Indoor Location Sensing Using Smartphone Acoustic System

Combining Acoustic and WiFi localization

Dāvis Kažemaks

Supervisor: Qun Song

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Dāvis Kažemaks
Final project course: CSE3000 Research Project
Thesis committee: Qun Song, Jorge Martinez Castaneda

Abstract

Indoor localization is an important field of research for advancing robotics and providing more accurate estimations of indoor locations for users. There are many indoor localization algorithm implementations, but many of them underperform under certain environmental changes or restrictions. This research will present a way of combining already existing indoor localization techniques to more accurately deduce the user's location within a building. An experiment was conducted within campus building Pulse, where multiple fingerprints of locations where gathered, and then used to train and test the combined classification models. By fusing active acoustic location sensing and WiFi localization using weighted averaging, ensemble stacking, and 2-step localization, the combination of classifiers was able to outperform individual classifiers by up to 5% of localization accuracy. Additionally, 2-step localization and weighted averaging methods did not add any performance overhead.

1 Introduction

Outdoor location sensing on mobile devices has been used for decades, but in recent years, there has been a significant demand for introducing indoor localization. Having reliable indoor location sensing can be a viable tool for navigating office spaces, museums, or any other large buildings. In hospitals, room-level localization can be particularly useful for finding patients quickly that need urgent medical care, the same idea can be applied to any other high-risk environment. Location sensing has additionally been used in robotics, to navigate a robot through workspaces or domestic environments [12].

Global positioning system (GPS) is the most widely used method for location sensing [11], providing accurate outdoor localization. Unfortunately, GPS heavily relies on satellite signal strength, which is significantly weaker indoors [11].

To address this, other methods of location sensing have been developed. These methods include radio frequency-based, barometric-pressure pattern, geomagnetism, and acoustic localization. Radio frequency-based approaches require an infrastructure to enable localization [1; 19; 10], barometric-pressure pattern localization algorithms do not perform well in same-level floor classification [28; 26], geomagnetism [7] and acoustic location sensing [24; 25; 23] room classification accuracy is heavily influenced by the environment and surroundings. This suggests that combining different localization methods can increase localization accuracy, by increasing the number of unique features that can be extracted from the environment, and reducing the dependency on each individual environmental attribute. Combining localization fingerprints and algorithms has been attempted before in [25; 4] but it only covers a very small subset of feasible localization method combinations.

This research will present a way of combining already existing localization algorithms, specifically, active acoustic location sensing and WiFi location sensing, to further increase the accuracy of classifying rooms or determining a position in large open indoor spaces. This implementation can be used as proof of concept in future research and presents ways to improve consistency for indoor localization algorithms. The following sub-questions will be addressed:

1. How can combining WiFi and active acoustic localization algorithms increase the accuracy of room-level localization?
2. Does combining WiFi and active acoustic localization algorithms make the application significantly slower?
3. Can the combination of WiFi and acoustic localization algorithms decrease the accuracy in specific scenarios?

Acoustic-based indoor localization uses the acoustics of the environment to determine the user's location. Acoustic localization can be split into 2 main subcategories: passive and active. Passive acoustic location sensing utilizes background acoustics to classify the current location, while active location sensing emits chirps and listens for their echo. Active acoustic sensing has been shown to outperform passive acoustic sensing both in accuracy and robustness against interferences in the environment [24; 25]. Recent studies have shown that utilizing a Convolutional Neural Network (CNN) in active sensing can produce an accuracy as high as 99% in certain environments [24]. Additionally, active acoustic location sensing is an infrastructure-free way of accurately assessing a smartphone's location. This allows it to be easily deployed in any indoor environment without any external physical setup. Because of these reasons, active acoustic sensing was chosen.

WiFi localization utilizes the nearby WiFi routers and their signals to locate the user. WiFi location sensing algorithms have shown great accuracy in classifying rooms in a static indoor environment with multiple WiFi access points [1], and having less than a meter localization error within a building [27]. Most public places in Europe have multiple WiFi access points spread around within a building. In the building where the experiment will take place, it has been measured that on average there are 22.32 WiFi beacons within an area (see Appendix A). Additionally, it has been shown that passive acoustic sensing misclassification errors are not geospatially clustered [25], while WiFi localization tends to misclassify nearby locations. By combining both of these localization approaches, it can be possible more accurately deduce the location.

