Logan Gall

# Spatiotemporal Analysis of WiFi-based Indoor Localization

**Abstract**
GPS technology has made it easy to track absolute location in outdoor environments, however, it is often unusable for indoor location tracking. This paper explores WiFi-based indoor localization as an alternative method for accurate positioning within indoor spaces. This project collects and analyzes WiFi signal data using a Raspberry Pi device under low-noise conditions in the Geocommons lab room at the University of Minnesota. Spatial analysis shows that Received Signal Strength Indicator (RSSI) values decrease with increasing distance from a WiFi access point, while temporal analysis examines the stability of RSSI over time. Additionally, neural network models are trained on the collected data to predict device distance and (X, Y) coordinate location, achieving high accuracy with an $R^2$ value of approximately 0.97 in the best model. The results demonstrate that WiFi signals can be effective for precise indoor localization, despite limitations in environmental interference and signal variability. This study provides a proof-of-concept for WiFi-based indoor localization and sets up future research to enhance scalability and reliability in more complex environments.

## Introduction

GPS technology has made it easy to track absolute location in outdoor environments, however, it is often unusable for indoor location tracking. GPS on smartphones can have an error of around 3-5 meters, making it impractical for accurate indoor positioning. Other methods are needed to find a location indoors. One alternative is the use of WiFi signals. Similar to how GPS uses distance trilateration for positioning, distance could potentially be estimated using the Received Signal Strength Indicator (RSSI) from WiFi access points. If distance can be inferred reliably from RSSI, we may be able to determine a device’s location within a room. Indoor location has been a major interest from companies such as ESRI, Google, and Apple, who are developing their own indoor localization techniques to improve mapping services.

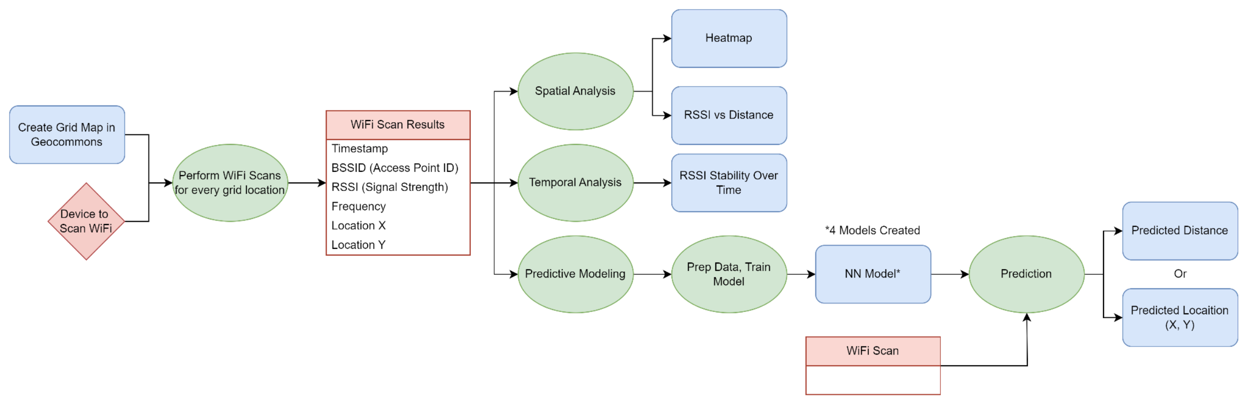
Other works also go into detail evaluating the accuracy and reliability of location estimation. For example, Ranacher et al. (2015) shows that measurement errors in GPS can lead to overestimation of traveled distances, especially at high sampling rates, further showing that GPS may be unreliable in indoor environments. The thorough analysis techniques of this paper are valuable and aided in planning for this project. There are also studies that focus specifically on indoor localization, comparing a wide range of technologies, including WiFi-based approaches. WiFi offers a potential low-cost solution due to existing infrastructure and ubiquity. Obeidat et al. (2021) highlights how WiFi and other technologies can be used alongside methods like triangulation to get reliable indoor localization accuracy. Zafari et al (2019) showcases methods that assess indoor localization systems for efficiency, scalability, and accuracy. All of these research articles give inspiration and real-world importance for exploring WiFi-based localization.

Building on this prior research, this project acts as an additional proof-of-concept for using WiFi signals in a small indoor environment. By performing a methodological test in the Geocommons room at the University of Minnesota, this project aims to analyze WiFi signals, showing that indoor localization can not only be precise but also implemented rapidly using only WiFi data. Furthermore, while most existing research primarily focuses on the spatial aspects of localization, this project introduces a temporal dimension, examining how temporal variability may affect localization model stability. This combined spatiotemporal analysis is significant, as it can contribute to refining indoor navigation systems and guide further improvements in indoor positioning.

