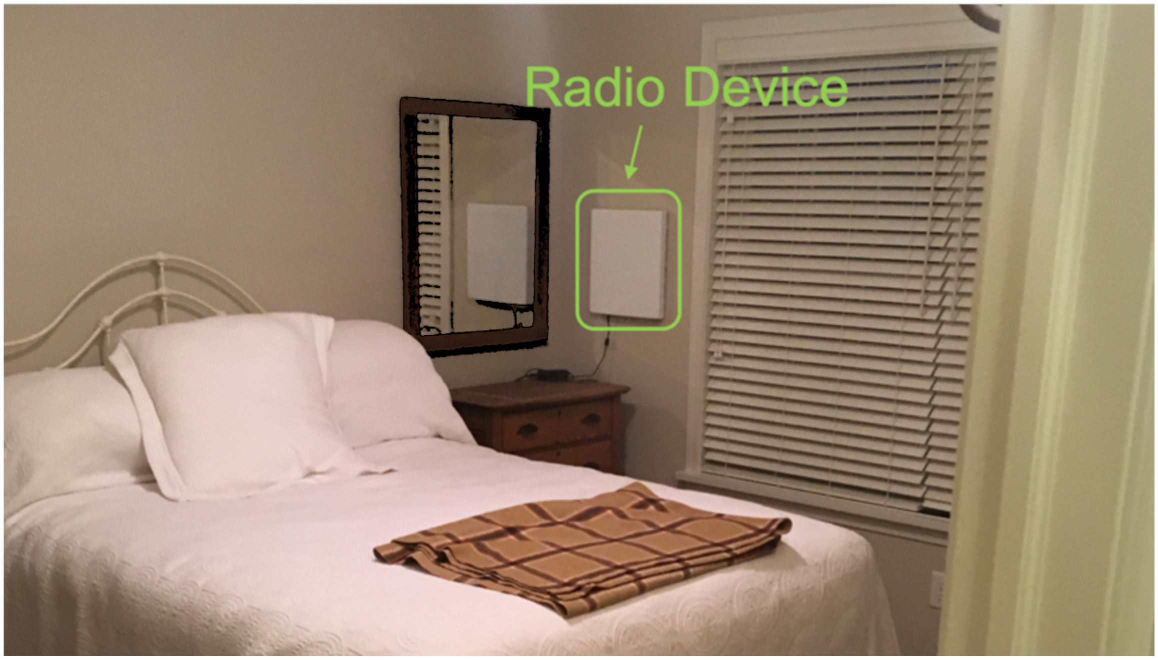

Figure 1-2: **Example of a deployed device in a patient's bedroom.**

solution. We also leveraged the device to characterize the impact of scratching on sleep quality.

In short, this work presents the following:

- Development of a system for measuring nocturnal scratching in an objective, sensitive, and privacy-preserving way, using radio signals and machine learning.
- We ran a clinical study and collected hundreds of nights of sleep and nocturnal scratching data using both an infrared video camera and radio signals.
- The proposed model can achieve a ROC AUC of 0.9956, Precision-Recall AUC of 0.8595, and can accurately extract meaningful metrics that characterize itch.
- We compare patients' self-reported itch severity, a subjective metric, with the ground truth scratching from infrared cameras, an objective metric, and show that the two metrics are only weakly correlated.
- We extract sleep metrics from the chronic itch patients, and show that nocturnal scratching has a significant negative impact on their sleep efficiency, sleep latency and WASO (wake after sleep onset).

Chapter 2

Related Work

2.1 Patient-Reported Outcomes

Assessment of itch is very often performed through patient self-reporting. The patient is asked to give a subjective evaluation of their condition. Several different patient-reported outcome (PROs) metrics have been developed to quantify itch severity. The most prominent self-assessment metrics available are the numerical rating scale (NRS) [7], the visual analog scale (VAS) [33], and the 5-D itch scale [10]. The NRS is a popular choice due to its simplicity; the patients are only required to assign a number to their itch on the scale of 0 (“no itch”) to 10 (“worst imaginable itch”). The VAS is another common choice for clinical trials. It consists of a 10cm long line, and patients are asked put a mark on the line depending on their itch intensity. The leftmost end represents "no itch" while the rightmost end represents "worst imaginable itch". The 5-D itch scale is slightly more involved. It asks the patients to assess their itch on 5 different dimensions, specifically, degree, duration, direction, disability, distribution.

Although valuable tools, PROs suffer from three key limitations: 1) They are highly subjective and hard to generalize across patients [8]. 2) They lack sensitivity because humans are not tuned to quantify small changes in their condition [8, 9]. This is compounded by the fact that itching and scratching are often subconscious and may occur during sleep, and hence patients may not be fully aware of the morbidity associated with their chronic itch. 3) Such self-assessment is unreliable in children and older adults with cognitive impairment, yet itch is common in those populations [34, 16].

In this work, we do ask the recruited patients to report their peak NRS, however, this was done only for the sake of comparing our extracted scratching metrics with it. PROs provide a subjective evaluation of itch, while our system tries to objectively quantify and measure scratching.

2.2 Wearables

Throughout the years, there has been several attempts at using wearable wrist actigraphy or accelerometers to capture nocturnal scratching [8, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 22]. Despite these efforts, actigraphy methods usually have poor accuracy as patients may scratch with either hand, or even their feet, making it impossible to capture these scratching events [34, 45]. Additionally, subjecting chronic itch patients to wearing wrist actigraphy devices can potentially irritate existing skin conditions [15], as the wrist area is very often affected by conditions such as atopic dermatitis. Furthermore, wearables are hard to enforce in children and cognitively impaired patients.

2.3 Cameras

The golden standard for quantifying nocturnal scratching is certainly through infrared cameras [15, 13, 14]. However, it's easy to understand that such an approach raises privacy concerns, as an infrared camera needs to be installed in the patient's bedroom, and a human labeler is tasked with watching and annotating scratching occurrences. This brings up another issue with this approach, which is the huge overhead in time and cost incurred by the need to watch and annotate full nights of infrared footage. Also, poor infrared lighting conditions and the fact that patients often scratch with their back turned to the camera, or under the blankets, makes it hard for human labelers to capture all scratching events taking place.

Our approach tries to achieve the accuracy and objectiveness of measuring nocturnal scratching through infrared videos, but without the associated privacy invasion, and huge manual effort required to watch and annotate them.

Chapter 3

Method

We have designed a system that allows for detecting and quantifying nocturnal scratching in a contactless manner, without requiring people to wear sensors on their bodies. Our has two components: an off-body radio sensor and a machine learning model, which we describe below.

3.1 Off-body radio sensor

For our sensing needs, we utilize a radio sensor that has been developed in the NET-MIT group (depicted in Fig. 1-2). The radio device is placed on the wall, facing the subject's sleeping location. A low power radio signal is transmitted (1000 times weaker than that of a typical home Wi-FI device), and its reflections are captured and processed to be used as input to our machine learning model. The main technique used to analyze the collected radio signals is called Frequency Modulated Continuous Wave (FMCW) [25, 26]. This technique allows us to focus on a specific location in the room (e.g., the subject's sleeping location) and discard other sources of reflections (e.g., fans, roommates, etc.).

3.2 The machine learning model

We propose a neural network that takes as input RF signals (coming from the subject's bed location), and produces a time series of predictions that characterize each time instance to one of the following classes: static, scratching, or motion (referring to any movement other than scratching). The proposed model consists of three stages: a feature extractor, an encoder, and a decoder (see Fig. 3-1).

Feature extractor: The feature extractor aims to capture temporal information and identify repetitive motions. We use the knowledge that scratching itself is a repetitive pattern, hence providing a strong bias for the model. This stage accepts as input windows of radio signal, with a duration of one minute. The given window is processed through several 1D convolutional layers (7 layers). This outputs a time series of feature vectors, which we then correlate (using cosine similarity) to create a temporal similarity matrix (TSM). The TSM is responsible for emphasizing repetitive motions (e.g., scratching). An example TSM is visualized in Fig. 3-2 (**a**).

Encoder: The encoder takes as input the extracted temporal similarity matrix from the feature extractor, and outputs a feature pyramid. This feature pyramid gives a high-level description of the input (TSM) at different resolutions (scales). The encoder is based on the EfficientNet architecture, adjusted to operate on 1D data [46].