To combine the localization methods, 2-step localization, weighted averaging, and ensemble stacking will be explored. 2-step localization first performs WiFi location sensing to gather an ordered list of the likeliest locations. Afterward, the acoustic localization method is used to pick the likeliest label from the supplied list. This method has been shown to increase the accuracy of localization when combining passive acoustic location sensing and WiFi location sensing [25]. Weighted averaging between multiple sources of predictions has been used in forecasting for decades, and it has shown that using aggregates of information produces more accurate results in comparison to only using a single source of information [6; 17]. To implement this in localization, the probabilities of each label from both classifiers are obtained, then

the labels are scaled by an assigned weight of the classifier, and finally summed up to create a weighted sum. The label that has the largest weighted sum is then picked as the prediction for the fingerprint. Ensemble stacking is a method that is used to create a meta-classifier that uses other classifier outputs as its input. This method has been shown to improve classification accuracy in different fields of research [2; 20], but has not been sufficiently explored in indoor localization.

The rest of the paper is organized as follows. Section 2 goes into more detail on previously done research in this field. Section 3 will provide reasoning on why certain measurements are taken and how to craft them into fingerprints. Section 4 describes the methodology that was followed to achieve the answer to the research question. Section 5 will show and explain the proposed implementation that will test the hypothesis of this research. Section 6 will discuss the testing environment where the implementation will be deployed, and analyze the obtained results. Section 7 will discuss the ethical implications of my research and the reproducibility of the results. In Section 8, results will be compared to previous work. And finally, in Section 9 the main research question will be summarized and answered.

2 Related work

When it comes to indoor localization using a smartphone, there are 2 main categories they fall into infrastructure-dependent and infrastructure-free. Infrastructure-dependent localization methods rely on a preexisting framework that allows users to derive their location. These localization approaches tend to use Radio frequency-based location sensing since it enables the mobile device to connect to external networks. Infrastructure-free localization does not require any physical setup for location sensing and relies on built-in sensors that are available on the mobile device. Most common approaches use an acoustic system, barometer, or magnetometer built inside the phone to measure and classify the room.

2.1 Infrastructure-depednet

One of the early Radio frequency-based localization implementations is RADAR which used triangulation with 3 access beacons to determine the indoor location [5], and had an accuracy error of 2-3 meters. In more recent studies, local networks and WiFi are utilized as access points, which makes deploying algorithms in office spaces more convenient [1; 27; 14]. Horus used access point signal strength to construct radio maps of locations within a building, during discrete location estimation approximated the signal strength using parametric distribution and then applied them to the radio map to locate the user. It was able to outperform RADAR's accuracy by 89% on certain testbeds [27].

In [1] the Gaussian fit was used to more efficiently store the signal strengths of base stations in specific states (locations). Then Bayesian localization was applied to obtain the state (location) of the user. Wireless location sensing for classifying rooms within a building was able to reach an accuracy of 95% in a static environment. Still, the performance seems to be highly reliant on the infrastructure [1]. By decreasing the number of WiFi beacons, the accuracy decreases by more than 20%. Within the same study, it was also shown that civil traffic in the hallways or rooms can heavily influence the reliability of location estimation.

In [14] it was shown that WiFi signal strength fluctuates over time, which requires the models to be retrained multiple times for them to give accurate results. This would require users to actively participate in model training, which defeats the purpose of localization since the user would have to be able to locate themselves. WASP [16] addresses the signal strength fluctuations by introducing weighing and filtering of access points based on their visibility. Access points that are occasionally visible will be filtered out or get a lesser weight assigned to them to reduce the reliance on access points with inconsistent signal strengths. With this approach, WASP was able to significantly outperform Redpin [8] algorithm, which was considered state-of-the-art at that time.

