Proximity Tracing App

Report from the Measurement Campaign 2020-04-09

Jackie Ma, David Neumann, Felix Sattler, Ralf Schäfer, Patrick Wagner, Thomas Wiegand

Executive Summary

The main questions that is studied in this project is the following

Given a time series of physical distances between a diseased and a healthy person what is the probability that the diseased person transmitted the virus to the healthy person over the time period?

In order to answer this question we perform several measurement campaigns to develop and validate our proposed solutions.
This report contains a detailed description of the methods and results of the measurement campaign on 2020-04-09conducted at the barracks of the German Bundeswehr. The measurements were performed **using 28 Samsung A40 smartphones and 20 Android phone from different manufacturers** that were carried by 48 protected soldiers, respectively. The Samsung A40 smartphone have Bluetooth instabilities that effected the results slightly. This problem can be solved by a software patch. The soldiers also wear protection mask in order to avoid possible disease infections. The experiment took several hours and thousands of datapoints were collected as well as video material for the labeling process.
Based on the acquired data we aim to predict dangerous and non-dangerous contacts between two subjects. To this end, we have to define models that describe dangerous and non-dangerous contacts. We have considered different models based on *linear*-, *box*-, and *sigmoid*-functions. In fact, our methods are not limited to on any particular model. However, for this problem we have decided to address the epidemiological problem with an epidemiological model. More precisely, for this report we rely on the epidemiological model proposed by Fraser et al. [1] which uses a quadratic diffusion model combined with a risk score model that controls the number of notified individuals.
Assuming this epidemiological model our model achieves the following results:

- if we alert 3 people on average: True Positive Rate of 74 percent and an AUC of 0.94
- if we alert 6 people on average: True Positive Rate of 67 percent and an AUC of 0.89
- if we alert 9 people on average: True Positive Rate of 76 percent and an AUC of 0.88

[1] *Defining an epidemiologically meaningful contact from phone proximity events: uses for digital contact tracing*, Christophe Fraser, David Bonsall, Robert Hinch and Anthony Finkelstein, Draft, April 2020.

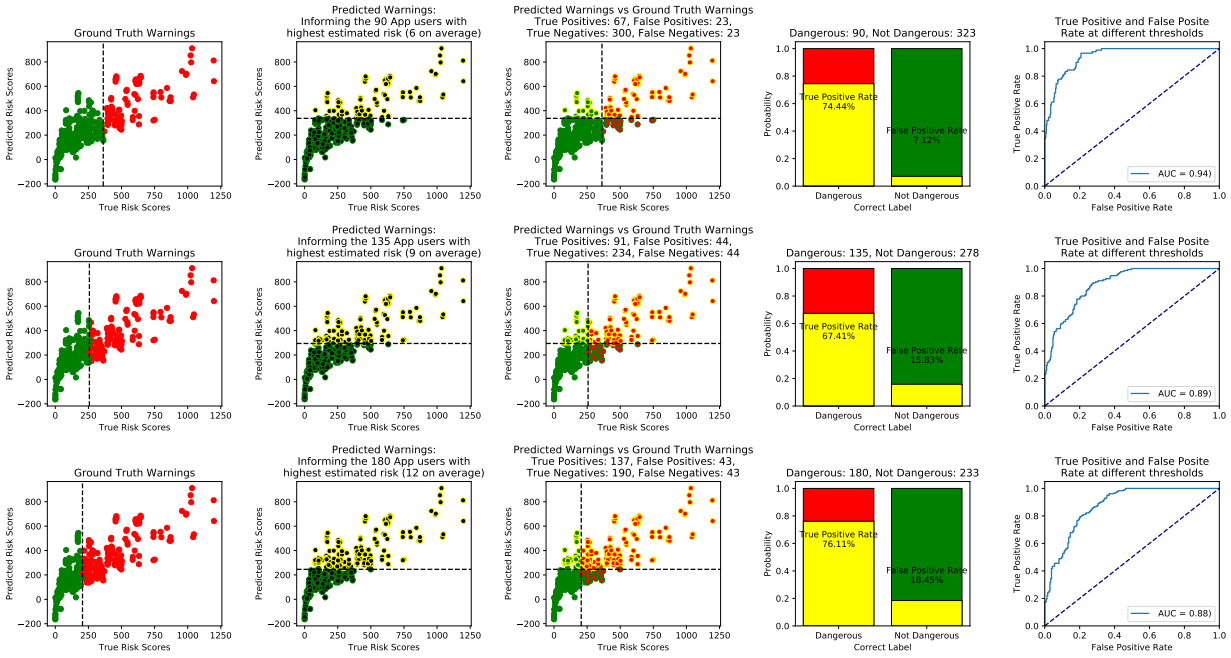

Figure 1: Visualization of the ground truth with respect to the epidemiological model and prediction accuracy. First row: 3 contacts will be notified on average. Second row: 6 contacts will be notified on average. Third row: 9 contacts will be notified on average.