**Methodology**

To start, Appendix Table A1 notes important terms that are used regularly through the writing, which may be useful for later reference. The full data for this project, including scanning scripts, analysis code, and raw data can be found at: [https://github.com/logan-gall/IndoorLocationModeling](https://github.com/logan-gall/IndoorLocationModeling) . Note: Generative AI was used when making modifications to code (OpenAI, 2024). Figure 1 below shows an overview data flow diagram for this project, showing the process from data collection, processing, analysis, and predictive modeling.

Figure 1: Overview Project Structure Data Flow



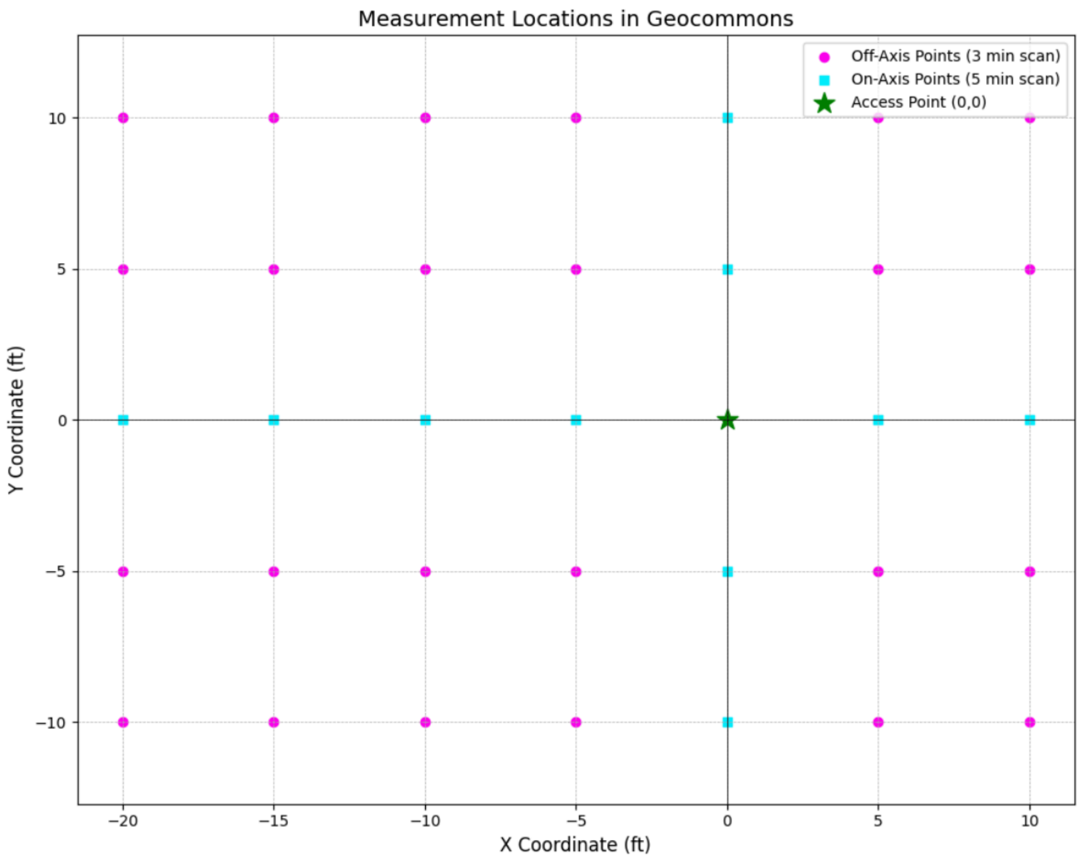

***Setup***

To perform tests of indoor localization, a new dataset has to be created to log WiFi signals. There were multiple devices that were initially tested for this job, such as a Windows 10 Laptop, an Arduino microcontroller, a Raspberry Pi 4b micro-computer, and an Android cell phone. The Raspberry Pi device was decided as the device to use for WiFi logging, because it is able to easily be developed with Python, and creates consistent, accurate, and complete WiFi scans at a ~4 second interval. Other devices fail to remove non-existent access points, are too difficult to develop on, or produce incomplete scan results.

For this study, the “Geocommons” lab room in Blegen Hall at the University of Minnesota, Twin Cities is chosen as the location to perform WiFi scans and analysis. This is an open lab/classroom space with a WiFi access point towards the center of the room. An initial test of the wifi scanning hardware performed scans at a few hand-picked locations in the Geocommons room, but due to a limited data size these results were dropped for a methodological approach to WiFi scan locations.