Less common Radio frequency-based solutions include Bluetooth [3; 19] and FM-based [10] localization. Bluetooth location sensing has similar results to WiFi location sensing but does require static Bluetooth beacons to be set up. FM-based localization has shown great success in room-level recognition, having almost the same level of accuracy as WiFi localization and having less fluctuation of signal strength over time. But this approach requires the mobile device to have an antenna or earphones, and the quality and position of these antennas can severely impact the accuracy.

2.2 Infrastructure-free

Some smartphones come with built-in barometers that can be used for location sensing. While barometers have shown success in vertical positioning [28; 26], they are not able to detect horizontal changes in position.

Geomagnetic localization is another very active field of research, and some studies have shown more consistent localization accuracy than WiFi [13]. Unfortunately, these implementations are very sensitive to local magnetic waves produced by the environment, which reduces the overall accuracy of these systems [13; 7].

A more accessible approach is to use acoustic location sensing to position the device. Acoustic localization can be split into 2 classes: passive and active location sensing.

Passive acoustic sensing only uses background noise to fingerprint and recognize the rooms. A successful implementation of that is described in [25] which is able to achieve room-level localization accuracy as high as 69%. The downside of passive sensing is that it is very susceptible to changes in the environment, for example, turning off the AC, switching off appliances in the room, amount of people in the room, etc.

Active acoustic sensing uses the speaker of the mobile device to emit chirps, which create echoes in the room that can be captured with the microphone. Since room sizes and shapes differ per room, it has been shown that these echos carry distinguishing features that can help classify a room. A study has shown that using the active approach and feeding the spectrograms into a two-layer convolutional neural network, it is possible to reach room-level localization accuracy as high as 99% [24]. RoomSense [23] used a Support Vector

Machine classifier instead of a convolutional neural network. It was able to reach a similar accuracy to RoomRecognize [24]. The downside of RoomSense in comparison to RoomRecognize is that it uses the entire audible band, which makes it more susceptible to interference. It has been additionally shown that the time-of-flight of the reflected chirps can be used to determine the location within a room with a median error of 12.4cm [15].

2.3 Combined Localization

There are multiple ways of combining localization methods. In [25] and [4], fingerprints were combined with linear combination distance or joined the fingerprints in a single tuple. These new fingerprints were then supplied to the classifier to obtain the label. Even though this approach has been shown to increase accuracy, it will be not used in this research, since WiFi localization methods and acoustic location sensing methods use different types of classification models.

Multi-step localization was used in [25], which was able to significantly increase the localization performance. It uses 2 or more separate localization methods that each perform localization, compiles an ordered list of likeliest labels, and supply this list to the next classifier to then again compile the likeliest labels from the provided list. After each level of localization is complete, the likeliest label in the final list is chosen. In the same study, it was shown that 2-step localization can achieve higher room-level classification accuracy than using individual localization implementations.

Weighted averaging or information aggregation has been used in forecasting for combining different sources of predictions. This method allows one to combine multiple sources of estimates and weigh them according to their reliability. Studies have shown that using aggregates of information produces more accurate results in comparison to only using a single source of information [6; 17].

Ensemble machine learning is a general approach to increasing predictive accuracy, by combining multiple machine learning models. In ensemble stacking, multiple models are run on the same data points, and each of their predictions is supplied to the meta-model for final classification. This approach has shown great accuracy performance in other domains [2; 20], so it is expected that it can additionally improve localization.

3 Measurement collection

This section will describe the measurements that were conducted for this experiment. The acoustic data collection and transformation into a fingerprint will be presented, which then will be followed by the same analysis but with WiFi localization.

3.1 Acoustic data

There are two main methods of acoustic data collection: passive and active. The passive approach collects all ambient sound, while the active approach produces a chirp from the speaker and waits to receive the echo that will be reflected by the room.