Contents

Table 1: Experimental Setup Overview.

Bundeswehr Experiment 04-09:	
Short Description: Test with devices from different manufacturers and new app versions.	
Number of Participants	48
Devices	• 28 Samsung SM-A405FN • 20 Android phone from different manufacturers
Number of Peer-to-Peer Connections	302
Number of Samples	1164
Total Recording Length in Hours	266.11

Room	Number of Samples (Timeseries)	Number of Users
1	216	8
2	384	10
3	263	9
4	120	7
5	181	7

1 Experimental Setup

1.1 Test Preparation

Tests were carried out with 48 test subjects at five different locations within the Julius Leber barracks in Berlin. There are three rooms within a conference center and two outdoor locations. In the three indoor locations and one outdoor location, ten subjects took part in the test. Meanwhile, eight subjects took part in the second outdoor location. All test subjects were equipped with breathing masks so that there was no risk of infection.
All test subjects were labelled with numbers and color codes (on their chest and back) and with color codes on their legs and arms. The numbers were 101... 110, 201... 210, 301... 310, 401... 410, and 501... 508, with the first digit indicating the test location.
The floor of the test areas was marked (Fig. 2). These markings consisted of a 5 m x 5 m grid with lines spaced 50 cm apart. Markings and arrows are also shown on this grid. From the starting point (box within a box) to the ending point (multiplication sign), the test subjects had to walk through markings and stay on each marker for a predetermined amount of time (2, 4, 6, or 10 min).
The markings are numbered on the green path from 1 to 9 and on the black path from 2 to 10 (Fig. 1, right).
One camera was installed at each location to take photographs from the test location. The photographs were taken at the time instances were all persons reached their positions and just before they left their locations. This turned out to be sufficient to get reliable time intervals where the test persons were at fixed positions and made the evaluation of the time intervals much simpler than with video.

1.2 Test Execution

The test was carried out in four runs. The duration of the first three runs was $\leq$ 30 min. The fourth run lasted 60 min. In response to a signal from the test coordinator, the test subjects moved from their assigned square?following the direction indicated by the arrow?to the next square. Here they remained for a predetermined amount of time (2, 4, 6, or 10 min). The test subjects with an odd number started on the green path; the test subjects with an even number started on the black path.

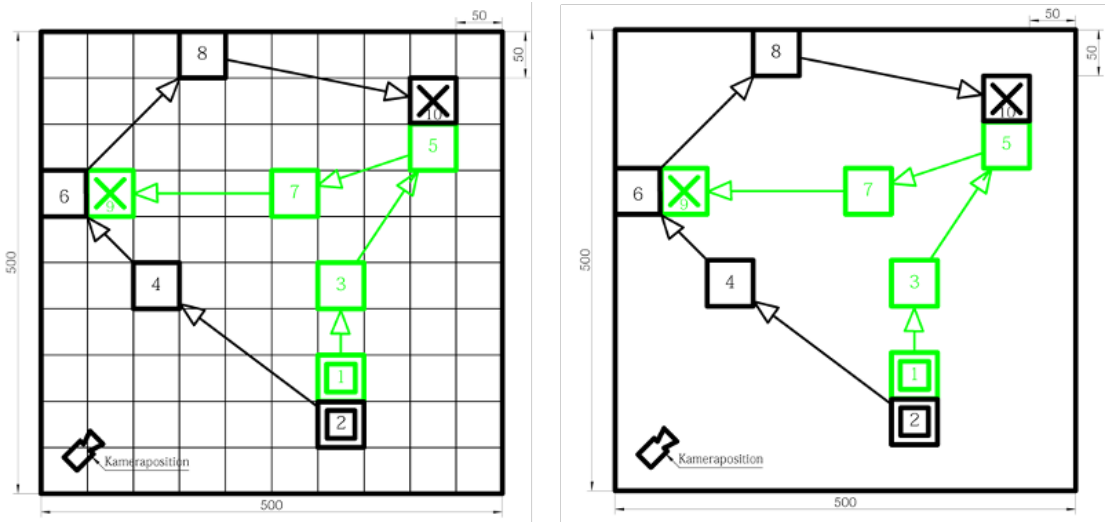

Figure 2: Test pattern on the floor of the five test areas (left with grid, right without grid).