Figure 2 below shows the locations of WiFi scans in the Geocommons lab room. A local coordinate system is used centered on the location of the access point in the geocommons room. This study is initially small in scope, so the ease of understanding of a local system is more important than adapting and mapping to a universal/global coordinate system. The location of the access point is considered X=0 feet and Y=0 feet. For reference, the doorway to the Geocommons room is located at approximately (-10, -14), the North wall of the room is located towards X=-23, and a large dividing wall starts close to (0, 0) and goes towards (10, 0). A grid pattern through the rest of the room is created in 5 foot increments. A laser tape measure is used to measure distance relative to the starting point and place markers at the given locations throughout the lab space.

Figure 2: WiFi Scan Locations.




***Data Collection***

Once all locations are marked off, the Raspberry Pi device is placed on the ground at a given marker location. A Python script to scan WiFi is used that takes input of current location (X, Y coordinate) and a logging time. The script continuously scans, searching for WiFi access point data for 3-5 minutes, and saves that data to a csv file. Important data columns that are logged can be found in Table 1 below. Each WiFi scan takes approximately 4 seconds to complete, so in a 3 minute log, there are approximately 45 signal scans with anywhere from 25-75 access point signals collected per scan (about 2,500 rows).

The data logging for the initial study was performed in one evening with minimal people and devices in the Geocommons room and Blegan Hall building. This is done to have ideal, repeatable conditions with minimal wireless signal interference. Data is logged for 5 minutes at each location that is ‘on-axis’ of the primary access point, shown as blue dots in Figure 2. Data is logged for 3 minutes at each location that is ‘off-axis’ of the primary access point, shown as pink dots in Figure 2. Points (-20, 10), (-20, 5), (-20, -5), and (-20, 10) were planned but not logged in the initial study due to time restraints.

Table 1: Reported data columns.

| Data Column Name | Description |
|---|---|
| Timestamp | Timestamp of the WiFi data scan |
| Location_X | The X coordinate location in the Geocommons local coordinate system |
| Location_Y | The Y coordinate location in the Geocommons local coordinate system |
| BSSID | The unique identifier for the given WiFi network signal. Each individual access point can have multiple BSSID corresponding to a WiFi network and Frequency. |
| SSID | The WiFi network name associated with the given BSSID, this does not uniquely identify the given wifi signal. |
| Frequency | The frequency at which the wireless signal is transmitted, a number near either 2.4 GHz or 5.8 GHz. The different frequencies have different physical properties, like how far the signal can travel. |
| RSSI (dBm) | Received Signal Strength Indicator, measured in dBm. This shows how strong the quality of the WiFi signal is. This value is negative. Lower values mean weaker signal |

| | |
|---|---|
| | strength. |

The access point inside the Geocommons room transmits six unique WiFi signals. There are three WiFi networks: “eduroam”, “UofM-Guest”, and “UofM-IoT”. Each network has a 2.4GHz and 5.8GHz frequency. The BSSID’s of these six signals are manually noted by performing a brief scan directly next to the access point device. This information is used later to filter raw data. However, since these signals all come from the same access point device, they tend to have equivalent RSSI values. The only variance in signal strength comes from the different frequencies. So, only two BSSID values are necessary to look at in later analysis.

***Analysis & Modeling***

The final logged data is brought into a Python Jupyter Notebook file for spatial analysis, temporal analysis, and predictive modeling. Spatial and temporal analysis seek to understand information about RSSI attenuation over space and stability over time. Data is filtered to show only results from the two BSSID values associated with “eduroam” inside the Geocommons room (“eduroam” at 2.4GHz, and “eduroam” at 5.8GHz). For spatial analysis, two plots are made: A heatmap of the average RSSI inside the Geocommons room, and a scatter plot showing the average RSSI at different distances. For temporal analysis, plots are made to evaluate RSSI stability over time at different measurement locations.

A simple neural network model is trained on the data from WiFi scans. It takes input of signals from a single 4-second scan. It looks at BSSID, Frequency, and RSSI values as input. The raw data contains observations of all WiFi signals the Raspberry Pi device detects. This can be from any number of access points or other devices (like a phone hotspot). If there is enough information that is relatively consistent, this data can be used as additional input to predict location. Four neural network models are created and evaluated using RMSE, MAE, and R2 values:

- Prediction of distance from origin, trained on geocommons access point signals.
- Prediction of distance from origin, trained on the full dataset.
- Prediction of (X, Y) coordinates, trained on geocommons access point signals.
- Prediction of (X, Y) coordinates, trained on the full dataset.