Passive acoustic location sensing can utilize many subsets of frequency ranges for data collection. In [25], different subsets of frequency bands were tested against each other to find the most optimal accuracy in the environment. It was shown that in quiet environments, the frequency band between 0 - 7kHz was the most optimal, with an accuracy of 69%. Changing to a noisy environment, there was a substantial drop in accuracy to just 3%. This was due to the fact that this frequency range included the speech band, which is around 300 - 3000Hz. When reducing the band to 0 - 300Hz, the accuracy increased to 63.4%. These results suggest that reducing the frequency band range omits certain unwanted ambient noises, but negatively impacts the accuracy when there is no ambient noise since there is less data to use for the fingerprinting.

Active acoustic localization depends on the frequency that the chirp is emitted. RoomSense [23] emitted a chirp that spanned the frequency range 0 - 24kHz. While giving great results in quiet environments, in [24] it was shown that RoomSense underperforms in environments with obstructing ambient noise. In the same research, the frequency range of 19.5 - 20.5kHz was analyzed, while emitting a chirp only in 20kHz frequency. This approach was more robust to ambient noises, and only slightly decreased the overall classification accuracy.

Combining these observations, the data collection using the RoomRecognize [24] approach was chosen. Additionally, the frequency close to 20kHz is inaudible, which should make the fingerprinting process silent. Unfortunately, this is not the case, since audio systems on smartphones are unable to correctly recover from emitting the chirp, producing a ”tick” noise.

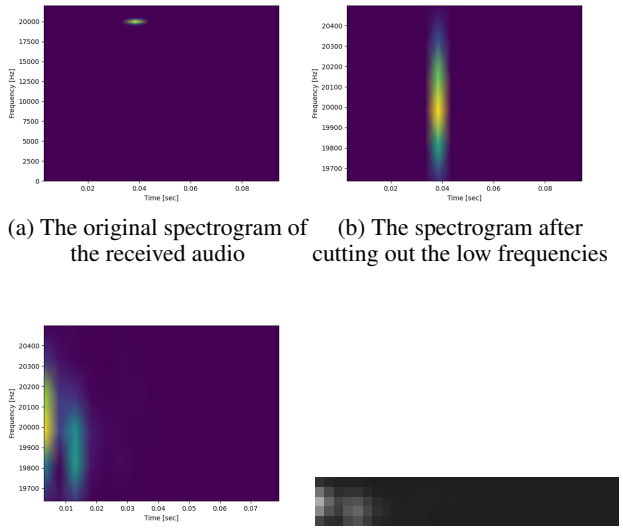
3.2 Acoustic fingerprint

The active acoustic localization will be closely following the implementation of RoomRecognize [24]. To gather the echo dataset, a large audio file will be recorded while chirps are being emitted every 100ms. The chirp is a 2ms 20kHz sound wave.

To modify and extract echoes from the full audio file such that it fits the input requirements of the CNN, the following transformations need to be applied. First, the correct frequency range of 19.5 - 20.5kHz will be focused on and the rest will be omitted from the spectrogram. Afterward, the spectrogram will be shifted in such a way, that the original chirp is cut out and the spectrogram only contains echos. During experimentation, it was found that the chirp takes 20ms to fully fade out due to direct propagation from the speaker to the microphone. Once this is finished, the spectrogram must be changed into grayscale and downscaled to 32x5 (time and frequency respectively). This procedure is repeated until all the echo segments from the audio file have been extracted. This procedure is visually represented in Figure 1.

3.3 WiFi data

There are 2 main methods used for WiFi data collection: Received Signal Strength Indicator (RSSI), and Round Trip Time (RTT). Received Signal Strength has been the most widely used method of WiFi data collection for localization [16; 27; 1; 14; 8]. But in more recent studies it has



(a) The original spectrogram of the received audio

(b) The spectrogram after cutting out the low frequencies

(c) The spectrogram after applying an offset and cutting out the original chirp

(d) The final fingerprint that will be fed to the convolutional neural network

Figure 1: Transformation of the original spectrogram into a fingerprint

been shown that RTT is more accurate for indoor localization and more robust to signal change over time than RSSI [22; 29]. Unfortunately, RTT support is only available on very few models of smartphones, so it would limit the application to a very small user base. For this reason, RSSI will be the method of WiFi data collection.