During the runs, the test subjects were instructed not to move to much but to hold the positions of the mobile phones relatively stable and to stand within the square. They held the cell phone in their hand the whole time and were not permitted to put it in their pocket.
For each test area, a run was initiated by the two subjects with the smallest numbers (e.g., 101 and 102). Then, the two subjects with the next highest numbers (e.g., 103 and 104) entered the test area, etc., until all ten (eight) subjects were in the test area. The two test subjects who initiated the run were also the first to leave the room, so that the test area gradually emptied ($10 \rightarrow 8 \rightarrow 6 \rightarrow 4 \rightarrow 2$ people). Before entering and after leaving the test area, the test subject was kept at least 6 m from the test area. Figure 3 illustrates the test run. Table 2 shows the control sheet for the test procedure in the first test area.

1.3 Evaluation

The fixed positions of the markings resulted in fixed distances between these markings. These are shown in Figure 3. Using the video recordings, it was possible to determine when test subjects reached and left a fixed position. Consequently, it was also possible to determine the time intervals during which a test cell phone was in a fixed position. Based on the values given in Table 2, the actual distances (ground truth) of all the mobile phones involved in the test can be determined. These values were used to validate the extent to which the distances determined by the apps match the actual distances (see data analysis).

2 Infection Risk Estimation

Based on this data, our goal is to predict whether a "dangerous" proximity between two soldiers has occurred over the time period $t = 1, .., T$.

To do so we first need to define which proximities should be considered as dangerous. This raises the following fundamental question:

Given a time series of physical distances, between a diseased and a healthy person $d_t, t = 1, .., T$, what is the probability that the diseased person transmitted the virus to the healthy person over the time period? ($\rightarrow$ see Fig. 4 for an illustration)

We address this question, by computing an "infectiousness" value depending on the time and proximity of

Position 1 - Indoor												
10 Participants												
Participants with uneven/even ID start on the green path, black path.												
2 minutes												
No	Time	1	2	3	4	5	6	7	8	9	10	clock time
1	2	101	102									
2	4	103	104	101	102							
3	6	105	106	103	104	101	102					
4	8	107	108	105	106	103	104	101	102			
5	10	109	110	107	108	105	106	103	104	101	102	
6	12			109	110	107	108	105	106	103	104	
7	14					109	110	107	108	105	106	
8	16							109	110	107	108	
9	18									109	110	
4 minutes												
No	Time	1	2	3	4	5	6	7	8	9	10	clock time
1	4	101	102									
2	8	103	104	101	102							
3	12	105	106	103	104	101	102					
4	16	107	108	105	106	103	104	101	102			
5	20	109	110	107	108	105	106	103	104	101	102	
6	24			109	110	107	108	105	106	103	104	
7	28					109	110	107	108	105	106	
6 minutes												
No	Time	1	2	3	4	5	6	7	8	9	10	clock time
1	6	101	102									
2	12	103	104	101	102							
3	18	105	106	103	104	101	102					
4	24	107	108	105	106	103	104	101	102			
5	30	109	110	107	108	105	106	103	104	101	102	
10 minutes												
No	Time	1	2	3	4	5	6	7	8	9	10	clock time
1	10	101	102									
2	20	103	104	101	102							
3	30	105	106	103	104	101	102					
Break												
4	10	107	108	105	106	103	104	101	102			
5	20	109	110	107	108	105	106	103	104	101	102	
6	30			109	110	107	108	105	106	103	104	

