2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Machine Learning Estimation of COVID-19 Social Distance using Smartphone Sensor Data

Oleksandr Semenov, Emmanuel Agu, *Member, IEEE*, Kaveh Pahlavan *Life Fellow, IEEE*

Abstract— **Airborne infectious diseases such as COVID-19 spread when healthy people are in close proximity to infected people. Technology-assisted methods to detect proximity in order to alert people are needed. In this work we systematically investigating Machine Learning (ML) methods to detect proximity by analyzing data gathered from smartphones' built-in Bluetooth, accelerometer and gyroscope sensors. We extracted 20 statistical features from raw sensor data, which were then classified (< 6ft or not) and regressed (distance estimate) using ML algorithms. We found that elliptical filtering of accelerometer and gyroscope sensors signal improved the performance of ML regression. The most predictive features were z-axis mean and fourth momentum for the accelerometer sensors, z-axis mean y-axis mean for the gyroscope sensor, and advertiser time and mean RSSI for Bluetooth radio. After rigorous evaluation of the performance of 19 ML classification and regression methods, we found that ensemble (boosted and bagged tree) methods and regression trees ML algorithms performed best when using data from a combination of Bluetooth radio, accelerometer and the gyroscope. We were able to classify proximity (< 6ft or not) with 100% accuracy using the accelerometer sensor and with 62%-97% accuracy with the Bluetooth radio.**

I. INTRODUCTION

COVID-19 is a highly infectious airborne transmittable disease that currently has over 150 million people infected with a total of 3.2 million deaths globally till date, costing over 28 trillion dollars to manage. The risk of airborne infectious diseases such as COVID-19 increases when healthy people are within 6 feet of infected people for longer than 15 minutes [25]. This has led to research interest in to estimate the distance between smartphone users by analyzing data from built in sensors such as Bluetooth, accelerometer and gyroscope.

Sensing social interactions has been explored in prior work. The main objective of such research is to track human mobility and epidemic spreading [2], human behavior in organization settings and how it shapes individuals and organizations [3], and the propagation of information [4]. Researchers in the past experimented using combination of various sensors to detect social interactions or estimate proximity including the accelerometer [5, 6, 7, 3], gyroscope [5, 6], magnetometer [8], microphone [6, 3], GPS [7], Wi-Fi [9, 8, 10] and Bluetooth [11, 9, 5, 12, 13, 3] radio. Some distance estimation approaches utilized mobile Received Signal Strength Information (RSSI) [11, 9, 5, 12] and Time Difference of Arrival (TDoA) [14]. Data analyses has utilized both classical proximity estimation techniques such as Path Loss Model (PLM), as well as ML algorithms such as: AdaBoost [11], XGBoost, Linear Regression, SVM, and Random Forest [5]. Some prior work focused on recognizing the user context [6]. Much of prior work was pre-COVID and focused on broader goals such as detecting social interactions. We focus on COVID-19 proximity estimation and classification.

In this paper, we investigate how accurately the proximity of two smartphones can be estimated using data from their built-in accelerometer, gyroscope sensors and Bluetooth radio. We analyzed these data using ML algorithms to estimate range. A novel aspect of our work is that we are the first to explore whether context-specific ML models are more accurate than general ones. Specifically, we explore whether first recognizing the user's context such as whether the user is indoors or outdoors, room size, user's pose and location of the transmitting device on the body and providing this context information as an input feature improves ML proximity estimation. We found that adding user context information as a feature improved the accuracy of ML regression (distance estimate) but not ML proximity classification (< 6ft or not).

Range estimation using Bluetooth radio is challenging because the RSSI varies continually due to multipath fading, the transmission

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

environment, room size, the presence of obstructions and the number of people in the room.