3.4 WiFi fingerprint

RSSI data can be crafted into multidimensional vectors, where the dimension of the vector is the total amount of routers. Each signal strength received from the router ranges from 0 to 100, 0 meaning that the router is not in range, and 100 meaning that the router is right next to the device. This kind of fingerprint is acceptable as input for most machine-learning models.

4 Methodology

To increase the room-level localization accuracy, the following steps will be outlined. First, the acoustic localization method will be described and analyzed. Then the WiFi localization algorithms will be explored and compared. Lastly, the fusion of both localization algorithms will be discussed and implementation will be proposed.

4.1 Acoustic localization

RoomRecognize [24] implementation will be used as the baseline for the active acoustic classifier. In the study, it was able to reach a very high room-level classification accuracy and is more resistant to interferences than any other active acoustic localization implementation [24; 23].

RoomRecognize uses Convolutional Neural Network as its machine learning model and uses echo spectrograms as training data. The procedure of how to create an acoustic fingerprint is described in subsection 3.2. CNN is a deep-learning algorithm used for classifying imagery. It consists of multiple layers of filters for feature extraction and then supplies the feature maps to dense layers, where the classification happens. Detailed implementation of the CNN model can be seen in [24] (Chapter 4.2).

4.2 WiFi localizatoin

WiFi localization will use Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN) models for labeling the data. SVM and k-NN were chosen as classification methods for WiFi fingerprints because they have shown great room-level classification accuracy [16] and are fairly simple to implement.

SVM is a supervised machine learning algorithm that attempts to create a decision boundary that can segregate n-dimensional space into classes. When classifying a fingerprint, the algorithm checks which side of the boundary the fingerprint belongs to. K-NN is a machine-learning model that is trained with labeled fingerprints. During classification, it compares the given fingerprint to the k nearest neighbors and uses a majority vote to assign a label. These machine-learning algorithms are a part of scikit-learn python library [1], and this library will be used in the implementation.

More complicated algorithms like Redpin or WASP described in [16; 8] will not be implemented due to the limited time of this research. Additionally, high individual accuracy of classifiers is not necessary, since the focus of this research is to improve the accuracy of fusion.

Since KNN and SVM are discriminative models, they need an additional implementation layer to turn them into probabilistic models, which is a necessity for weighted averaging. This is fortunately implemented within the scikit-learn library. For k-NN, the probability is calculated by dividing the class label neighbor amount by the total neighbor amount k. For SVM, Platt scaling is used to map classified labels to probability scores.

4.3 Fusing WiFi and acoustic localization

To combine both localization techniques, weighted averaging, 2-step localization, and ensemble stacking will be used.

Weighted averaging between multiple sources of predictions has been used in forecasting weather and it has shown that using aggregates of information produces more accurate results in comparison to only using a single source of information [6; 17]. To implement this in localization, the probabilities of each label from both classifiers are obtained, then the labels are scaled by an assigned weight of the classifier, and finally summed up to create a weighted sum. The label that has the largest weighted sum is then picked as the prediction for the fingerprint. The hyperparameter that will be tweaked is the weight that will be assigned to each classifier. Since there are 2 classifiers, only 1 of the weights needs to

[1] https://scikit-learn.org/stable/

be tweaked to change the proportions of probabilities. The weight for acoustic predictions was chosen in this paper.

2-step localization was used to combine passive acoustic location sensing with WiFi localization [25], and it was able to significantly increase the overall localization accuracy. Hence, it is expected that replacing passive acoustic localization with active will yield similar results. 2-step localization works by first performing WiFi location sensing to gather an ordered list of the likeliest locations. Afterward, the acoustic localization method is used to pick the likeliest label from the supplied list. The hyperparameter that will be tweaked is the number of labels in each supplied list.

Ensemble stacking will be performed by taking the output of both acoustic and WiFi classifiers and supplying it to another machine-learning model. Models that will be used as the meta-model are k-NN, SVM, and logistic regression.