The geocommons access point training data consists only of signals from the six BSSIDs that emit from the Geocommons room access point. The full dataset consists of all available WiFi signals, minus a known noisy signal that was emitted from a cellular phone’s hotspot. The hotspot negatively impacts results of neural network training when included in the data.

Finally, results from all these analyses can be combined into a fully interactive plot, allowing for an in-depth analysis of indoor localization using WiFi signals.

**Results**

Figure 3 shows the heatmap of the average RSSI value for two signals in the geocommons room. While it is not perfect, there is a general trend where the RSSI signal becomes lower as the distance increases. RSSI is on a logarithmic scale from -30dBm to -90dBm, and every 3dBm is considered a halving of signal power. Directly near the access point device, there is good signal strength. Though, at location (5, -5), there is a low point, this could be due to slight interference from the wall that blocks direct line of sight at that point. A dropoff in signal strength as distance increases can especially be seen in the corners of the right plot, which shows 5.8GHz WiFi signals, which tend to have a shorter range than 2.4GHz WiFi.

Figure 3: Heatmap of Geocommons Room for RSSI Signal Strength.

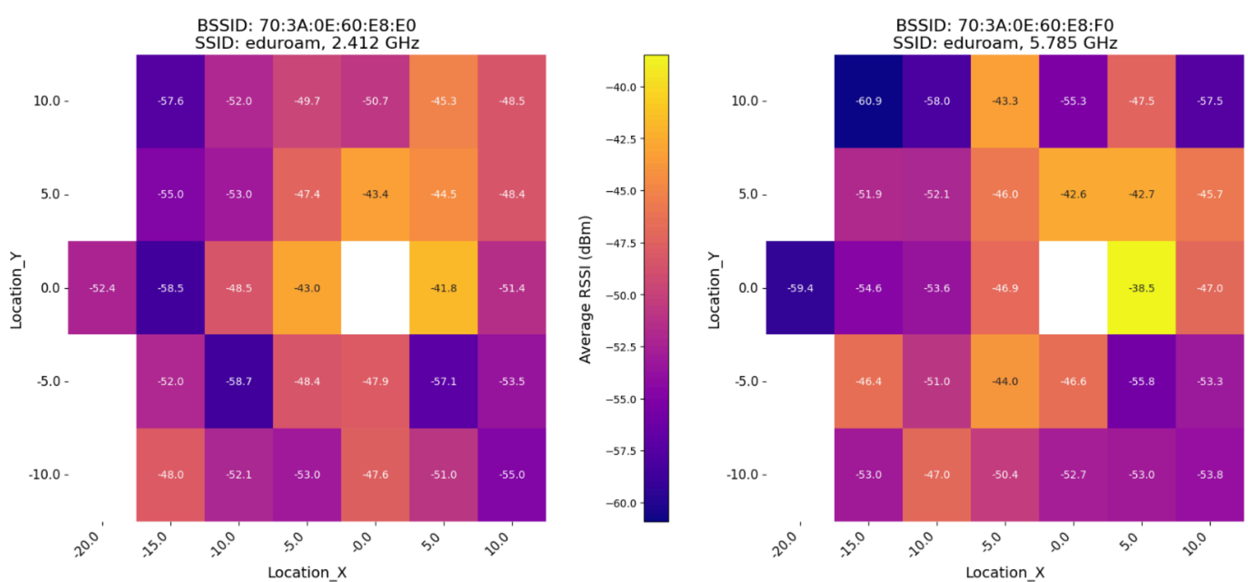
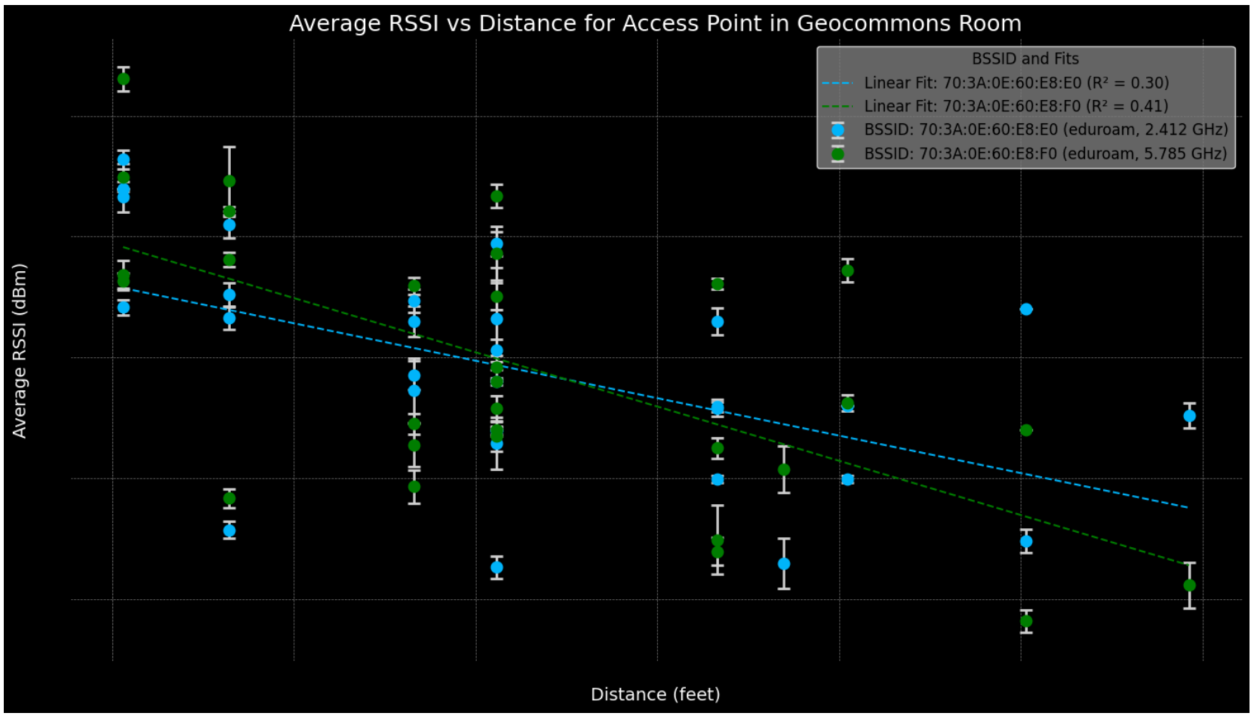