II. Approach

TABLE 1. SUMMARY OF THE MITRE RANGE AND ANGLE (UNSTRUCTURED) DATASET

Gyroscope			**Context**	
x-axis	y-axis	z-axis	Indoor	Sitting
Accelerometer			Outdoor	Standing
x-axis	y-axis	z-axis	Large Room	Hold to Right Ear
Bluetooth			Medium Room	Front Pants Pocket
RSSI	TSSI	Advertiser Timestamp	Small Room	In Hand
Response			Center Congested	In Purse
Range	Angle		Center Open	Rear Pants Pocket
			Near Wall Congested	Shirt Pocket
			Near Wall Open	

TABLE 2. STATISTICAL FEATURES COMPUTED FROM ACCELEROMETER, GYROSCOPE, AND BLUETOOTH DATA

Sensor	*Feature*	*Formula*	*Ref*
A, G	***Magnitude***	$magnitude = \sqrt{x^2 + y^2 + z^2}$	-
A, G, B	***Mean***	$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i n$	-
A, G, B	***Standard Deviation***	$S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\lvert X_i - \mu\rvert^2}$	-
A, G, B	***Third and Fourth Momentum***	$m_k = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^k$, k=3, 4	-
A, G, B	***Percentile***	The score at k percentile for k = 25, 50, 75	-
A, G, B	***Value Entropy***	$H_x(y) = -K\sum_{i=1}^{N} p_i \log p_i$	15
A, G	***Time Entropy***	$H_x(y) = -K\sum_{i=1}^{N}\lvert m_i\rvert \log(\lvert m_i\rvert)$	15
A, G, B	***Autocorrelation***	$r_k = \frac{c_k}{c_0}$	18
A, G	***Autocovariance***	$K_X(t,s) = cov[X(t), X(s)] = E\{[X(t) - \mu_X(t)][X(s) - \mu_X(s)]\}$	19
B	***Delta***	$delta = RSS - TSS$	-

Smartphone proximity dataset: In our experiments, we utilized the publicly available PACT Mitre Range and Angle (Unstructured) (MRAU) dataset [1] to develop our ML proximity classification and regression models [1]. Table 1 summarizes the data in the MRAU dataset. The RSSI signal measurements were taken at increments of 2 feet from 2 to 16 feet between transmitting and receiving devices. However, not all subjects recorded RSSI signal strength at all distances. The data was collected for 60 seconds at each distance at ~4Hz frequency for accelerometer and gyroscope sensors and ~10Hz for Bluetooth. The main steps in our machine learning pipeline are shown in figure 1. First, the raw sensor signals were filtered using a low-pass filter. Next various statistical features that prior work has found to be predictive were extracted. These features including the user's context were then classified or regressed to predict user proximity/range using ML methods.

Signal filtering: We evaluated the utility of 5 filter types for ML range estimation: Butterworth [16], Chebyshev [16], Elliptical [16], Median [17], moving average and moving average with overlapping windows. In order to determine frequencies of interest, the FFT of the sensor signal were computed. Most of the signal energy were found to be concentrated in the 0-0.2Hz, 0.3Hz-0.5Hz, and 1.3Hz-1.5Hz bands. The Kaiser window FIR filter was used to construct a band-pass filter to test combinations of each of the three bands.