Other fusion methods like linear combination [25] or fingerprint joining [4] were used to fuse together the fingerprints, rather than combining the predictions. These kinds of methods will not be looked into in this research, since it would require changing the CNN model of the acoustic localization model, which is not feasible to implement due to time restrictions.

5 Implementation

This section will describe the implementation of the application that uses 2-step localization and weighted averaging to classify the current room. The source code can be found in Github [2]. In subsection 5.1 the overall architecture will be presented. Next in subsection 5.2 the most important features of client-side will be showcased. Finally, server-side implementation will be explained.

5.1 Client and server architecture

The application is split into 2 parts: client-side and server-side. The server was added to the architecture so that the client can offload training and labeling tasks to a more powerful machine.

Each component in the architecture has its own set of features. The client is responsible for displaying an easy-to-use front end to the user for labeling and classifying rooms. The server is responsible for capturing the client's requests and handling them.

5.2 Client implementation

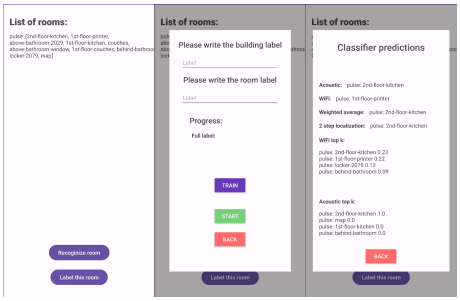
The client was built using Android Studio[3] using Java as the programming language. Since this was quite a large task to do within this research, the client side of the application was developed collaboratively within the research group.

The client has 3 main screens: home, labeling, and recognition. Each screen encapsulates features that are necessary for creating an effective localization tool. All of these screens can be seen in Figure 2

The home screen is the first screen the user interacts with. The screen shows all the buildings and their rooms in a list. This screen helps the user understand whether the location they are in right now is already labeled, or requires labeling. From the home screen, it's possible to navigate to the labeling screen or recognition screen.

The labeling screen allows the user to input building and room labels, and then collect data samples for the current location. The labels and data samples are then sent to the server, where the information is crafted into a fingerprint. The data collection happens in 10 cycles. In each cycle, the first action is to collect the WiFi fingerprint and set a timeout for 10 seconds. The timeout is necessary for the WiFi API to finish scanning the local network, otherwise, the previous scan result will be returned. Afterward, 20 chirps are emitted from the mobile device while recording the audio. At the end of the cycle, all the collected data is sent to the server to properly process the data into fingerprints. Note that the server drops the first 5 and last 5 chirps, since they tend to contain a lot of noise. After the 10 cycles have been completed, the user can issue a query to the server to retrain the model with the new dataset.

The recognition screen allows the user to acquire their location within the list of rooms. It performs a single scanning cycle, where the WiFi fingerprint and the audio with the echoes are collected and sent to the server. The server then responds with all the predictions from all the classifiers that are running on the server. For WiFi top 3 and Acoustic top 3, the probabilities are also added next to the label that ranges from 0 to 1.



(a) The home screen of the application (b) The labelling screen of the application (c) The recognition screen after the data has been classified

Figure 2: The 3 main screens of the client-side application

5.3 Server implementation

The server is written in Python, using Flask as the web framework for enabling communication between the client and server. The hosting of the server has to be set up manually. When the server is launched, it will display its IP address, which must be manually copied into the client's source code for it to connect to the server.

The server consists of 3 main components:

- Local data storage - since this is not a large-scale project, all the necessary data is stored locally. The spectrograms

[2] https://github.com/kazemaksOG/IndoorNavigationRP

[3] https://developer.android.com/studio

of the echoes for training are stored in `/images`, the WiFi fingerprints are stored in `/wifi`, and some metadata like diagrams and recognition echos are stored in `metadata`.

- Server communication - The server has 4 endpoints that handle requests: `get_rooms`, `train`, `add_room`, and `recognize_room`. Each endpoint handles a specific request from the client and sends back a corresponding response message.
- Classifiers - the server has 5 different classifiers: acoustic, wifi, weighted average, ensemble stacking, and 2-step localization. Each of them must be trained using the fingerprints that are in the local data storage before they can be used for classification. After training, when the classifiers receive a fingerprint(s), they will assign the most likely label to it. The implementation of these classifiers strictly follows the design that was described in Section 4. Additionally, the top 3 likeliest labels for both WiFi and acoustic localization are also computed and sent to the client during recognition.