Figure 4 supports the heatmap observations by looking at the RSSI over distance. There is a downward trend, showing that as distance increases, RSSI also decreases. In a relatively small room like this, the signal strength stays fairly strong, meaning there is not quite enough data to show the entire RSSI attenuation from perfect signal to zero signal. We can see that the WiFi signals are not perfectly consistent, there is a fair amount of variance at every distance, hinting that other factors than just distance can affect the quality/strength of a signal. Overall however, the reduction in RSSI over distance shows potential to predict distance based on RSSI value.

Figure 4: Distance vs RSSI for Access Point Located in Geocommons Room.




For temporal analysis, plots showing the RSSI value over time were created at each location. Figures 5 and 6 show two example plots at the locations (-10, -10) and (10, 10). In Figure 5, the RSSI value for both 2.4GHz and 5.8GHz signals remained fairly constant. In this 3 minute scan, there was little interference or changes in the WiFi. In contrast, Figure 6 has significantly more variance in the data, especially on the 5.8GHz network. There does not seem to be any regular interval in the variance, which may make modeling a greater challenge. The difference in these results could be due to many factors, such as random network congestion, multipathing (wifi signals bouncing off walls/windows), or, most likely, a human body being a physical interference to the signals. This measurement inconsistency damages the quality of the full dataset, but gives valuable insight to the sensitivity of these WiFi signals. Minor changes in the environment can very likely cause significant variations in the measured RSSI value.

Figure 5: RSSI Over Time at Location X=-10, Y=-10.