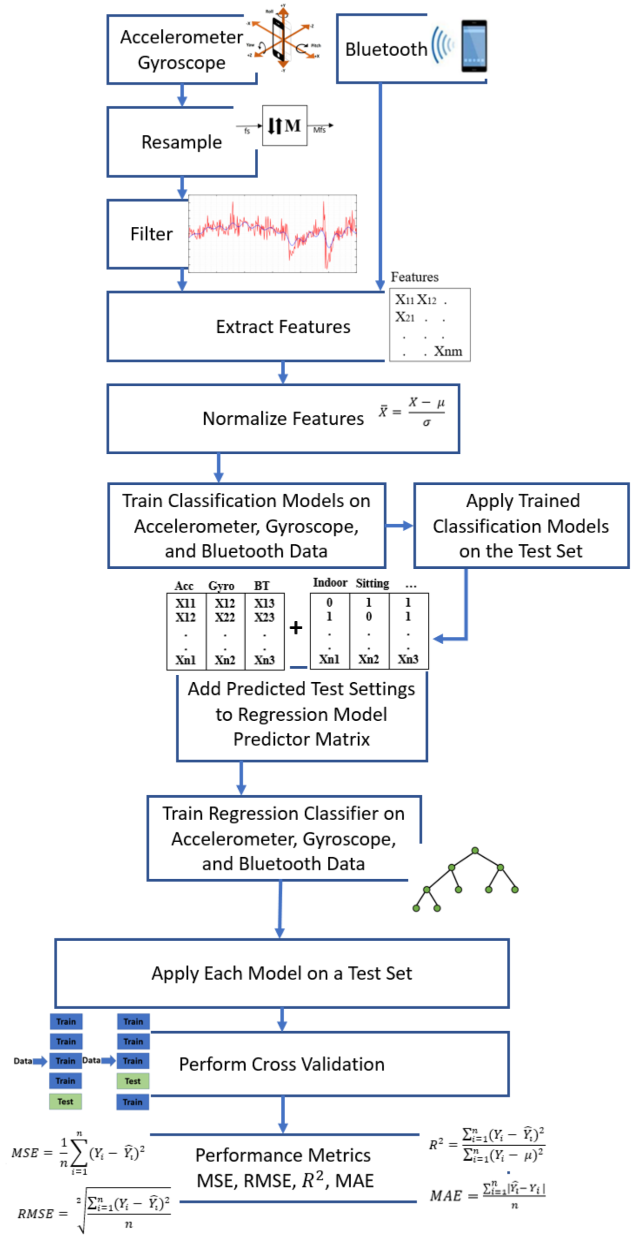

Figure 1. Overview of our ML proximity regression pipeline using Accelerometer, Gyroscope, and Bluetooth data.

Feature extraction: Table 2 summarizes statistical features we computed from accelerometer, gyroscope, and Bluetooth data. Additionally, for accelerometer and gyroscope sensors, mean and standard deviation for each axis were computed as well as autocorrelation between xy, xz, and yz axes. A total of 20 features were extracted for each sensor. The advertiser time feature was also extracted for Bluetooth, yielding a total of 14 features for Bluetooth radio.

TABLE 3: EVALUATION METRICS

Mean Squared Error (MSE)	$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$
Root-Mean-Square Error (RMSE)	$\mathrm{RMSE} = \sqrt[2]{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}}$
R^2	$R^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \mu)^2}$
Mean Absolute Error (MAE)	$\mathrm{MAE} = \frac{\sum_{i=1}^{n}\lvert\hat{Y}_i - Y_i\rvert}{n}$
Balanced Accuracy (BA)	$\mathrm{BA} = \frac{TPR + TNR}{2}$
F1	$F1 = \frac{2 * TPR * precision}{TPR * precision}$
True Positive Rate (TPR)	$\mathrm{TPR} = \frac{TP}{TP+FN}$
True Negative Rate (TNR)	$\mathrm{TNR} = \frac{TN}{TN+FP}$
Precision	$precision = \frac{TP}{TP + FP}$

TABLE 4. F1 AND BA CLASSIFIER PERFORMANCE

Classification Model Accuracy	**Accelerometer**		**Gyroscope**		**Bluetooth**	
	F1	***BA***	***F1***	***BA***	***F1***	***BA***
Indoor	0.8512	0.9014	0.7595	0.8371	**0.9981**	**0.9991**
Outdoor	0.7812	0.8626	0.4939	0.7098	**0.9999**	**1**
Large Room	0.7946	0.8708	0.7152	0.8465	**0.9971**	**0.9988**
Medium Room	0.9035	0.9041	0.7793	0.7843	**0.9992**	**0.9992**
Small Room	0.8466	0.9005	0.5839	0.74	**0.9999**	**1**
Center Congested	0.5835	0.7511	0.1614	0.59	**0.9936**	**0.9999**
Center Open	0.8511	0.8921	0.6514	0.7466	**0.9995**	**0.9995**
Near Wall Congested	0.8657	0.8993	0.7512	0.8113	**0.9999**	**1**
Near Wall Open	0.9035	0.9309	0.7357	0.8026	**0.9999**	**0.9999**
Sitting	0.8815	0.8987	0.7401	0.782	**0.9979**	**0.998**
Standing	0.9151	0.8987	0.8195	0.782	**0.9979**	**0.998**
Hold to Right Ear	0.8546	0.9033	0.7712	0.8362	**0.9992**	**0.9996**
Front Pants Pocket	0.635	0.7677	0.3487	0.6125	**0.9943**	**0.9975**
In Hand	0.7518	0.8048	0.5636	0.6611	**0.9929**	**0.9929**
In Purse	0.6756	0.7956	0.2771	0.5835	**0.9796**	**0.992**
Rear Pants Pocket	0.6292	0.7419	0.456	0.6573	**0.985**	**0.9895**
Shirt Pocket	0.5803	0.7863	0.1325	0.5656	**0.989**	**0.9987**
Average:	0.7826	0.8535	0.573	0.7264	**0.9955**	**0.9978**