Table 2: The control sheet for the test procedure in the first test area

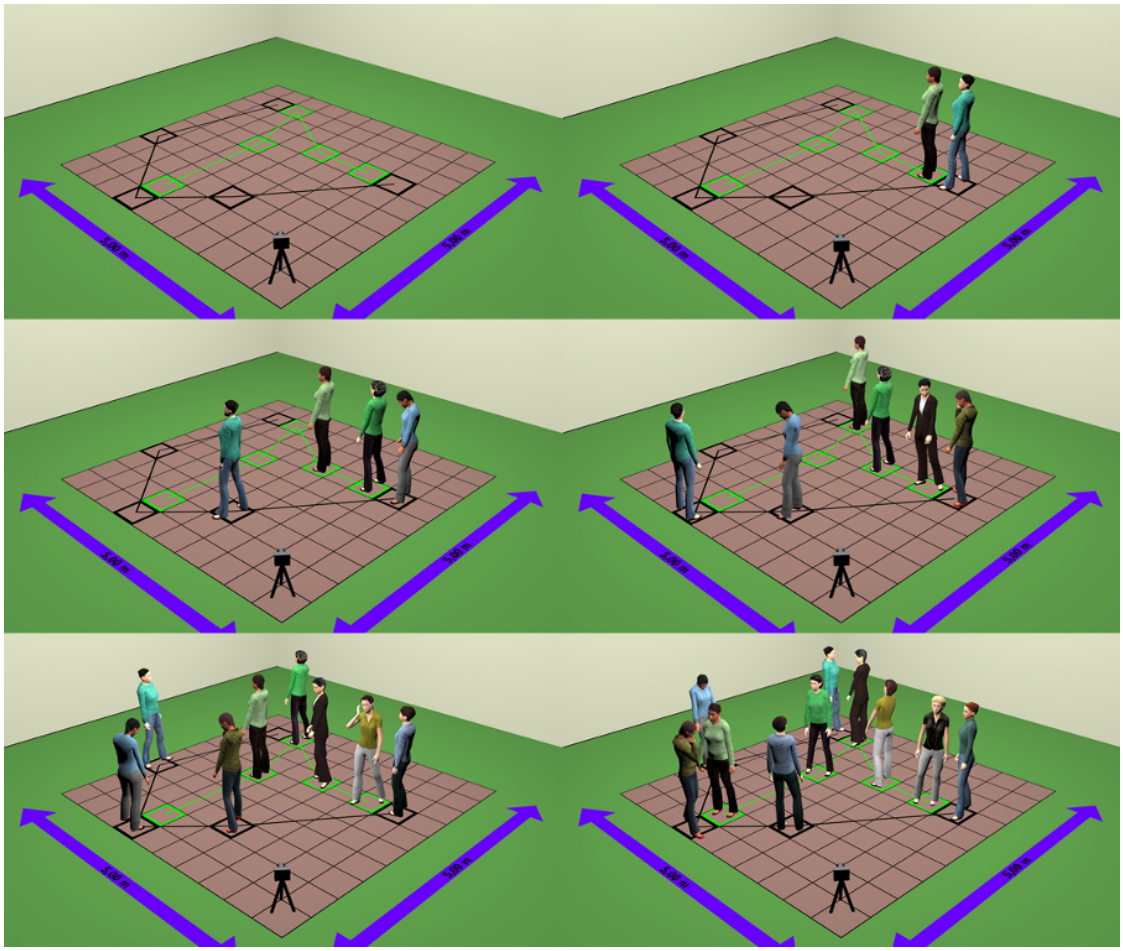

Figure 3: Illustration of the test runs.

contact according to the equation

$$I_f(d_1, .., d_T) = \sum_{t=1}^{T} f(d_t). \tag{1}$$

There are many different choices for the infectiousness function f. These simple models based on, for instance, linear functions, box functions or sigmoid functions. As the underlying problem is an epidemiological problem we have decided to use an epidemiological model to do the risk prediction. In fact, we will use the model proposed by Fraser et al. [2] which uses a quadratic diffusion model combined with a risk score model that controls the number of notified individuals. However, our methods are not limited to a certain model and can be applied on other functions as well.

2.1 Risk Estimation

The function proposed by Fraser et al. is monotonically decreasing with distance to account for the fact that infectiousness decreases as the distance between people increases. Furthermore, for a model parameter $d_0 > 0$ that models the short distance we use the hybrid diffusion function

[2] *Defining an epidemiologically meaningful contact from phone proximity events: uses for digital contact tracing*, Christophe Fraser, David Bonsall, Robert Hinch and Anthony Finkelstein, Draft, April 2020.

	1	2	3	4	5	6	7	8	9	10
1	-	50.00	100.00	223.61	269.26	360.56	206.16	380.79	320.16	316.23
2	50.00	-	150.00	250.00	316.23	390.51	254.95	427.20	353.55	364.01
3	100.00	150.00	-	200.00	180.28	316.23	111.80	291.55	269.26	223.61
4	223.61	250.00	200.00	-	335.41	141.42	180.28	254.95	111.80	360.56
5	269.26	316.23	180.28	335.41	-	403.11	158.11	269.26	353.55	50.00
6	360.56	390.51	316.23	141.42	403.11	-	250.00	212.13	50.00	412.31
7	206.16	254.95	111.80	180.28	158.11	250.00	-	180.28	200.00	180.28
8	380.79	427.20	291.55	254.95	269.26	212.13	180.28	-	180.28	254.95
9	320.16	353.55	269.26	111.80	353.55	50.00	200.00	180.28	-	364.01
10	316.23	364.01	223.61	360.56	50.00	412.31	180.28	254.95	364.01	-

Table 3: Distances between all markings in a test area. The numbering corresponds to those in the markings in Fig. 2 (right).

Figure 4: Exemplary distance profile between a diseased and a healthy person over time for the scenario "short encounter in the U-Bahn". What is the probability of infection for this event?