Decoder: The decoder takes as input the feature pyramid of the encoder, and outputs a time series of predictions that characterize each time instance as one of the

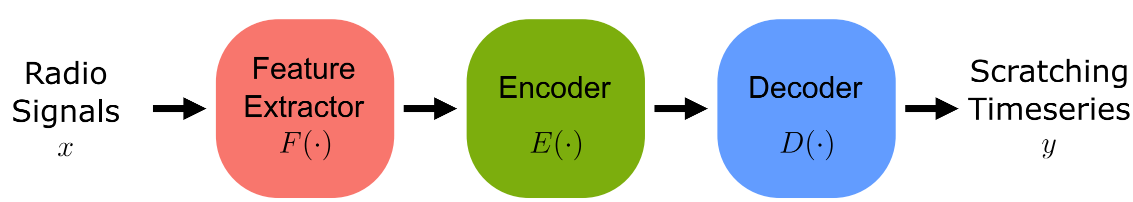

Figure 3-1: **High-level overview of the proposed neural network architecture.** The model consists of three stages: a feature extractor, an encoder, and a decoder. The model takes as input radio signals that are reflected from the subject's bed location, and outputs a time series of scratching predictions.

following classes: static, scratching, motion. The length of the output time series is the same as that of the original radio signal input (one minute). The decoder is based on the UNet architecture [47].

3.3 Training details

We trained the model using cross-entropy loss and Adam optimizer. The model was implemented in PyTorch [48]. The weights of the model are randomly initialized at the start of the training. The batch size was set to 32. The model was trained for 30K iterations on eight NVIDIA GeForce RTX 2080 Ti graphical processing units with distributed data parallelization. The model was trained for 30K iterations. The initial learning rate was set to 1e-3 for the feature extractor and 5e-3 for the encoder and decoder. A 10x learning rate decay was applied (to all three stages) after 5K, 10K and 20K iterations.

3.4 Model interpretability

For the design of the proposed neural network, we make use of the fact that scratching is a repetitive behavior. Hence, we choose to focus the first stage of the model on extracting temporal similarities in the input signal, which would correspond to repetitive motions. As already discussed, the first stage produces a temporal similarity matrix (TSM). This matrix highlights instances when the model discovers repetitive patterns, thus giving us the ability to interpret the model's predictions.

In Fig. 3-2 we show an example of a TSM, the matching model predictions, as well as the ground truth scratching annotations. The horizontal axis of Fig. 3-2 (**a**) corresponds to the time axis, while the vertical axis is a shorter (myopic) time axis that indicates how far ahead we are looking (relative to the position on the horizontal time axis), to uncover repetitive patterns (i.e., scratching). The horizontal black and yellow stripe pattern in the TSM indicates a repetitive motion, and the separation between them is equal to the repetition interval (i.e., how fast the subject scratches).

For the specific example shown in Fig. 3-2, the subject is initially static. This is also shown in the ground truth activity annotations. A static subject results into a static scene, which in turn creates high correlation over time. This matches with the bright yellow patch at the left side of the visualized TSM. Next, the subject moves their arm to begin the scratching motion. This brief motion is not correlating with past or future time instances, which results to a darker triangular patch appearing (immediately after the initial yellow patch). Then, the patient starts performing the scratching action. Due to the repetitive nature of scratching motions, e.g., hand moving up and down, the temporal similarity matrix shows horizontal black and yellow stripes. The separation of the stripes indicates that the scratching cycles is about 0.5 seconds long. After this scratching event, the subject returns their arm back to initial resting position. Similarly, this movement creates a darker patch in the TSM (after the first black and yellow stripe pattern). Finally, the patient stays still for a moment (small, yellow triangle in the middle the TSM), and then they perform a second scratching event.

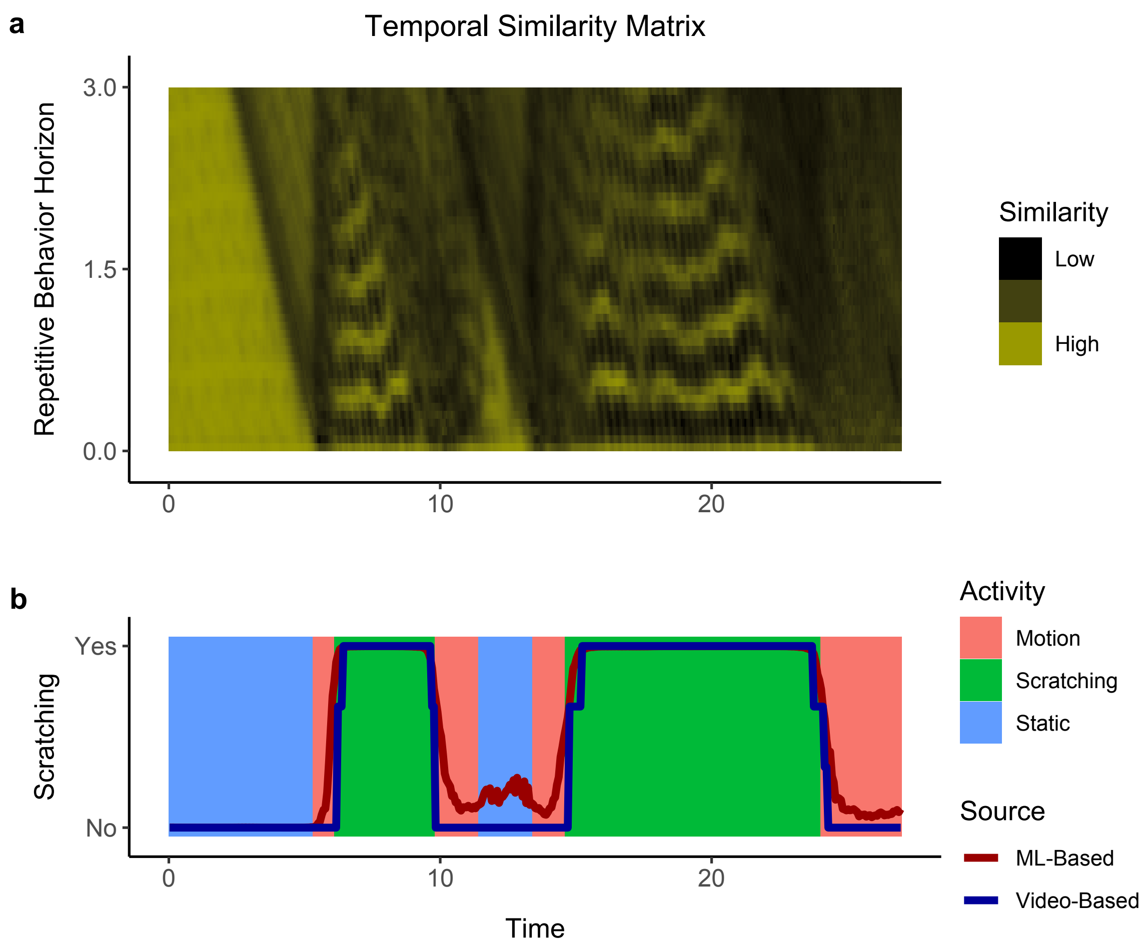

Figure 3-2: **Model interpretability.** This figures visualizes a 26-second example during which the subject scratches twice. (**a**) shows the temporal similarity matrix, which highlights repetitive patterns (e.g., scratching). The horizontal axis is the time axis, while the vertical axis a shorter time axis that corresponds to how far we are looking in the future to discover those repetitive patterns. Bright yellow means high similarity, while black means low similarity. (**b**) Shows the corresponding model predictions (red line), the human of scratching (blue line), as well as some activity labels based on what type of movement the subject performs (pastel colored rectangles)

Chapter 4

Results

4.1 Dataset

We collected data from 31 individuals in the span of 11/19/2019 to 04/06/2022; 11 of them were healthy control subjects, and 20 were patients of chronic pruritus associated with atopic dermatitis (AD), prurigo nodularis (PN), or chronic pruritus of uknown origin (CPUO). All of the 20 recruited patients have reported itching for at least 6 weeks prior to the study enrollment. We offer demographic and clinic characteristics of the two groups (control vs. chronic pruritus) in Table 4.1. The control individuals were selected so that race and gender are evenly balanced, and their mean age was 54.9 years old. In the chronic pruritus group, 70% were women, 80% were white, and the mean age was 57.5 years old. The baseline mean numerial rating scale (NRS) itch score reported was 0 and 7.8 for control and patients, respectively.