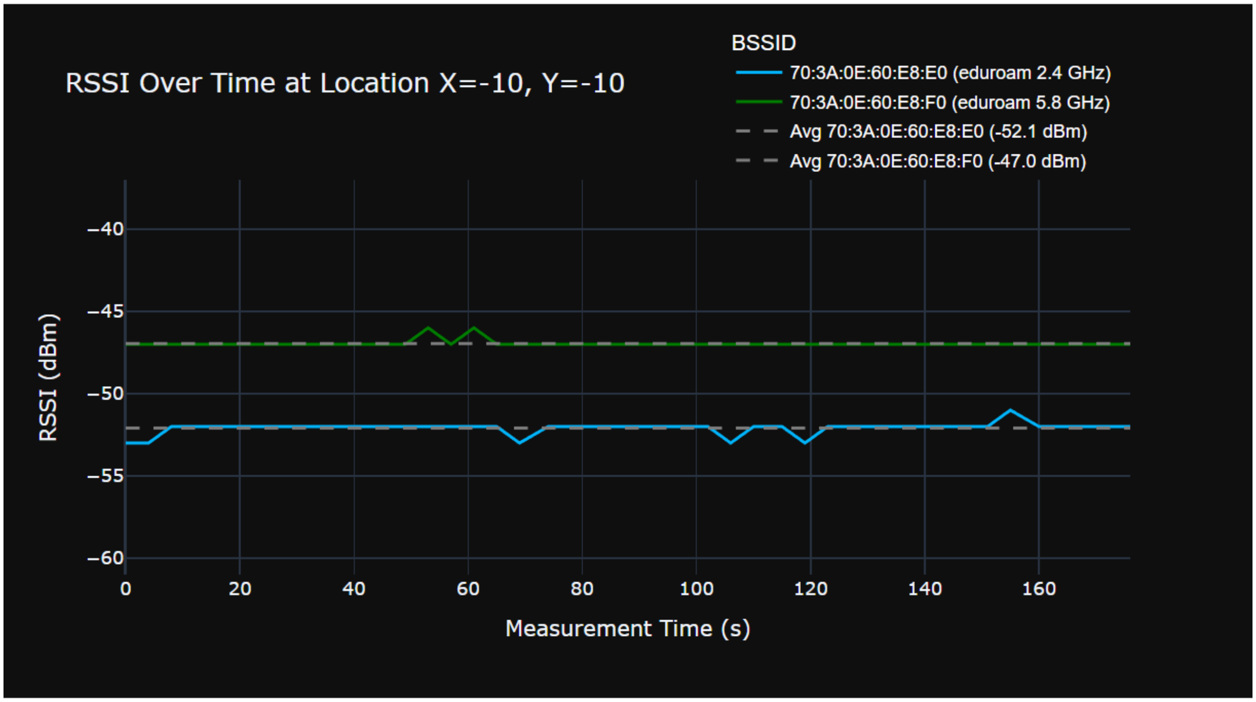



Figure 6: RSSI Over Time at Location X=10, Y=10.

Table 2 below compares the four neural network models that were trained on the data. The first model was trained on only data from the access point in the Geocommons room and predicted distance from the origin. This has a decent accuracy, showing that it can predict distance within a few feet even with data from a single access point. When we expand the training dataset to all access points that were detected, the distance predictions become much more accurate to about 1 foot of mean absolute error. This extra training data significantly increases the overall prediction accuracy, even if the additional wireless signals have to travel through walls & other materials.

The second set of models try to localize the device's location within the room. When training on just the access points within the Geocommons room, the model does not have extremely high accuracy, with 4-5 feet of average error. This is understandable, since the RSSI value likely spreads equally in all directions, it is difficult to model which direction a device is located relative to one access point signal. This explains why the "Distance" predictions trained on the same dataset are able to model to a higher accuracy score. This limited model still has an $R^2$ of 0.65, showing some ability to explain variances. There also could be some dataset overfitting present due to the model picking up on measurement error with the limited one-night of measurement.

The final model predicts X & Y coordinates based training data from the entire WiFi scan dataset. It was able to predict with high accuracy. There was only a couple feet of error in predictions, and an $R^2$ of about 0.97. This is a strong model score, showing that, based only on a single 4 second scan, the model has the potential to locate a device within a few feet of its actual location.