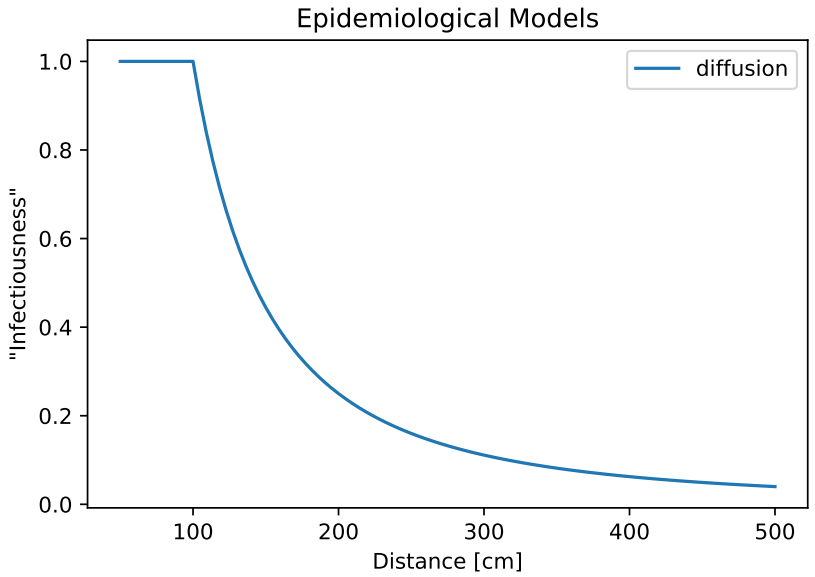

Figure 5: The Diffusion Infectiousness Function, which aims to approximate the relative risk of getting infected as proposed by Fraser et al.

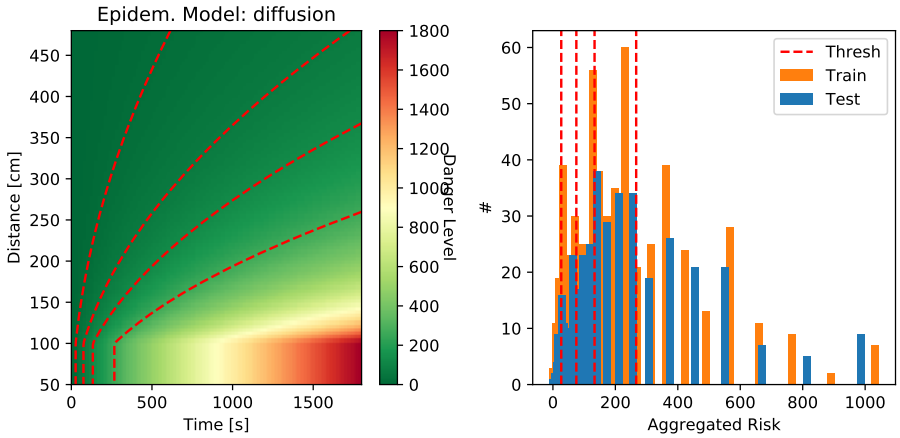

Figure 6: Left plot: Regions of the time-distance plane, which are considered as dangerous of not according to our infectiousness model and different thresholds. Right Plot: Histogram of the aggregated proximities of all points in the data set and thresholds obtains by four different reference sequences.

$$f_{diffusion}(d[cm]) = \begin{cases} 1 & \text{if } d[cm] < d_0 \\ d_0^2/d[cm]^2 & \text{if } d[cm] \geq d_0 \end{cases}$$

as proposed by Fraser et al.
For every infectiousness function we then compute an infectiousness threshold by inserting the reference sequence d^{ref}, with

$$d_t^{ref} = 200cm, t = 1, .., 900s \tag{2}$$

into equation (1). This results in an infectiousness threshold

$$I_f^{th} = I_f(d^{ref}) = \sum_{t=1}^{T} f(d_t^{ref}) \tag{3}$$

We choose the sequence of distances (2) as a reference sequence to determine the threshold, because, according to the Robert Koch Institute, for the flu, a physical proximity between two people of less than 2 meters over a time period of 900 seconds (15minutes), results in a high risk of being infected.
By selecting the infectiousness function and the infectiousness threshold we can determine, which time series of distance measurements should be considered dangerous and which shouldn't:

$$\text{Dangerous}(d_1, .., d_T) = \begin{cases} True & \text{if } I_f(d_1, .., d_T) > I_f^{th} \\ False & \text{if } I_f(d_1, .., d_T) \leq I_f^{th} \end{cases} \tag{4}$$

2.2 Macroscopic Risk Estimation

As there is currently no consensus on which reference sequence d_t^{ref} is the correct one for the current Sars-Cov-19 pandemic, we can also take a different labeling approach which is outlined in Fraser. Instead of setting a fixed threshold I_f^{th}, they propose to adapt the value of I_f^{th} according to the global transmission rate of the virus. Say we assume a reproduction number of k, e.g. that one infected person passes the virus on to k other people on average. Then the total number of infected people will be

$$N_{infected-total} = kN_{tested-positive} \tag{5}$$

i.e. k times the number of people that are tested positive. Our approach allows us to set the risk threshold I^{th} in such a way that we will warn exactly $N_{infected-total}$ people. This makes it possible to steer the warning sensitivity of our App according to the macroscopic parameters, such as the reproduction number.