Chronic itch participants were monitored in-home for 4 weeks using both the aforementioned radio device and an infrared camera. These two devices were installed in the patients' bedrooms. They were asked to collect video data for at least 4 nights per wekk, and also to report their daily NRS and sleep scores for each night (on a scale of 0 to 10). Data were analyzed only for nights for which both the infrared camera and radio device were turned on. In total, we analyzed 370 nights, including 24,780 occurrences of scratching.

Table 4.1: **Demographic and clinical characteristics.**

Group	**Control**, N = 11[1]	**Pruritus**, N = 20[1]
Age	54.9 (17.3)	57.5 (16.1)
Gender		
Female	6 (55%)	14 (70%)
Male	5 (45%)	6 (30%)
Race		
Asian	3 (27%)	0 (0%)
Black	3 (27%)	4 (20%)
White	5 (45%)	16 (80%)
Diagnosis		
AD	0 (0%)	4 (20%)
PN	0 (0%)	2 (10%)
CPUO	0 (0%)	14 (70%)
NRS	0.0 (0.0)	7.8 (2.0)

[1]Mean (SD); n (%)

The control individuals were monitored for a minimum of 2 weeks, using only the radio device. In total, we collected data from 190 nights. Note that we didn't deploy any infrared cameras with these individuals.

To evaluate the accuracy of our model on quantifying scratching using RF data, we compare its output with the ground truth annotations (from humans watching the IR footage).

In order to extract the annotations of scratching instances, we developed a labeling server. Once a labeler logs into the system, they are served video clips (10s of seconds each), and they are asked to watch the videos and annotate the start and end of each scratching event. Naturally, the videos have been de-identified by blurring the face of the participants. An example of the developed labeling server can be see in Fig. 4-1.

Figure 4-1: **An example of the labeling server developed for annotating scratching events based on the collected infrared footage.**

The resulting data are split into training sets. The data from each subject appear exclusively to either training or testing set. A 4-fold cross-validation strategy is used to train and test the model.

4.2 Evaluation of the Scratching Model

The scratching model's ability to quantify scratching was evaluated using two metrics: a) scratching time per hour (STH), and b) scratching bouts per hour (SBH). Both STH (R = 0.97, p < 2.2e-16) and SBH (R = 0.93, p < 2.2e-16) as measured by the radio wave device correlated very strongly and significantly with the video-based manually annotated scratching assessments (Fig. 4-2 (**a**), (**b**)). Additionally, we investigate the exact matching between the model's predicted time series of scratching and the manual human annotations of scratching from the video footage.

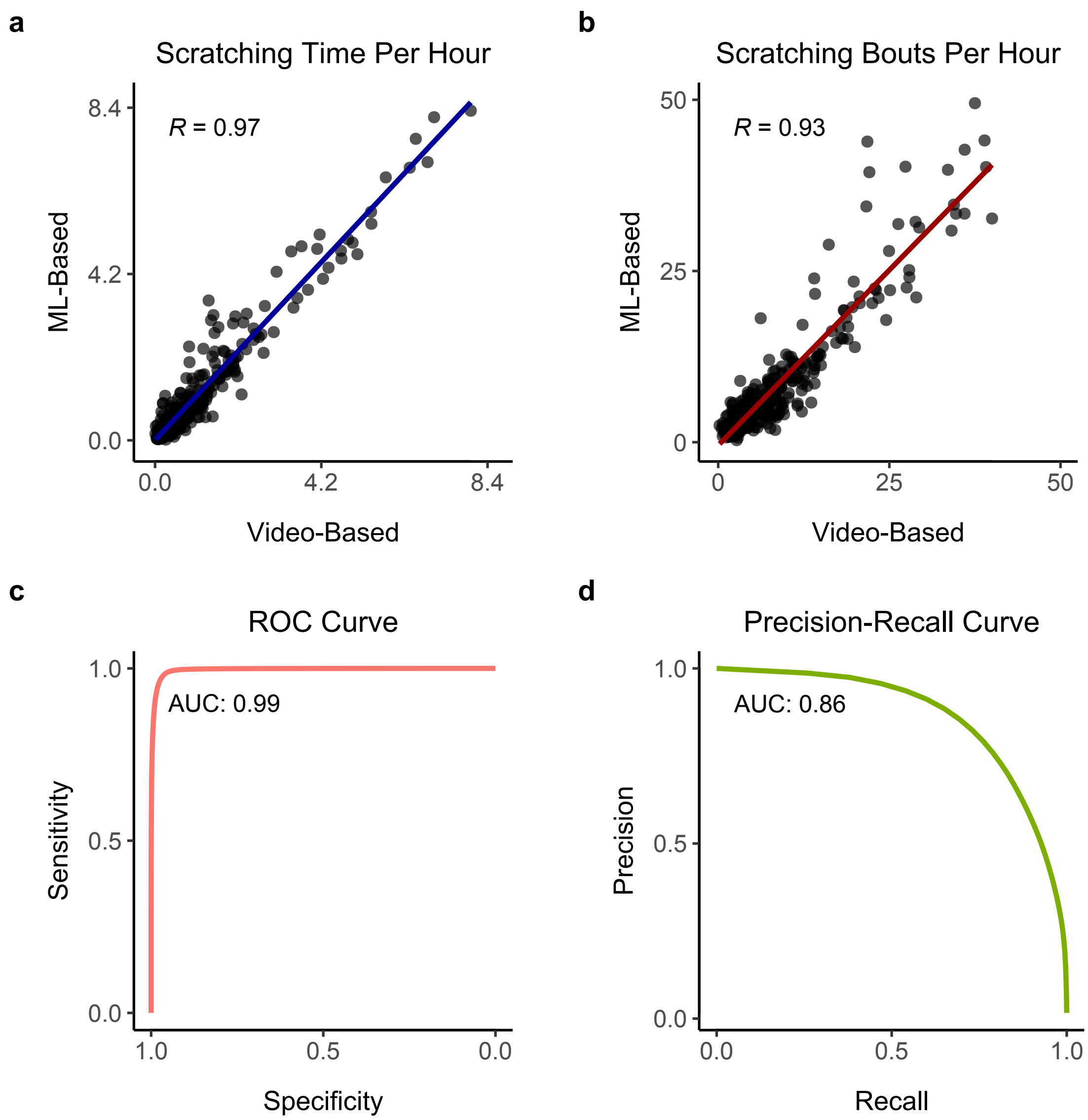

Figure 4-2: **Performance evaluation of the ML-based scratching model.** (**a**), (**b**) Investigate the scratching time per hour, and the scratching bouts per hour, respectively, as estimated by our ML-based approach compared to the gold standard video-based human annotations. Each point in these plots corresponds to a full night of data (n = 20, 370 nights). Very strong and significant correlation is achieved for both metrics (R = 0.97, $p <$ 2.2e-16 for scratching time per hour, and R = 0.93, $p <$ 2.2e-16 for scratching bouts per hour). (**c**) Shows the receiver operating characteristic curve, along with the area under the curve (AUC of 0.9957). (**d**) Shows the precision-recall curve, along with the area under the curve (AUC of 0.8595).

The model's predictions match the video annotations with a receiver operating characteristic (ROC) area under the curve (AUC) of 0.9956 (Fig. 4-2 (**c**)) and a precision-recall AUC of 0.8595 (Fig. 4-2 (**d**)). For a selected threshold of 0.665, the model achieves a specificity of 99.693% (95% CI [99.692%, 99.694%]), sensitivity of 75.574% (95% CI [75.518%, 75.629%]), precision of 79.798% (95% CI [79.745%, 79.851%]), and F1-score of 77.628%. These results corroborate the ability of the model to detect time-instances of scratching with high accuracy, across the population.