In addition to statistical features, we employed classification models to estimate the environment in which measurements were taken. There were five types of environment labels: 1) Indoors or outdoors, 2) Room size: large, medium or small room. 3) Transmitting device's location in the room: center congested, center open, near wall congested, and near wall 4) Pose of the test subject: sitting or standing, and 5) Phone placement: held to right ear, front pants pocket, in hand, in purse, rear pants pocket, or shirt pocket.

All the features were computed over windows of 10 samples of continuously sampled sensors' signals. Before training regression classifier on the dataset, the data was normalized using one of two methods: 1) z-score $\bar{X} = \frac{X-\mu}{\sigma}$ or 2) min-max normalization $\bar{X} = \frac{X-X_{min}}{X_{max}-X_{min}}$.

Machine learning algorithms: We evaluated linear regression [20, 22], regression trees [21], Support Vector Machines (SVM) [23], ensemble methods [20], and Gaussian Process Regression (GPR) [24]. We validated the results of regression models using 5-fold cross validation technique to avoid overfitting and to robustly determine the optimal ML model. The best performing model was also evaluated using leave-one-out cross-validation with subject level splitting.

Evaluation metrics: Table 3 summarizes evaluation metrics used to measure performance of regression and classification models.

III. EVALUATION AND RESULTS

A. *Signal filtering*

We found that the Elliptical filter of the 9th order performed best. Various cutoff frequencies ranging from 0.1 Hz to 1.8 Hz were tested. The highest regression R^2 for the accelerometer was 0.63 at cut-off frequency of 0.2 Hz and 0.27 for the gyroscope at a cut-off frequency of 0.1 Hz. This was a significant improvement over the best R^2 achieved using unfiltered data: 0.31 for the accelerometer and 0.21 for the gyroscope.

B. *Classification performance on various labels*

Table 4 summarizes the results of classifying various labels in the MRAU dataset. Overall, Bluetooth radio had very high accuracy for recognizing all the context variables with BA and F1 values of 0.99 for all. Accelerometer data yielded good accuracy for recognizing when the subject was sitting or standing and also in detecting the size of the room. Figure 2 shows the importance of each feature for various range estimation tasks. Feature importance calculated as difference in the node risk between parent and children's nodes

$\frac{R_1 - R_2 - R_3}{N_{branch}}$ where risk is defined as a node error. We discovered that including additional features improved the ML model's performance of the model when validated using 5-fold cross-validation.

TABLE 5. RESULTS OF AN OPTIMAL CLASSIFIER SEARCH

Model	Method	RMSE score		
		A	*G*	*B*
Linear	Linear	4.4223	4.6607	4.7696
	Interactions Linear	4.6399	7.938	5.82E8
	Robust Linear	4.4308	4.6633	12.719
	Stepwise Linear	3.9749	4.6461	7.0528
Tree	Fine Tree	3.5293	5.1107	**0.2289**
	Medium Tree	3.3486	4.7299	0.3087
	Coarse Tree	3.4537	4.4817	0.4765
SVM	Linear SVM	4.528	4.7148	4.6823
	Quadratic SVM	4.2603	7.1004	19.766
	Cubic SVM	30.453	105.29	26940
	Fine Gaussian SVM	3.234	4.2502	3.7406
	Medium Gaussian SVM	3.3354	4.2912	4.2992
	Coarse Gaussian SVM	4.2538	4.5428	4.5327
Ensemble	Boosted Trees	3.5044	4.2923	2.1337
	Bagged Trees	**2.837**	**4.0989**	0.5902
GPR	Squared Exponential	3.0078	4.1947	3.6705
	Matern 5/2	2.9709	4.1685	3.5152
	Exponential	2.9223	4.1153	2.0701
	Rational Quadratic	2.9249	4.1399	2.1051