2.3 Features and Scoring Model

Before computing risk scores we compute different set of features from resampled data such that each sample has same dimensionality independent of recording length. In particular we tested **len_max_mean** i.e. length, maximum and mean of resampled received RSSI values resulting in three-dimensional features. In our experiments we tested for multiple other features but observed no major effects on the performance. Once features have been extracted from the raw RSSI data via the data preprocessing pipeline described above, we can input these features into a logistic regression model in order to obtain a "risk" score.
This logistic regression model is parametrized by a set of parameters, which may be different for every combination of receiving and transmitting device, due to differences in the bluetooth antennas and low level preprocessing done by the operating system (see introduction). The input to the logistic regression thus comprises a matrix of parameters and a vector of features. The risk score is then computed by first performing the vector-matrix multiplication and then applying the Softmax function to the 2-dimensional output to get the risk level. The value of the risk level is always between 0 and 1 and can be interpreted as the probability of a possible dangerous contact.

3 Data

3.1 Raw Data

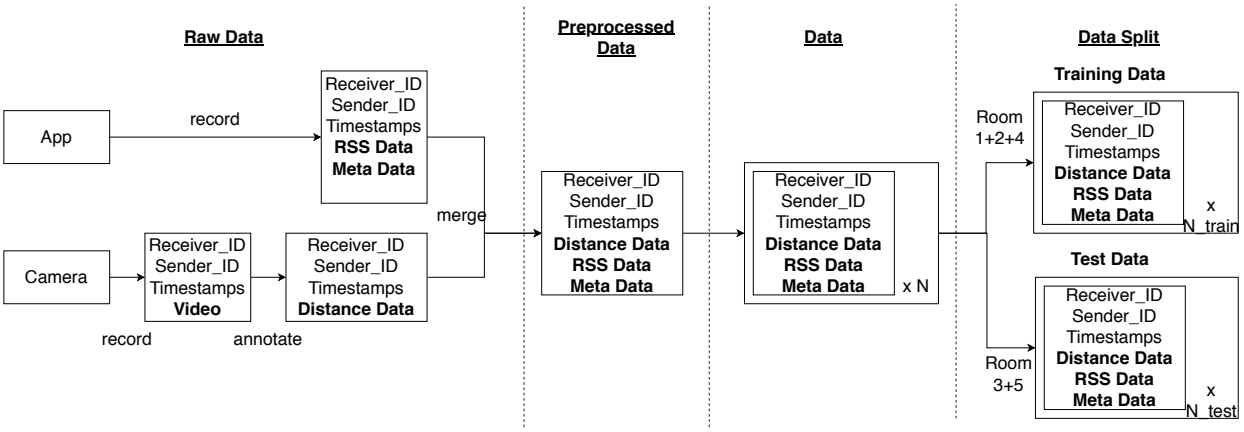

Figure 7: Data processing. Final data will be split in train and test data.

As described in Section1.2 "Test execution", the raw data consists of two types. Signal strength data were collected via the prototype of the Proximity Tracing App [Received Signal Strength (RSSI) data]. The RSSI data were sampled at a random and potentially varying frequency (between 0.1 Hz and 10 Hz). The cameras provided recorded video data. In order to collect Bluetooth signal strength data and ground truth distance measurements, experiments were performed by 50 soldiers at Julius-Leber-Kaserne Berlin from April 01 to April 07, 2020 in Berlin. These experiments resulted in data of the following format, where for every pair of soldiers we collected:

- A time series of distances $d_t, t = 1, .., T$
- A time series of Bluetooth Low Energy (BLE) **r**eceived **s**ignal **s**trength (RSSI) values $RSSI_t, t = 1, .., T$, recorded by mobile phones held by the soldiers

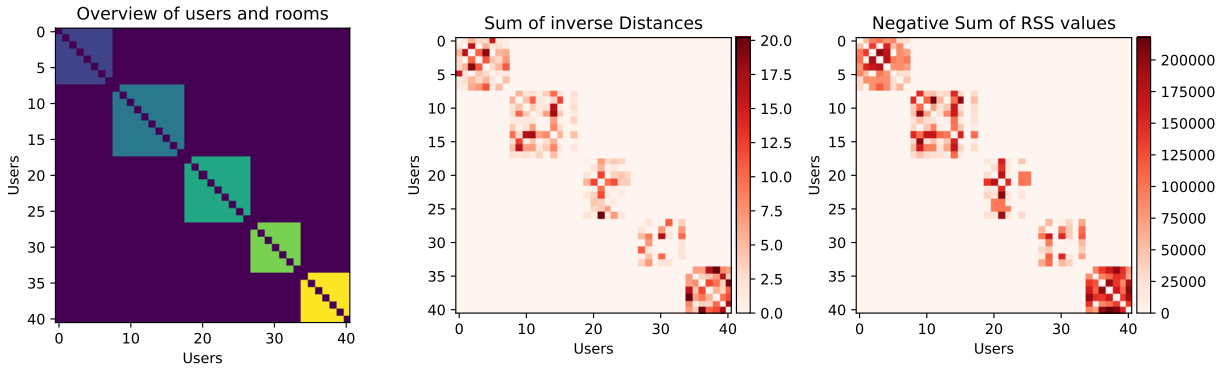

Figure 8: Overview of experimental setup. There is a high correlation between the infection risk and the sum of recorded signal strengths.