We now seek to measure the model's ability to effectively assess the scratching activity on a per-individual basis. This would be helpful not only for clinical trials, but also for informing both clinician and patient about the individual burden on chronic itch. To examine this, we plot each patient on a different plot, with the patients sorted in decreasing order of their maximum scratching. Additionally, different chronic pruritus diagnoses are color-coded (AD - red, PN - green, CPUO - blue). Figure 4-3 and Fig. 4-4 show that the model can strongly correlate with the ground truth assessments of STH and SBH ($p < 0.05$), even when focusing on a single individual (i.e., $n = 1$). All in all, these results demonstrate that our model can accurately and precisely quantify scratching activity, and therefore infer nocturnal itch burden in an objective, privacy preserving, and touchless manner.

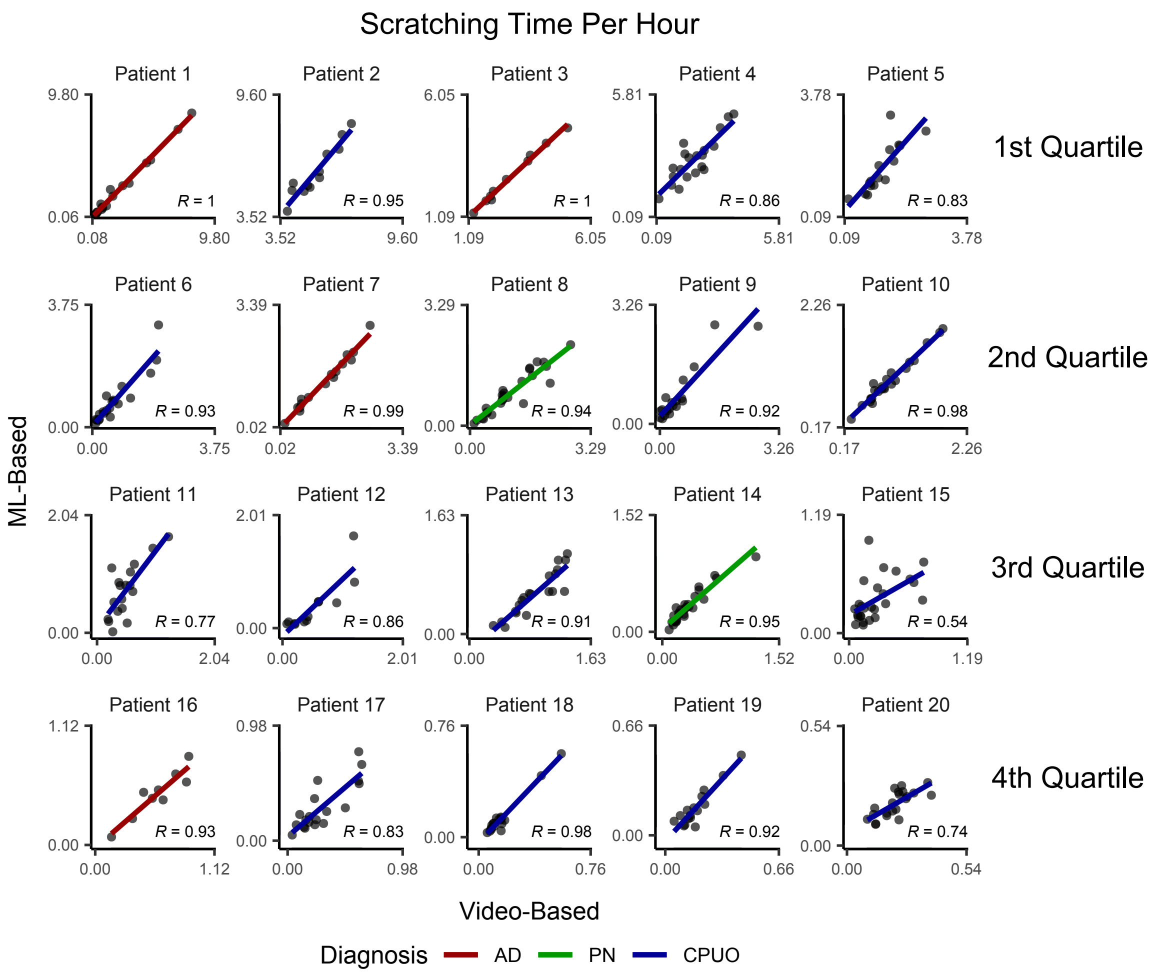

Figure 4-3: **Per subject accuracy evaluation at measuring scratching time per hour.** The model can accurately infer the scratching time per hour across subjects with varying disease severity. Strong Pearson correlation (with $p < 0.05$) is shown between our ML-based predictions and the ground truth from video annotations across all 20 patients.

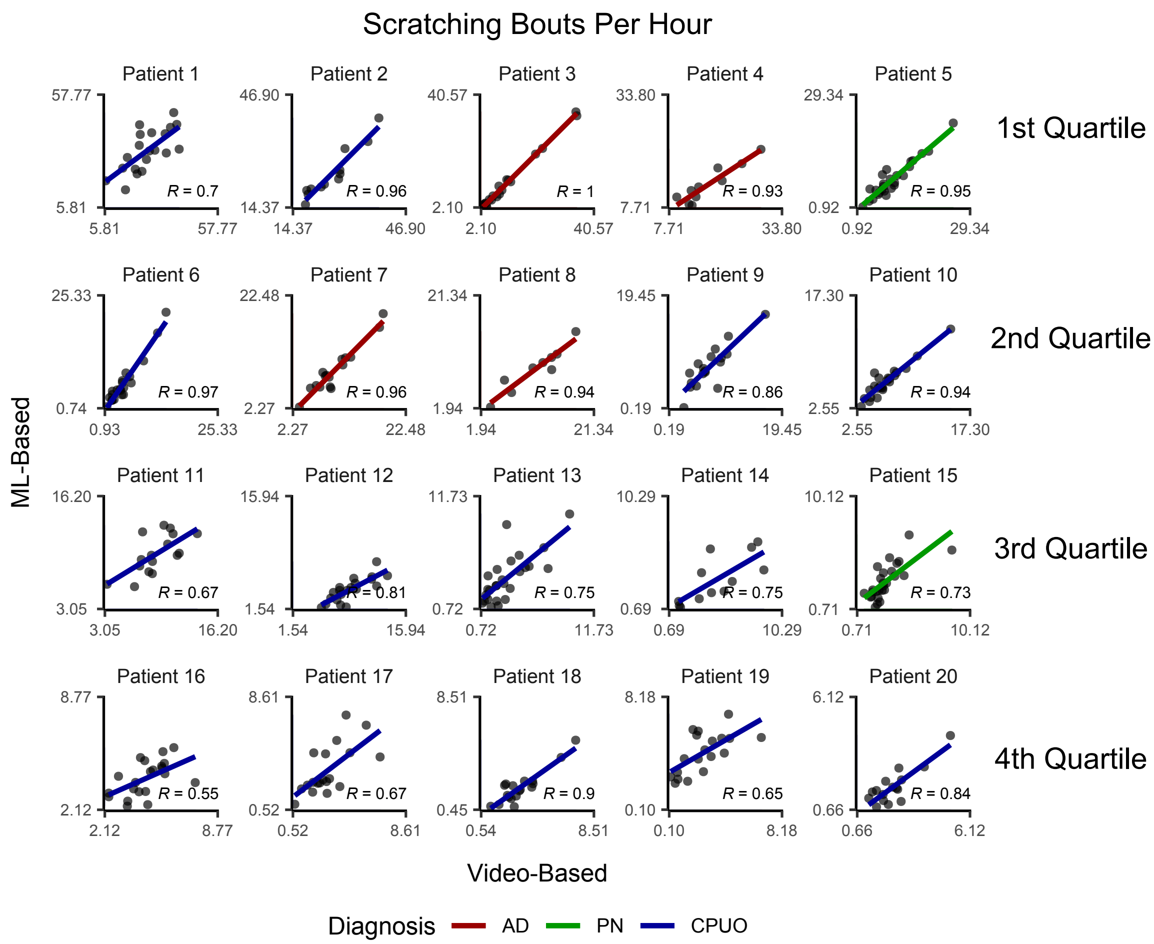

Figure 4-4: **Per subject accuracy evaluation at measuring scratching bouts per hour.** The model can accurately infer the scratching bouts per hour across subjects with varying disease severity. Strong Pearson correlation (with $p < 0.05$) is shown between our ML-based predictions and the ground truth from video annotations across all 20 patients.