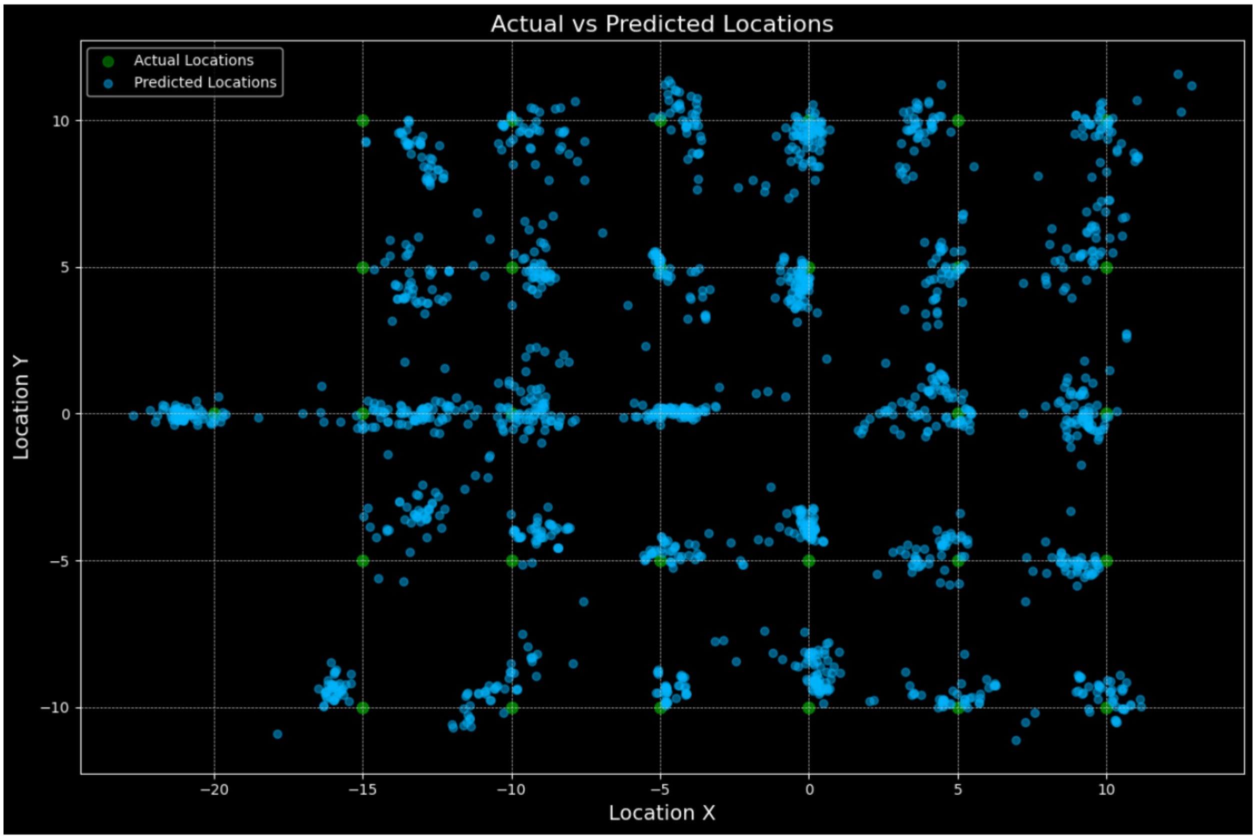
Table 2: Neural Network Model Accuracies

| Model | RMSE | MAE | $R^2$ |
|---|---|---|---|
| Train: Geocommons Room<br>Output: Distance | 2.30610 | 1.74366 | 0.72421 |
| Train: Full Dataset<br>Output: Distance | 1.33227 | 1.00731 | 0.90795 |

| | | | |
|---|---|---|---|
| Train: Geocommons Room<br>Output: X, Y Location | 4.62025 | 3.66511 | X Location: 0.65452<br>Y Location: 0.63565<br>Average: 0.64508 |
| Train: Full Dataset<br>Output: X, Y Location | 1.26567 | 0.87259 | X Location: 0.97238<br>Y Location: 0.97558<br>Average: 0.97398 |

Figure 7 visualizes the final model's predictions to a map of the Geocommons room. In this image, we can see where the model tends to miss predictions. In general, it seems that the model misses predictions "in-line" to the access point. For example, look at points (-5, 0) and (-10, 0). The predicted locations near these points are distributed in a way that seems to point towards the origin. That makes it seem that the model is good at figuring out the general direction from the origin, though still has small errors with absolute distance/location. It also tends to under-predict the location, as many predicted locations are closer to the origin than the actual location.

Figure 7: Full Dataset Neural Network Model Location Predictions




For a final analysis, the visualizations from Figures 5/6 and Figure 7 were brought together into a single interactive plot. An example is shown in Figure 8 and the plotting application can be accessed at the website https://z.umn.edu/IndoorLocationTrackingModel . Clicking on points on the left plot displays the neural network predictions and a time-series of RSSI values that were associated with that location. Clicking through multiple points, it is a little bit difficult to figure out the causes of prediction inaccuracy, though it is known that RSSI from within the geocommons room is only a partial factor in model predictions. This interactive plot helps to visualize outliers and show as a proof of concept for further localization analysis applications.

Figure 8: Screenshot of Interactive Plotting Application.

**Wi-Fi Signal Strength Analysis with NN Predictions**


### Conclusion

Overall, the results from this indoor localization study support the idea that WiFi signals can be used as a measure to improve location estimation within indoor environments. By utilizing the Received Signal Strength Indicator (RSSI) value from WiFi access points, it is possible to estimate distance from the access point and coordinates of a device. Prediction of distance from the access point is useful for things such as presence detection and determining when a device is within a certain proximity of the access point in the room.