6 Experimental setup and results

This section will give an overview of how the experiments were conducted, and showcase the obtained results. In subsection 6.1 the details of how and where the data was collected are presented. Then how the data needs to be split for effective accuracy estimates is discussed in subsection 6.2. Finally, in subsection 6.3 and subsection 6.4 the accuracies of each classifier are compiled and analyzed.

6.1 Data collection

The experiment will be conducted in Pulse, which is a building within the Delft University of Technology campus. As shown in Appendix A, the building is equipped with multiple WiFi access points, which is a requirement for the WiFi localization method. Additionally, the building has a significant number of rooms that do not require employee access rights, and rooms tend to be empty during early mornings and late evenings.

The data collection was conducted using Samsung Galaxy A32 (SM-A325f/DS). While performing sampling, the WiFi of the phone must be turned on, and the volume has to be set to maximum. Since acoustic fingerprinting is very sensitive to the environment, the phone was always held horizontally, with the speaker and microphone facing the user, and the user had to stand still.

10 locations in the building were chosen (see Appendix B). In each location, 10 fingerprinting cycles were performed. In each cycle, 1 WiFi fingerprint and 20 acoustic samples (from which only 10 are used) are collected. After each cycle, it is suggested to slightly move the smartphone up, down, left, or right while keeping the same orientation. This is done to reduce overfitting on a very specific location and enhance feature extraction. In total, 100 WiFi fingerprints and 1000 acoustic fingerprints were collected.

6.2 Dataset preperation

To obtain the accuracy of the acoustic and WiFi localization algorithms, the data needs to be split into training, validation, and test sets. The training set is used to train the model for classification. During training, a validation set is used to assess the quality of the trained model and tweak the parameters accordingly. Lastly, the test set is used to obtain an unbiased estimate of the model's accuracy against never before seen samples.

To accurately access the accuracy of all 5 classifiers, the test set must be shared. This is achieved by combining the test set of acoustic fingerprints and the test set of WiFi fingerprints by matching them on corresponding labels. Since there is a much larger amount of acoustic fingerprints than WiFi, the test set for acoustic fingerprints is reduced to be the same size as WiFi fingerprints.

For the acoustic model, the dataset is split into 80% training set, 10% validation set, and $\leq$10% test set. While training the model, the training set is used to train the model, and at the end of each epoch, the validation set is used to access the accuracy of the model. In the end, the test set is used to obtain the unbiased accuracy of the model.

The WiFi model's dataset is split into 50% training set, and 50% test set. The models used by WiFi do not require a validation set while training, so the validation set and test set are merged. WiFi classifier is fit onto the training set, and its accuracy is then evaluated with the test set.

The combined classifiers need both acoustic and WiFi fingerprints for training and evaluating the accuracy. As mentioned before, the test set is already shared between all classifiers by combining them on the labels. The same method has to be applied to the training sets of WiFi and acoustic fingerprints. This combined training set will be then used to find the most optimal hyperparameters for the combined classifiers, and the test set will be used to assess the accuracy.

6.3 Individual accuracy

To avoid over-relying on a single train-test split, 3 random splits were evaluated. All the accuracies of the classifiers can be seen in Table 1. Split 1 will be analysed, while other split confusion matrices can be found in Appendix C

Classifier	Split 1	Split 2	Split 3	Average
WiFi	88%	90%	88%	89%
Acoustic	82%	88%	92%	87%
Weighted average	98%	90%	94%	94%
2-step localization	89%	92%	92%	91%
Stacking	92%	81%	94%	89%
WiFi top 3	100%	100%	100%	100%
Acoustic top 3	100%	98%	100%	99%

Table 1: Test set localization accuracies obtained from 3 different train-test splits