4.3 Impact of Itch on Sleep

We seek to quantify the impact of itch on sleep quality. Past work has demonstrated the feasibility to extracting accurate sleep information from radio signal using a device similar to the one used in this study. Thus in this section we use the same wireless signal collected by the device to both infer sleep quality (using the model by Zhao et al) and scratching (using the model developed herein). We compare various sleep quality metrics against the amount of scratching on a per night basis ($n = 19$, 256 nights). Note that one patient is removed from this analysis as they would wake up in the middle of night and sleep in a different room. Additionally, we exclude nights for which the metrics we investigate are zero (this is because of the logarithmic scale used for the analysis). We consider three sleep metrics: a) Sleep Efficiency which is the ratio of time spent sleeping to the total time spent in bed, b) Sleep Latency, which is the time spent awake before sleep onset, and c) WASO (wake after sleep onset) which is the amount of awake time after sleep onset. Figures 4-5 (**a**), (**b**), (**c**) plot the correlation between these sleep metrics and scratching on a logarithmic scale. We chose a log scale since our data shows that the relationship is non-linear. The figure reveals that all sleep metrics are impacted by scratching, i.e., the more the person scratches in a particular night, the lower their sleep efficiency ($R = -0.62$, $p < 2.2e\text{-}16$), the longer their sleep latency ($R = 0.66$, $p < 2.2e\text{-}16$), and the awake time they have at night ($R = 0.7$, $p < 2.2e\text{-}16$).

We also investigate whether scratching occurs only while the person is awake or it also happens during their sleep. For each night, we consider the percentage of scratching in each sleep stage: wake, light sleep, deep sleep, and REM. Fig. 4-5 (**d**) shows boxplots of the scratching time as a function of the sleep stage. On average, 92.08% of the scratching took place while in wake, 5.15% in the light sleep, 0.83% in deep sleep and 1.91% in REM. These results highlight the fact that scratching can be a conscious act as well as reflex that occurs subconsciously.

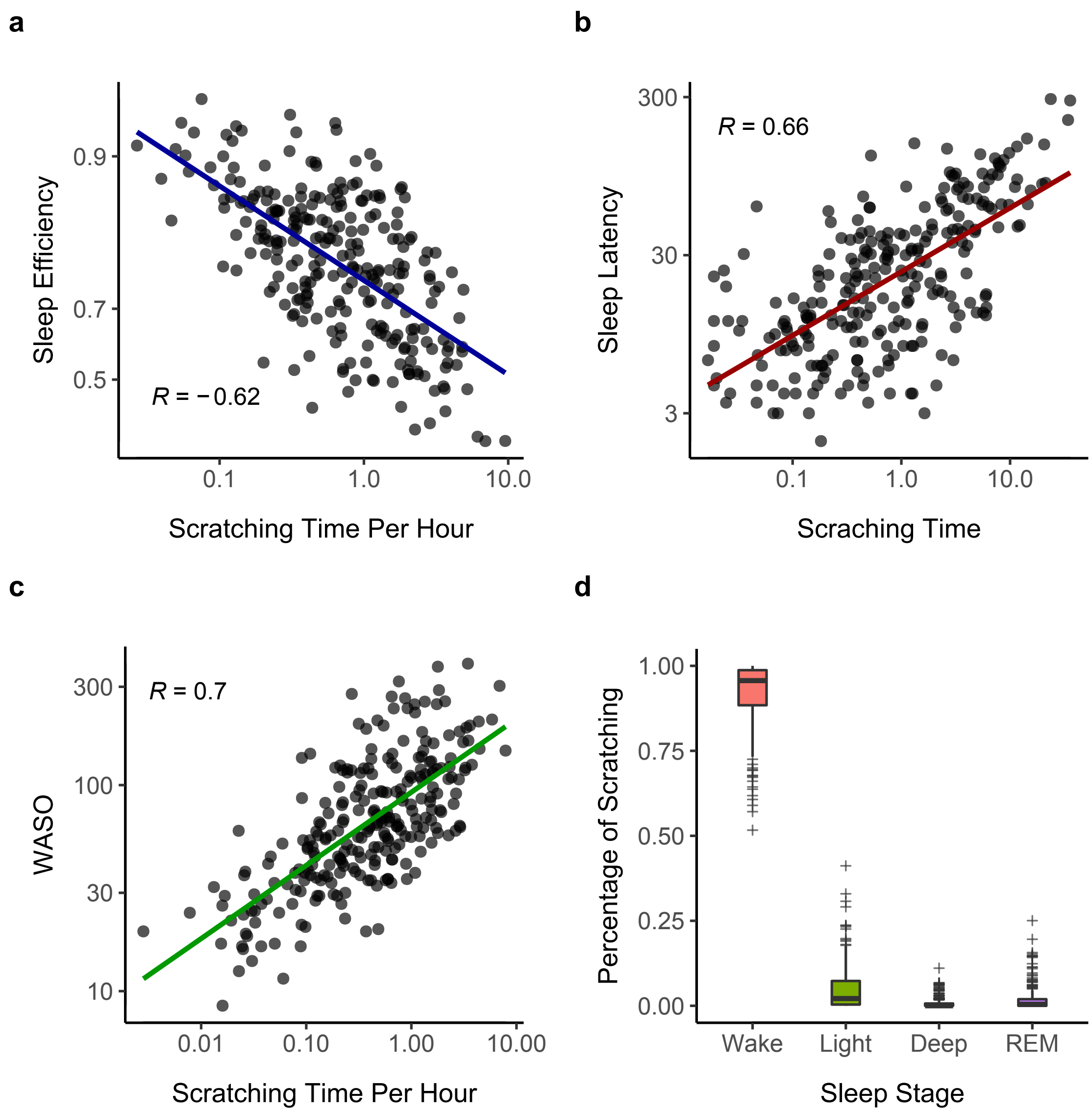

Figure 4-5: **Impact of itch on sleep quality.** This figures examines the impact of itch on sleep quality (n = 19, 256 nights). (**a**) Compares the sleep efficiency with the predicted scratching time per hour. Sleep efficiency is defined as the ratio of time spent asleep over the total time spent trying to sleep. Significant Pearson's correlation is shown between the two metrics (R = -0.62, $p < 2.2e\text{-}16$). (**b**) Compares the sleep latency with the predicted scratching time for all subjects, and nights. Sleep latency is defined as the time spent awake until sleep onset. Significant Pearson's correlation is shown between the two metrics (R = 0.66, $p < 2.2e\text{-}16$). (**c**) Compares WASO (wake after sleep onset), and scratching time per hour. Significant Pearson's correlation is shown between the two metrics (R = 0.7, $p < 2.2e\text{-}16$). Both axes are in logarithmic scale for (**a**), (**b**) and (**c**) . (**d**) Shows the percentage of scratching per sleep stage, for each night. In each box plot, the central line indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles respectively. The whiskers extend to 1.5 times the interquartile range. Points beyond the whiskers are plotted individually using the cross symbol.

4.4 Objective Scratching and Perception of Itch

We compared nocturnal scratching with the numerical rating scale (NRS) reported by the patients for nights that have NRS data and scratching measurements from the camera and the radio device ($n = 20, 340$ nights). Figures 4-6 (**a**)-(**b**) show, for each night, the NSR value and the corresponding scratching from the video annotations and the ML model, respectively. This result show that the NRS has only a mild correlation with nocturnal scratching (Pearson correlation is $R = 0.11$, $p = 0.048$ for the ground truth videos and $R = 0.12$, $p = 0.031$ for the ML model).

Next, we compare the change in scratching and NRS before and after receiving a dose of treatment with Dupimulab 300mg. Only one participant in our population started treatment during the study. The participant received the first dose of Dupimulab 300mg immediately before the beginning of the monitoring period. We compared the participant's status during the two weeks before and after the second drug administration. As shown in Figs. 4-6 (**c**), (**d**), there is a significant reduction in scratching time and bouts per hour after drug administration ($p = 4.16\text{e-}04$, $p = 6.07\text{e-}03$, respectively). Similarly, there is also a significant reduction in self-reported NRS ($p = 4.43\text{e-}04$), as shown in Fig. 4-6 (**e**). While this result is only for one patient, it exhibits statistical significance. The results also provide indication that scratching metrics show response to treatment.