TABLE 6. REGRESSION MODEL RESULTS WITH 5-FOLD CROSS-VALIDATION

	Accelerometer		Gyroscope		Bluetooth	
	Handcrafted Features	*All Features*	*Handcrafted Features*	*All Features*	*Handcrafted Features*	*All Features*
MSE	7.1917	7.4384	15.9857	16.2524	0.0409	0.0485
RMSE	2.6817	2.7273	3.9982	4.0314	0.2023	0.2202
R2	0.6874	0.6767	0.3051	0.2935	0.9983	0.998
MAE	1.8938	1.9426	3.2899	3.3445	0.0112	0.0122

TABLE 7. REGRESSION MODEL RESULTS WITH LEAVE-ONE-OUT CROSS-VALIDATION

	Accelerometer		Gyroscope		Bluetooth	
	Handcrafted Features	*All Features*	*Handcrafted Features*	*All Features*	*Handcrafted Features*	*All Features*
MSE	43.8925	43.2142	39.2562	39.0704	33.2625	12.9921
RMSE	6.6251	6.5738	6.2655	6.2506	5.7674	3.6045
R2	-0.2007	-0.1822	-0.0739	-0.0688	0.0901	0.6446
MAE	5.9464	5.9096	5.7377	5.7527	4.5635	2.2201

TABLE 8. REGRESSION MODEL RESULTS WITH SUBJECT LEVEL SPLITTING CROSS-VALIDATION

	Accelerometer		Gyroscope		Bluetooth	
	Handcrafted Features	*All Features*	*Handcrafted Features*	*All Features*	*Handcrafted Features*	*All Features*
MSE	11.7313	11.1554	24.4948	23.3507	35.929	41.4525
RMSE	3.4251	3.34	4.9492	4.8323	5.9941	6.4384
R2	0.5098	0.5339	-0.0235	0.0243	-0.5012	-0.732
MAE	2.8614	2.7524	4.3665	4.2637	4.5213	4.8695

C. *Best performing ML classification algorithm*

Table 5 summarizes compares the performance of various machine learning classification algorithms. Regression Trees performed best with Bluetooth sensor data, and Bagged Trees performed best on the accelerometer and gyroscope data. The highest ML regression fit was observed with Bluetooth sensor.

D. *Cross-validation*

In addition to 5-fold cross validation technique we validated performance of our model using leave-one-out cross validation, and subject level splitting. For leave-one-out cross-validation we split data into training and test sets by subject, where one subject's data was put in the test set and all others in the training set. This validation approach most closely mimics real world environments. For subject level cross-validation we used a 70/30 train/test split (12 subjects in training set, 5 subjects in test set). Tables 6 through 8 show the results of cross-validation.

IV. DISCUSSION

The key findings of our study include that:

1) Elliptical filtering of the accelerometer and gyroscope signals improves the regression R^2 by 0.32 and 0.06 using accelerometer and gyroscope data respectively.

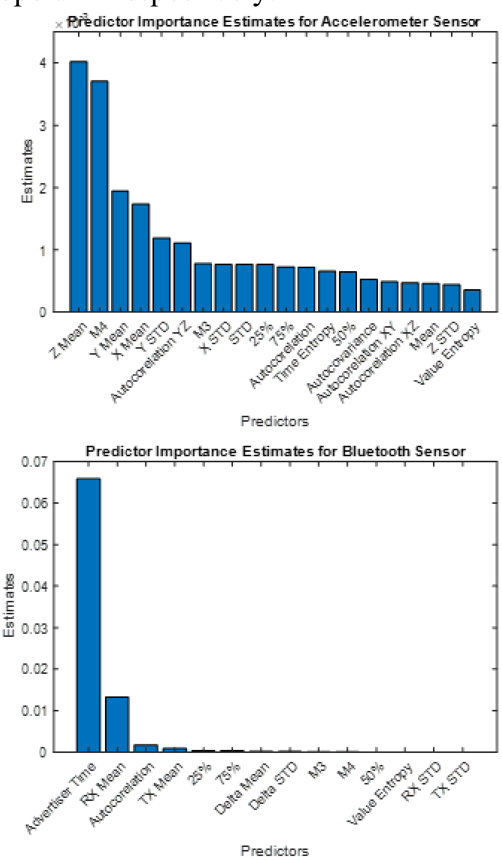

Figure 2. Predictor Importance

2) Ensemble ML classification methods (boosted and bagged trees) classified < 6ft or not between with 100% accuracy using accelerometer sensor data.