The experiments were performed on five different sites (indoor, outdoor) among seven to ten devices, in multiple measurement setups. To evaluate our approach, we consider the entire duration of the measurements. Figure 8 summarizes the experiments and shows the five different sites with their respective number

of devices, as well as the respective accumulated proximities and RSSI data. Both accumulated values correlate with each other and the relation between receiving and sending is symmetric.

3.2 Merging Measurements with Ground Truth Data

During the data collection experiments the events were recorded in videos. The videos were later analyzed by hand to acquire the exact times of position changes of the experiment participants, so that distances can be mapped to the RSSI data. In a merging step, the raw measurements are merged with their respective ground truths (result of the merge in Fig. 3). These preprocessed data contain the following information:

- 'model' – Device (telephone) name
- 'receiverid' – Unique ID of the recipient device
- 'transmitterid' – Unique ID of the transmitting device
- 'timestamp' – Unique time stamp
- 'RSSI' – Measured signal strength (Received Signal Strength)
- 'distance' – Actual distance (ground truth) between the transmitter and receiver devices
- 'scenario' – Scenario that was tested (the test location and how long the test subjects were located at a given point)

3.3 Preprocessing

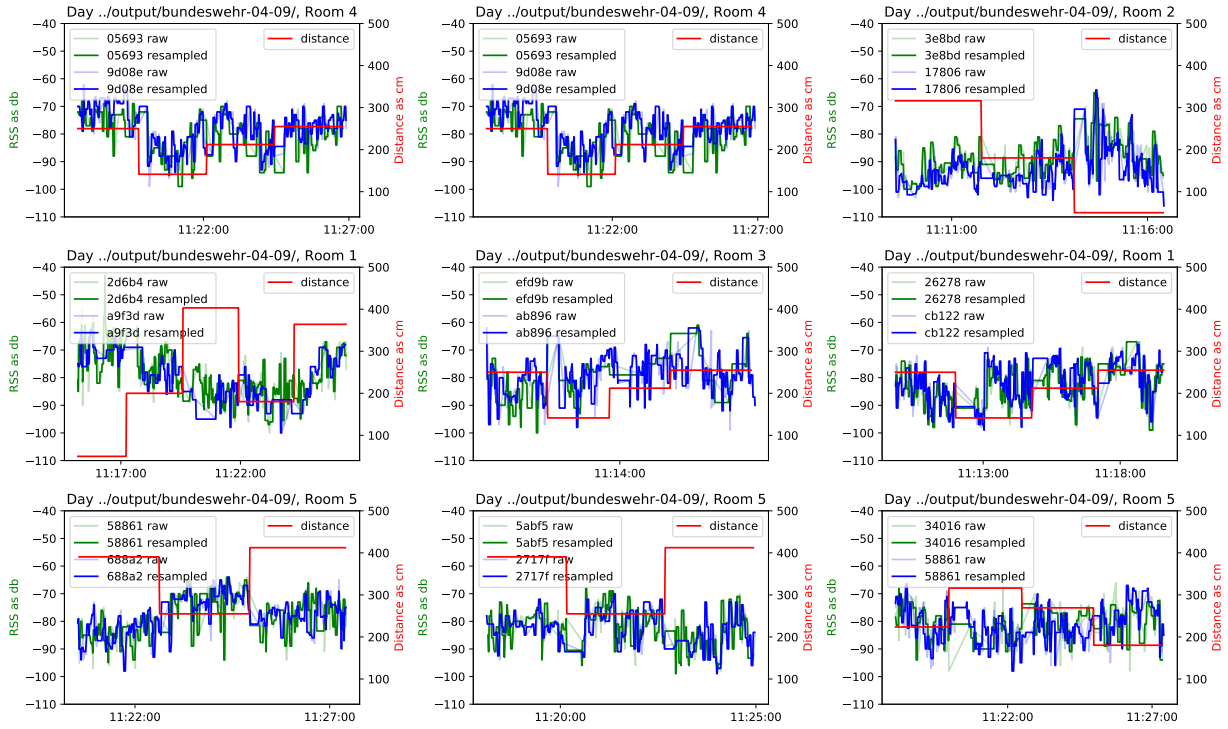

Figure 9: Random samples of data from the data set.