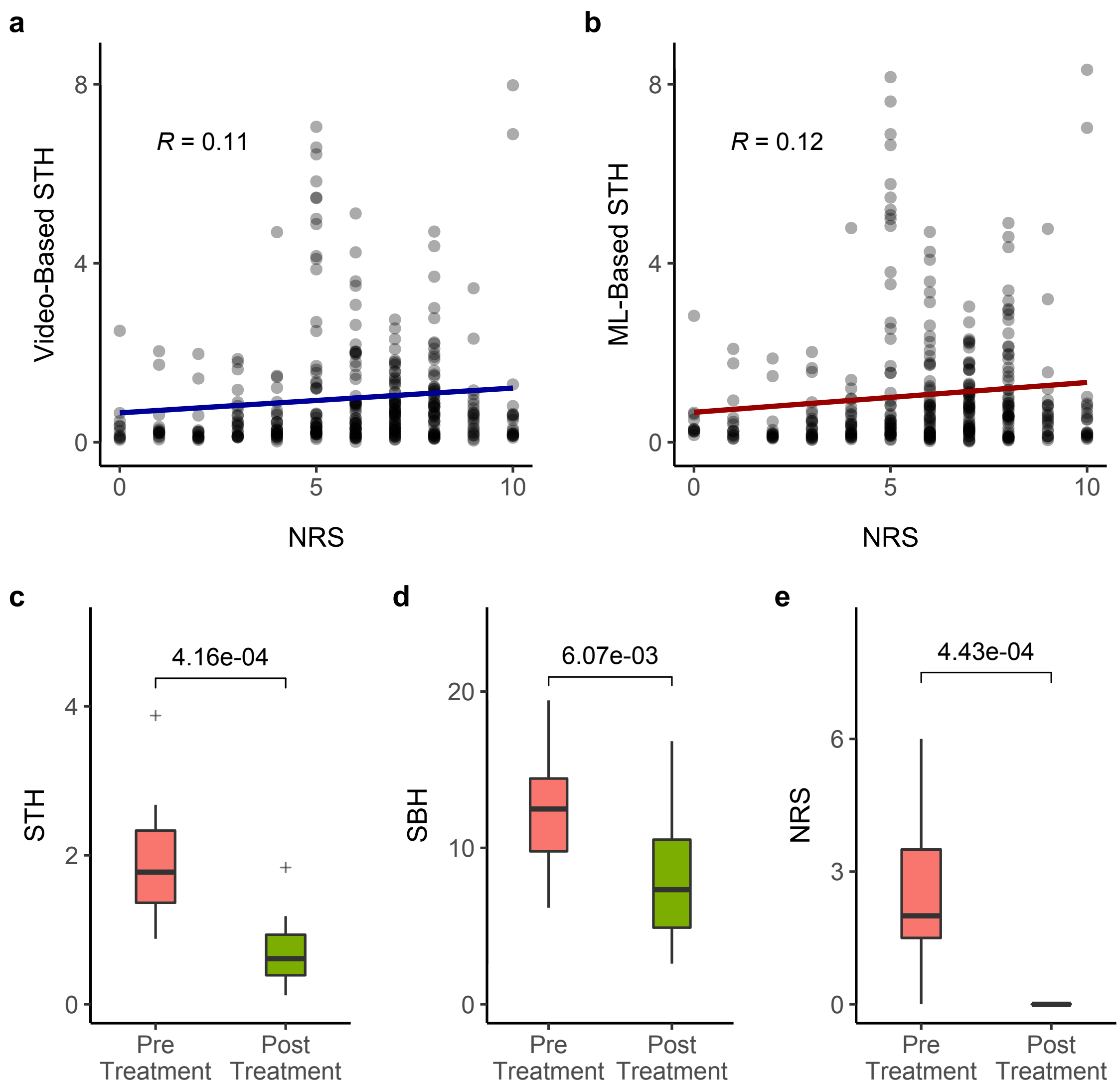

Figure 4-6: **Objective scratching vs. subjective perception of itch** This figures examines the relationship between the objective measurement of scratching and the subjective self-reported perception of itch. (**a**), (**b**) Show scatter plots of the scratching time per hour (from video and predictions, respectively) compared with the self-reported NRS of each patient (n = 20, 340 nights) Each point corresponds to a full night of data. In both cases we observe significant ($p < 0.05$) but weak correlation ($R = 0.11$ and $R = 0.12$, respectively). (**c**), (**d**), (**e**) Show the difference in inferred STH, SBH, and self-reported NRS before and after drug administration, dupimulab 300mg (n = 1, 27 nights). There is statistical difference for all 3 of them ($p = 4.16\text{e-}04$, $p = 6.07\text{e-}03$, $p = 4.43\text{e-}04$, respectively), however, the patient reports a constant NRS of 0 after the treatment, while there is still significant amount of scratching each night. On each box plot, the central line indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles respectively. The whiskers extend to 1.5 times the interquartile range. Points beyond the whiskers are plotted individually using the cross symbol.

4.5 Comparison between healthy control and chronic itch patients

In this section, we seek to compare the scratching amount and sleep quality between the chronic itch patients (n = 19, 337 nights) and age- and gender-matched healthy control. We discover in Fig. 4-5 (**a**), (**b**) that the scratching metrics (STH and SBH) are significantly elevated for the chronic pruritus patients, relative to the healthy control ($p = 4.05e\text{-}22$ and $1.21e\text{-}20$, respectively). Additionally, according to Figs. 4-5 (**c**), (**d**), (**e**) chronic itch patients demonstrate decreased sleep efficiency ($p = 3.17e\text{-}04$), increased sleep latency ($p = 7.17e\text{-}06$) and increased WASO ($p = 5.94e\text{-}06$).