3) Regression tree methods to estimate the actual distance between users when utilizing Bluetooth data, achieving an R^2 between 0.64 to 0.99.
4) The most important features were z-axis mean and fourth momentum for the accelerometer, z-axis mean and y-axis mean for the gyroscope sensor, and advertiser time and mean RSSI for Bluetooth.
5) Recognizing user context improved the performance of range regression but not classification.

Even though performance of the regressor trained on the Bluetooth radio data was as high as 0.99 F1 score when validated with 5-fold cross-validation, performance of the regressor dropped significantly when validated using subject level splitting. The R^2 of the best performing ML regression model using Bluetooth reduced by 0.35 R^2, and by 0.49 and 0.22 when using accelerometer and gyroscope data respectively. We believe that accelerometer and gyroscope are valuable sensors when estimating distance since these sensors have information about human motion and those can help identify pose and where the phone is on human body. We observed that adding context information as an input feature to the machine learning model improved regression model performance but not on classification models validated either using 5-fold cross-validation or with leave-one-out cross-validation. But context features helped during validation.

TABLE 9. ACCURACY OF ESTIMATION IF SUBJECTS ARE CLOSER THAN 6 FEET

30/70 Subject Level Splitting			
	Accelerometer	*Gyroscope*	*Bluetooth*
F1	0.9999	0.8452	0.5782
BA	1	0.5045	0.6238
Leave-one-out			
	Accelerometer	*Gyroscope*	*Bluetooth*
F1	0.9999	0.6666	0.9726
BA	1	0.5	0.9722

We believe that the reduction in regression model performance was due to inadequate training data. Figure 3 shows true the predicted vs actual when validated using leave-one-out cross validation. Accelerometer and gyroscope data have a high variance, which can be improved by training the regression model on more data. In contrast, Bluetooth regressor had both high variance and bias. Thus, adding more diverse Bluetooth data could improve regression performance. Overall, the approach presented in this paper is capable of detecting whether two subjects are within 6 feet of each other with 100% accuracy when using accelerometer sensor data. Table 9 shows the classification results for all sensors with leave-one-out cross-validation techniques.

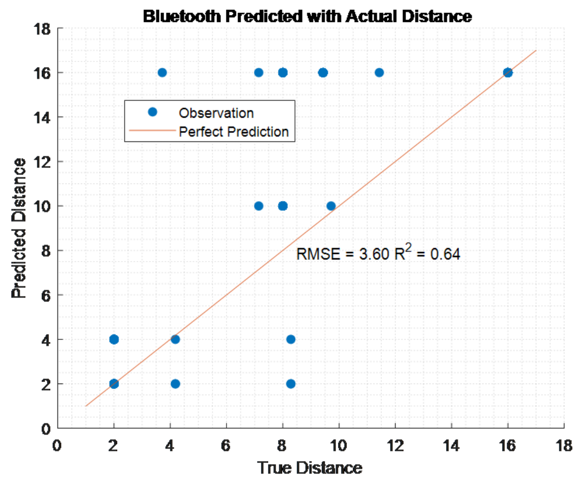

Figure 3. Predicted versus actual distance with leave-one-out cross-validation

V. CONCLUSION

In this paper we have presented research that demonstrates that accurate range estimation with accelerometer, gyroscope sensors and Bluetooth radio are possible with high accuracy in some cases considered. We found that ensemble ML models worked best with accelerometer and gyroscope sensors data, while regression trees performed best with Bluetooth radio data. We found that Elliptical low-pass filter of 9th order with cut-off frequency of 0.2 Hz for accelerometer and 0.1 Hz for gyroscope performed best. Z-axis mean and fourth momentum were the most important features in ML model developed for accelerometer sensor, z-axis mean y-axis mean worked best for gyroscope sensor, and advertiser time and mean RSSI worked best with Bluetooth radio. In addition to handcrafted features, this paper showed that adding context to predictor matrix could improve regression model performance. Our classification model was able to detect context in which measurements took place with an average BA of 0.85 using accelerometer sensor, 0.73 using gyroscope sensor, and 0.99 with Bluetooth radio. Finally, we presented results of an

ML classification model trained on accelerometer data that achieved 100% accuracy estimating if to subjects are closer than 6 feet.