For each device pair (transmitter and receiver), there are several time series of RSSI values corresponding to specific times and points in the test field (see Fig. 2). The time series and associated distances are shown

in color-coded form (Fig. 4). The data are then split into their individual time slots and the RSSI data of each time slot is resampled to 1 Hz. In Figure 9, 9 random samples from the data set are displayed. Note how the RSSI and distance data is symmetric between the two participants. This results in a data set consisting of 1164 data points. The new data points have the following information:

- 'receiver_id', – Unique ID of the recipient device
- 'transmitter_id', – Unique ID of the transmitting device
- 'time_series', – Time series of the time stamps
- 'rss_series', – Time series of the RSSI values
- 'distance_series', – Time series of the distances
- 'scenario', – Scenario that was tested (the test location and how long the test subjects were located at a given point)
- 'number_of_contacts', – Number of devices that were on the test grid at the same time
- 'origin_identifier' – Unique identifier that shows which data points belong to the same measurement

3.4 Training and Testing Data

For training and testing, the time series data are divided into training and validation data. Although there are multiple different approaches that could be used for the division, it was decided to used a mixed approach, where the two indoor data sets (room 1 and 2) and one outdoor data set (room 4) were used for training, and the third indoor data set (room 3) as well as the last outdoor data set (room 5) were used as the testing data. In previous tests multiple combinations of indoor and outdoor rooms were tested to investigate possible indoor and outdoor effects. No significant effects could be detected, so a mixed approach was used. The number of data points collected for these tests is shown in Table 3.

4 Experiments

In the following we label our data according to the scheme described in Section 2.2. Figure 10 shows the achieved results for this approach. Following the recommendation of Fraser et. al., we take conservative estimates of the reproduction factor of 4, 8 and 16 to ensure that all truly infected people will be alarmed.

List of Figures

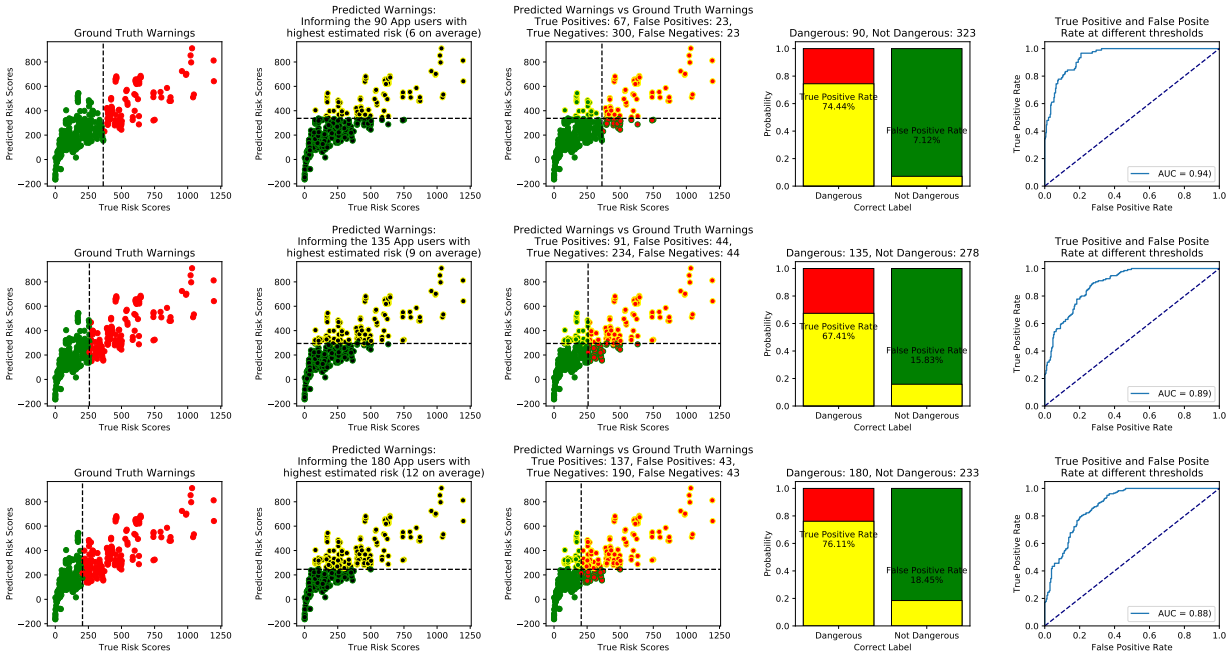

Figure 10: Visualization of the ground truth with respect to the epidemiological model and prediction accuracy. First row: 3 contacts will be notified on average. Second row: 6 contacts will be notified on average. Third row: 9 contacts will be notified on average.