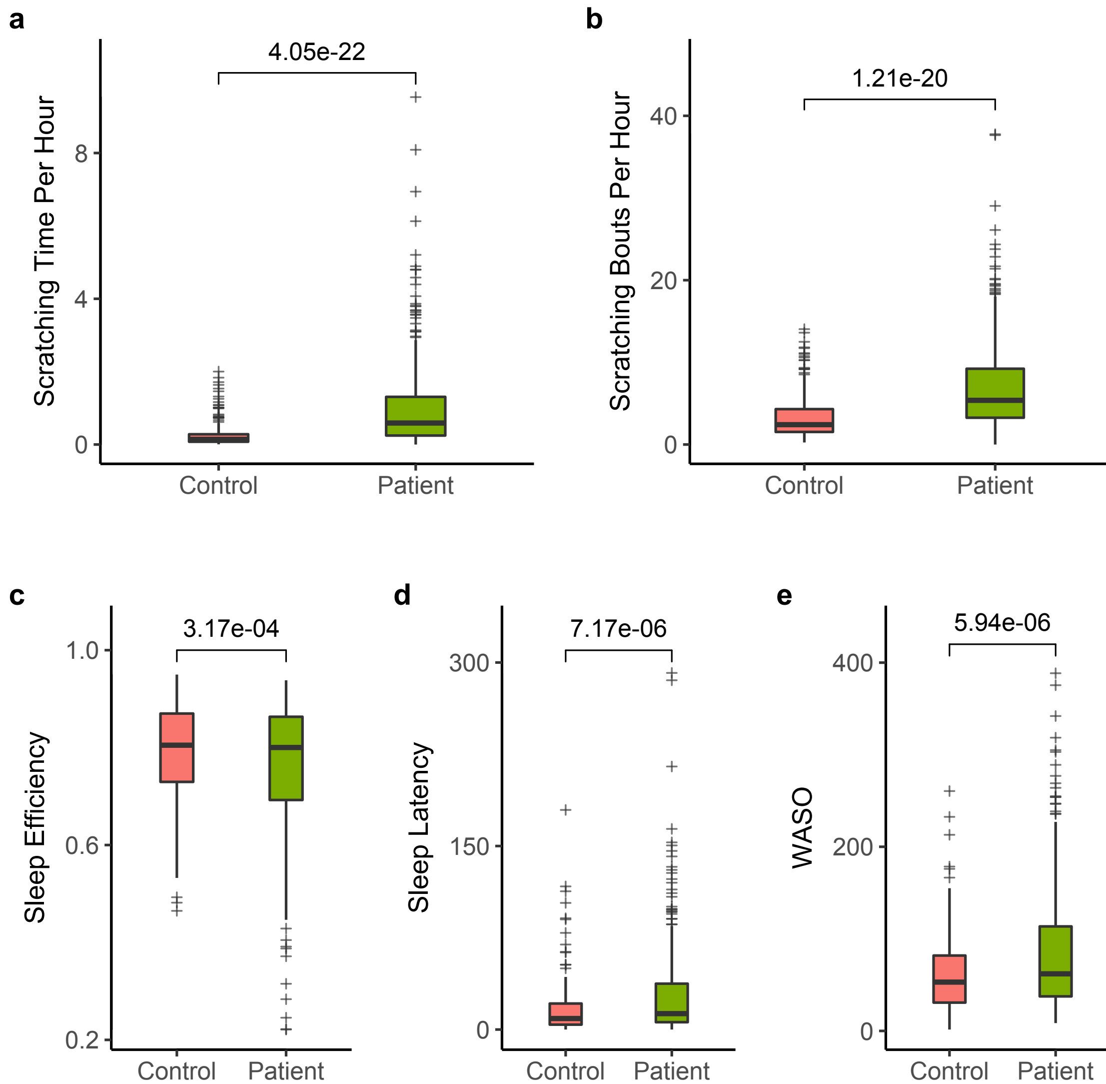

Figure 4-7: **Comparison between healthy control and chronic itch patients.** This figures examines the difference in scratching and sleep between control (n = 11, 190 nights) and chronic itch patients (n = 19, 337 nights). (**a**), (**b**) Compare the predicted scratching time per hour and predicted bouts per hour between the two groups. In both cases the chronic patients demonstrate significantly elevated scratching (p = 4.05e-22 and p = 1.21e-20, respectively). (**c**), (**d**), (**e**) Compare the sleep efficiency, sleep latency and WASO between the two groups. In all three cases, the chronic itch patients demonstrate deterioration in sleep quality (p = 3.17e-04, p = 7.17e-06 and p = 5.94e-06, respectively). For each box plot, the central line indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles respectively. The whiskers extend to 1.5 times the interquartile range. Points beyond the whiskers are plotted individually using the cross symbol.

Chapter 5

Conclusion

In this work, we collected a large nocturnal scratching dataset, consisting of 370 nights of infrared video footage, radio signals and, human annotations of scratching events. This dataset was then used to train and test a neural network capable of predicting nocturnal scratching using only radio signals, in an objective, sensitive and privacy-preserving way. We conducted analysis to assess the model's accuracy with respect to manual annotations of scratching, and demonstrate that the model can achieve very good results. Next, we showed that the proposed model can extract meaninful scratching metrics that correlate with the patients' perception of itch, while also providing promising initial results about the ability to track disease progression/medication efficacy in a sensitive and quantifiable way. Further, we utilized prior art on extracting breathing signals sleep stages from radio signals, to study the negative impact of nocturnal scratching on sleep quality. Finally, we demonstrated the increased scratching amount and deteriorated sleep quality of chronic itch patients by comparing them with a group of healthy age- and gender-matched individuals.

Despite all the advantages of our proposed solution and improvements over prior work, we also note that it has some limitations. First, our system was trained and tested only on 20 chronic itch patients. We understand that there might be scratching patterns or certain activities people do while in bed at night, that we haven't encountered in these 370 nights of collected data. Second, our model was trained with the assumption that patients sleep alone in their beds. This limits the model to clinical trials in which the participants are willing and able to sleep by themselves for the duration of the monitoring period. We anticipate however that future work can extend the model to deal with scenarios in which the participants may share the bed with a partner or even a pet. This is similar to earlier work on monitoring nocturnal breathing from radio signals, where earlier models were limited to one individual in bed [27] but later models allowed for monitoring breathing even when the person has a bed partner [29]. Finally, pets can be problematic. Pets often sleep in the same room as the patient, or even in the bed right next to the patient. This can create confusions for the breathing extraction and sleep staging part of the analysis. Additionally, dogs and cats tend to scratch, and this could also potentially confuse the model into thinking that the patient is scratching. For this work, people having pets was one of the exclusion criteria, thus the model hasn't seen any such examples.

In conclusion, we develop a system to quantify nocturnal scratching in an accurate, objective, sensitive and privacy preserving way. The model allows us to extract meaningful metric about scratching, that not only correlate with patients' perception of their itch but can also potentially track disease progression. The system also allows us to understand the impact of scratching on sleep quality. We believe that our system will have significant clinical implications. One major setback for drug development and treatment is the lack of quantifiable, objective and sensitive metrics for itch. We hope that with our solution we can provide a way for clinicians and researchers to track the disease progression of patients and other parts of the physiology of the patient, such as various sleep quality metrics, so that they can assess treatment efficiency, provide new prescriptions, or study the effects of a new drug.

Bibliography

[1] Raghav Tripathi, Konrad D Knusel, Harib H Ezaldein, Jeremy S Bordeaux, and Jeffrey F Scott. The cost of an itch: A nationally representative retrospective cohort study of pruritus-associated health care expenditure in the united states. *J. Am. Acad. Dermatol.*, 80(3):810–813, March 2019.

[2] Irem Eren Erturk, Ozer Arican, Imran Kurt Omurlu, and Necdet Sut. Effect of the pruritus on the quality of life: a preliminary study. *Ann. Dermatol.*, 24(4):406–412, November 2012.

[3] Jon Anders Halvorsen, Florence Dalgard, Magne Thoresen, Magne Thoresen, Espen Bjertness, and Lars Lien. Itch and mental distress: a cross-sectional study among late adolescents. *Acta Derm. Venereol.*, 89(1):39–44, 2009.

[4] Seema P Kini, Laura K DeLong, Emir Veledar, Anne Marie McKenzie-Brown, Michael Schaufele, and Suephy C Chen. The impact of pruritus on quality of life: the skin equivalent of pain. *Arch. Dermatol.*, 147(10):1153–1156, October 2011.

[5] U.S. Food, Center for Drug Evaluation Drug Administration, and Research. Reference id: 4828760, July 2021. https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/215498s000lbl.pdf.

[6] Jonathan I Silverberg. Public health burden and epidemiology of atopic dermatitis. *Dermatol. Clin.*, 35(3):283–289, July 2017.

[7] G Yosipovitch, M Reaney, V Mastey, L Eckert, A Abbé, L Nelson, M Clark, N Williams, Z Chen, M Ardeleanu, B Akinlade, N M H Graham, G Pirozzi, H Staudinger, S Plaum, A Radin, and A Gadkari. Peak pruritus numerical rating scale: psychometric validation and responder definition for assessing itch in moderate-to-severe atopic dermatitis. *Br. J. Dermatol.*, 181(4):761–769, October 2019.

[8] Caroline S Murray and Jonathan L Rees. Are subjective accounts of itch to be relied on? the lack of relation between visual analogue itch scores and actigraphic measures of scratch. *Acta Derm. Venereol.*, 91(1):18–23, January 2011.

[9] J A Savin, W D Paterson, I Oswald, and Kirstine Adam. Further studies of scratching during sleep. *Br. J. Dermatol.*, 93(3):297–302, September 1975.

[10] S. Elman, L.S. Hynan, V. Gabriel, and M.J. Mayo. The 5-d itch scale: a new measure of pruritus. *British Journal of Dermatology*, 162(3):587–593, December 2009.

[11] Stephen Erickson and Brian S Kim. Research techniques made simple: Itch measurement in clinical trials. *J. Invest. Dermatol.*, 139(2):264–269.e1, February 2019.

[12] Ngoc Quan Phan, Christine Blome, Fleur Fritz, Joachim Gerss, Adam Reich, Toshiya Ebata, Matthias Augustin, Jacek C Szepietowski, and Sonja Ständer. Assessment of pruritus intensity: prospective study on validity and reliability of the visual analogue scale, numerical rating scale and verbal rating scale in 471 patients with chronic pruritus. *Acta Derm. Venereol.*, 92(5):502–507, September 2012.

[13] Ebata, Aizawa, Kamide, and Niimura. The characteristics of nocturnal scratching in adults with atopic dermatitis. *Br. J. Dermatol.*, 141(1):82–86, July 1999.

[14] Toshiya Ebata, Hiroshi Aizawa, and Ryoichi Kamide. An infrared video camera system to observe nocturnal scratching in atopic dermatitis patients. *J. Dermatol.*, 23(3):153–155, March 1996.