REFERENCES

[1] "Pact datasets and evaluation," PACT, Boston, MA, 2020. [Online]. Available: https://mitll.github.io/PACT/datasets.html#datasets-submit

[2] Isella, Lorenzo et al. "What's in a Crowd? Analysis of Face-to-Face Behavioral Networks." Journal of theoretical biology 271.1 (2011): 166–180. Web.

[3] Olguin, D.O et al. "Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior." IEEE transactions on systems, man and cybernetics. Part B, Cybernetics 39.1 (2009): 43–55. Web.

[4] Pentland, Alex (Sandy). "Automatic Mapping and Modeling of Human Networks." *Physica A* 378.1 (2007): 59–67. Web.

[5] Katevas, Kleomenis et al. "Finding Dory in the Crowd: Detecting Social Interactions Using Multi-Modal Mobile Sensing." (2018): n. pag. Print.

[6] Vaizman, Yonatan, Katherine Ellis, and Gert Lanckriet. "Recognizing Detailed Human Context in the Wild from Smartphones and Smartwatches." IEEE pervasive computing 16.4 (2017): 62–74. Web.

[7] Ouchi, Kazushige, and Miwako Doi. "Indoor-Outdoor Activity Recognition by a Smartphone." ACM Ubicomp 2012. 600-1.

[8] Matic, A et al. "Multi-Modal Mobile Sensing of Social Interactions.", IEEE PervasiveHealth Workshop 2012. 105-14.

[9] Banerjee, Nilanjan et al. "Virtual Compass: Relative Positioning to Sense Mobile Social Interactions." Pervasive Computing. Berlin, Heidelberg: Springer Berlin Heidelberg. 1–21. Web.

[10] Sapiezynski, Piotr et al. "Inferring Person-to-Person Proximity Using WiFi Signals." ACM IMWUT 1.2 (2017): 1–20. Web.

[11] Palaghias, Niklas et al. "Accurate Detection of Real-World Social Interactions with Smartphones." 2015 IEEE ICC 2015. 579–585.

[12] Katevas, Kleomenis et al. "Detecting Group Formations Using iBeacon Technology." Proc. ACM Ubicomp 2016. 742–752.

[13] Ghose, Avik, Chirabrata Bhaumik, and Tapas Chakravarty. "BlueEye: a System for Proximity Detection Using Bluetooth on Mobile Phones." Adj. Proc.. ACM Ubicomp 2013. 1135–1142.

[14] Huang, William et al. "Opo: a Wearable Sensor for Capturing High-Fidelity Face-to-Face Interactions." In Proc ACM Conference on Embedded Network Sensor Systems.2014. 61–75..

[15] Shannon, C. "A Mathematical Theory of Communication." Mobile computing and communications review 5.1 (2001): 3–55.

[16] Schlichthärle, Dietrich. Digital Filters: Basics and Design. Second edition. Heidelberg: Springer, 2011.

[17] Tukey, John Wilder. Exploratory Data Analysis. Reading, Mass: Addison-Wesley Pub. Co., 1977.

[18] George E.P. Box. Time Series Analysis: Forecasting and Control. 5th ed. New York: Wiley, 2016.

[19] Hsu, Hwei P. Schaum's Outline of Theory and Problems of Probability, Random Variables, and Random Processes. Place of publication not identified: McGraw Hill, 1997.

[20] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2009.

[21] Breiman, Leo. Classification and Regression Trees. Belmont, Calif: Wadsworth International Group, 1984.

[22] Huber, P. J. Robust Statistics. New York: Wiley, 1981.

[23] Vapnik, Vladimir N. The Nature of Statistical Learning Theory. New York, NY: Springer New York, 1995.

[24] Rasmussen, Carl Edward, and Christopher K. I Williams. Gaussian Processes for Machine Learning. Cambridge: MIT Press, 2005. Print.

[25] Greenberg, C. and, D. (2020), NIST Pilot Too Close for Too Long (TC4TL) Challenge Evaluation Plan, NIST TC4TL Challenge, [online], https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=930486, https://www.nist.gov/itl/iad/mig/nist-tc4tl-challenge (Accessed May 3, 2021)