Additionally, the ability to localize a device to exact coordinates has even greater potential for many applications. The neural network models had a high accuracy in predicting (X, Y) coordinates with minimal error, especially when trained on the full dataset of all detectable WiFi signals. An $R^2$ value of approximately 0.97 indicates that the model can reliably explain the variance in the data, showing the effectiveness of using WiFi signal data for precise indoor localization.

The final results come with several limitations however. The initial study was conducted in an idealized scenario with minimal network interference, which may not reflect more active and crowded environments. Even in this controlled setting, minor interferences such as phone hotspots and movement of people or objects in the room have a noticeable impact on measured RSSI values. In active environments with more people, devices, and wireless noise, the accuracy of the model predictions could be significantly reduced. These factors highlight the sensitivity of WiFi signals to environmental changes and signal noise.

Despite these limitations, the good performance of the neural network model shows a high potential for this approach to be effective in future applications. With further refinement and adaptation, WiFi-based indoor localization can be enhanced to handle more complex and variable environments. Future studies could expand on this work in several ways, including:

- Scan the same area over longer periods of time to capture and account for additional temporal variation.
- Perform scans during busy days to see how increased network traffic and physical movement affect the data.
- Cover a larger area, such as an entire floor, to evaluate scalability.
- Analyze contributing factors, such as WiFi signals coming from nearby rooms, more deeply to further understand the model.
- Modify the neural network to incorporate multiple observations using Long Short-Term Memory (LSTM) models, which may better handle temporal variations in the data.
- Incorporate measurements from other sensors, such as Bluetooth beacons or accelerometers, to enhance precision and gain higher refresh rates for location tracking.

This study confirms the ability to utilize WiFi signals in indoor localization. The strong results, particularly with neural network modeling, demonstrate the ability for this approach to gain precise results from quick scans. Continued research and development can refine WiFi-based indoor localization to achieve greater accuracy and reliability, paving the way for adoption in a wide range of real-world applications.

## References


Obeidat, H., Shuaieb, W., Obeidat, O., & Abd-Alhameed, R. (2021). A review of indoor localization techniques and wireless technologies. *Wireless Personal Communications*, *119*(1), 289–327. https://doi.org/10.1007/s11277-021-08209-5 .

OpenAI. (2024). *ChatGPT* (GPT-4o & o1 models). https://chat.openai.com/ .

Ranacher, P., Brunauer, R., Trutschnig, W., Van der Spek, S., & Reich, S. (2015). Why GPS makes distances bigger than they are. *International Journal of Geographical Information Science*, *30*(2), 316–333. https://doi.org/10.1080/13658816.2015.1086924 .

Zafari, F., Gkelias, A., & Leung, K. K. (2019). A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials*, *21*(3), 2568–2599. https://doi.org/10.1109/comst.2019.2911558 .

**Appendix**

Table A1: Networking Terms of Interest. This table lists important terms used in the paper, useful for reference.

| Term | Description | Example |
|---|---|---|
| WiFi Network | A given WiFi signal connection that is used by many devices. | This is the thing that people think of when they “connect” to WiFi, like the “eduroam” network or “homeWiFi27” network. |
| Frequency | The wireless signal frequency that the WiFi network is using. | This is the airwaves that WiFi signals travel along. Sometimes this is shown by the 2.4G or 5G label that is seen on WiFi network names. |
| Access Point | A physical hardware device that provides WiFi connectivity and can host multiple WiFi Networks at multiple frequencies. | This is the physical device in a room that sends out WiFi network signals. This can send out multiple WiFi networks at the same time, such as “eduroam”, “UofM-Guest”, and “UofM-IoT”. Similar to a router that people have at home, but for larger-scale applications. |
| SSID | The text name for the wifi network | This is the actual name “eduroam”. |
| BSSID | A unique identifier for a WiFi network from the Access Point, WiFi Network, and | This uniquely identifies a specific WiFi signal, it looks like this: |

| | | |
|---|---|---|
| | Frequency. | “70:3A:0E:60:E8:F0”. Similar to a “MAC Address” that uniquely identifies phones/laptops/devices. |
| RSSI (Received Signal Strength Indicator) | This is a measure of signal strength for a WiFi signal being sent out, measured in dBm. -30dBm is considered a perfect signal, and -90dBm is considered no signal. | This is a factor in the quality of a WiFi network. When a person sees , this likely means the RSSI is low and far away from an access point signal, versus  likely has a high RSSI value and is near to an access point. |