[15] Mary Patricia Smith, Karen Ly, Quinn Thibodeaux, Thulasi Weerasinghe, Jashin J. Wu, Gil Yosipovitch, Tina Bhutani, and Wilson Liao. Emerging methods to objectively assess pruritus in atopic dermatitis. *Dermatology and Therapy*, 9(3):407–420, June 2019.

[16] Danny Camfferman, John D Kennedy, Michael Gold, Alfred J Martin, and Kurt Lushington. Eczema and sleep and its relationship to daytime functioning in children. *Sleep Med. Rev.*, 14(6):359–369, December 2010.

[17] Yoshikazu Gotoh, Yu Omori, Tsugunobu Andoh, and Yasushi Kuraishi. Tonic inhibition of allergic itch signaling by the descending noradrenergic system in mice. *J. Pharmacol. Sci.*, 115(3):417–420, March 2011.

[18] P E Beattie and M S Lewis-Jones. An audit of the impact of a consultation with a paediatric dermatology team on quality of life in infants with atopic eczema and their families: further validation of the infants' dermatitis quality of life index and dermatitis family impact score. *Br. J. Dermatol.*, 155(6):1249–1255, December 2006.

[19] Cheng-Hao Weng, Ching-Chih Hu, Tzung-Hai Yen, Ching-Wei Hsu, and Wen-Hung Huang. Uremic pruritus is associated with two-year cardiovascular mortality in long term hemodialysis patients. *Kidney Blood Press. Res.*, 43(3):1000–1009, June 2018.

[20] D J Buysse, C F Reynolds, 3rd, T H Monk, S R Berman, and D J Kupfer. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res.*, 28(2):193–213, May 1989.

[21] Anita L. Stewart. *Measuring functioning and well-being: The Medical Outcomes Study Approach*. Duke Univ. Press, 1998.

[22] Nikhil Mahadevan, Yiorgos Christakis, Junrui Di, Jonathan Bruno, Yao Zhang, E Ray Dorsey, Wilfred R Pigeon, Lisa A Beck, Kevin Thomas, Yaqi Liu, Madisen Wicker, Chris Brooks, Nina Shaafi Kabiri, Jaspreet Bhangu, Carrie Northcott, and Shyamal Patel. Development of digital measures for nighttime scratch and sleep using wrist-worn wearable devices. *NPJ Digit. Med.*, 4(1):42, March 2021.

[23] Bruce G Bender, Robert Ballard, Beth Canono, James R Murphy, and Donald Y M Leung. Disease severity, scratching, and sleep quality in patients with atopic dermatitis. *J. Am. Acad. Dermatol.*, 58(3):415–420, March 2008.

[24] Fadel Adib and Dina Katabi. See through walls with WiFi! In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*. ACM, aug 2013.

[25] Fadel Adib, Zachary Kabelac, and Dina Katabi. Multi-person localization via rf body reflections. In *Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation*, NSDI'15, page 279–292, USA, 2015. USENIX Association.

[26] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics*, 34(6):1–13, nov 2015.

[27] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, apr 2015.

[28] Tianhong Li, Lijie Fan, Mingmin Zhao, Yingcheng Liu, and Dina Katabi. Making the invisible visible: Action recognition through walls and occlusions, 2019.

[29] Shichao Yue, Hao He, Hao Wang, Hariharan Rahul, and Dina Katabi. Extracting multi-person respiration from entangled RF signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–22, jul 2018.

[30] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. Through-wall human pose estimation using radio signals. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2018.

[31] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. RF-based 3d skeletons. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. ACM, aug 2018.

[32] Mingmin Zhao, Shichao Yue, Dina Katabi, Tommi Jaakkola, and Matt Bianchi. Learning sleep stages from radio signals: A conditional adversarial architecture. 08 2017.

[33] A Reich, M Heisig, N Phan, K Taneda, K Takamori, S Takeuchi, M Furue, C Blome, M Augustin, S Ständer, and J Szepietowski. Visual analogue scale: Evaluation of the instrument for the assessment of pruritus. *Acta Dermato Venereologica*, 92(5):497–501, 2012.

[34] C I Wootton, K Koller, S Lawton, C O'Leary, K S Thomas, and on behalf of the SWET study team. Are accelerometers a useful tool for measuring disease activity in children with eczema? validity, responsiveness to change, and acceptability of use in a clinical trial setting. *Br. J. Dermatol.*, 167(5):1131–1137, November 2012.

[35] Carol Bringhurst, Karen Waterston, Olivia Schofield, Kenneth Benjamin, and Jonathan L Rees. Measurement of itch using actigraphy in pediatric and adult populations. *J. Am. Acad. Dermatol.*, 51(6):893–898, December 2004.

[36] Keum San Chun, Youn J Kang, Jong Yoon Lee, Morgan Nguyen, Brad Lee, Rachel Lee, Han Heul Jo, Emily Allen, Hope Chen, Jungwoo Kim, Lian Yu, Xiaoyue Ni, Kunhyuck Lee, Hyoyoung Jeong, Joohee Lee, Yoonseok Park, Ha Uk Chung, Alvin W Li, Peter A Lio, Albert F Yang, Anna B Fishbein, Amy S Paller, John A Rogers, and Shuai Xu. A skin-conformable wireless sensor to objectively quantify symptoms of pruritus. *Sci. Adv.*, 7(18), April 2021.

[37] Bruce G Bender, Susan B Leung, and Donald Y M Leung. Actigraphy assessment of sleep disturbance in patients with atopic dermatitis: an objective life quality measure. *J. Allergy Clin. Immunol.*, 111(3):598–602, March 2003.

[38] Akihiko Ikoma, Toshiya Ebata, Laurent Chantalat, Kimitoshi Takemura, Fabienne Mizzi, Michel Poncet, and Didier LeClercq. Measurement of nocturnal scratching in patients with pruritus using a smartwatch: Initial clinical studies with the itch tracker app. *Acta Derm. Venereol.*, 99(3):268–273, March 2019.

[39] Hiroyuki Fujita, Mayumi Nagashima, Yoshihiro Takeshita, and Michiko Aihara. Correlation between nocturnal scratch behavior assessed by actigraphy and subjective/objective parameters in patients with atopic dermatitis. *Eur. J. Dermatol.*, 24(1):120–122, January 2014.

[40] Takamasa Kogure and Toshiya Ebata. Activity during sleep measured by a sheet-shaped body vibrometer and the severity of atopic dermatitis in adults: A comparison with wrist actigraphy. *J. Clin. Sleep Med.*, 14(2):199–204, February 2018.

[41] C J Gustafson, J O'Neill, E Hix, D T McLaren, O M Buxton, and S R Feldman. Feasibility of actigraphy wristband monitoring of atopic dermatitis in children. *Skin Res. Technol.*, 20(4):510–514, November 2014.

[42] Johanna Feuerstein, Daniel Austin, Robert Sack, and Tamara L Hayes. Wrist actigraphy for scratch detection in the presence of confounding activities. *Annu Int Conf IEEE Eng Med Biol Soc*, 2011:3652–3655, 2011.

[43] Anna B Fishbein, Jennifer Lor, Frank J Penedo, Christopher B Forrest, James W Griffith, and Amy S Paller. Patient-Reported outcomes for measuring sleep disturbance in pediatric atopic dermatitis: cross sectional study of PROMIS pediatric sleep measures and actigraphy. *J. Am. Acad. Dermatol.*, June 2020.

[44] A Kalinska-Bienias, T Piotrowski, E Kowalczyk, A Lesniewska, M Kaminska, P Jagielski, C Kowalewski, and K Wozniak. Actigraphy-measured nocturnal wrist movements and assessment of sleep quality in patients with bullous pemphigoid: a pilot case-control study. *Clin. Exp. Dermatol.*, 44(7):759–765, October 2019.

[45] Albert F. Yang, Morgan Nguyen, Alvin W. Li, Brad Lee, Keum San Chun, Ellen Wu, Anna B. Fishbein, Amy S. Paller, and Shuai Xu. Use of technology for the objective evaluation of scratching behavior: A systematic review. *JAAD International*, 5:19–32, December 2021.

[46] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. 2019.

[47] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[48] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.