To obtain the most optimal WiFi classifier, GridSearchCV[4] was used. The hyperparameter list can be found in the app's

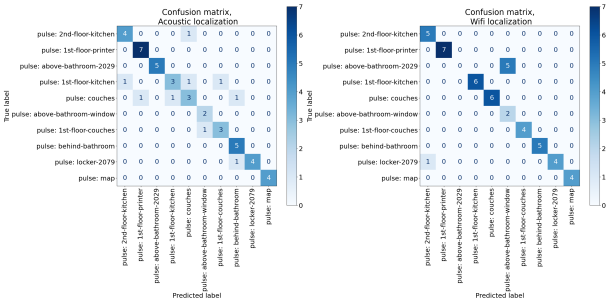
[4]https://scikit-learn.org/stable/modules/generated/sklearn.model$_s election.Gri$

implementation source code. During training, it was shown that SVM slightly outperforms k-NN, and has a more even probability distribution, hence it was chosen as the primary WiFi classifier (see Appendix D).

The acoustic classifier was trained using the CNN and only applying 10 epochs. The reason for such low epochs was to artificially make the classifier underperform slightly, so it does not dominate the combined classification.

The WiFi and acoustic classifiers were evaluated using the test set and the results are visualized in Figure 3. Both acoustic and WiFi localization were able to reach a room-level classification accuracy close to 90%.

By looking at the misclassification errors, a pattern can be observed. The WiFi misclassification errors seem to be geospatially correlated, since `above-bathroom-2029` and `above-bathroom-window` are only 3 meters apart, and between `locker-2079` and `2nd-floor-kitchen` there is only about 3-4 meters. The acoustic classifier does not exhibit the same kind of pattern, and misclassifications appear randomly.



(a) Confusion matrix of the acoustic classifier

(b) Confusion matrix of the WiFi classifier

Figure 3: Confusion matrices of individual classifiers

6.4 Combined accuracy

To get a better overview of the upper bound of the combined accuracy, the top 3 predictions of WiFi and acoustic localization were calculated. Since the top 3 predictions are the ones that will mostly affect the results of the combined accuracy, examining how often the correct label shows up in the top 3 can give a reasonable heuristic of the upper bound of these algorithms. Since the correct label almost always shows up in the top 3 of each classifier, the upper bound of the room-level localization accuracy for this experiment can be estimated to be 100%.

Combined classifiers were trained on the combined WiFi and acoustic fingerprint dataset. Having the same training and test sets helps us to more objectively evaluate each algorithm.

For the weighted average algorithm, weights between 0 and 5 were tested with a step of 0.1. Since there are only 2 sources of predictions, only one of them needs to be scaled by the weight to control the proportion, which was chosen to be the acoustic prediction probability.

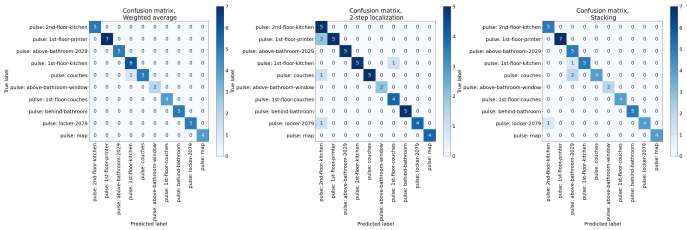
For 2-step localization, the number of predictions per list between 2 and 10 was tested with a step of 1. While training, the best prediction parameters were either 2 or 3.

For ensemble stacking hyperparameter tweaking GridSearchCV was utilized. The approach was used similarly to how it was done for WiFi. During testing, it was found that Logistic regression performed better than k-NN and SVM, hence it was chosen as the meta-model (see Appendix D).

The weighted average algorithm and 2-step localization were evaluated using the test set and achieved the room-level localization accuracies of 94% and 91% respectively. Ensemble stacking underperformed in the second split, which dramatically decreased its average classification accuracy. The confusion matrices of these algorithms can be seen in Figure 4.

Analyzing the confusion matrices, it can be seen that the misclassification errors tend to overlap with faults seen in individual classifier matrices but to a reduced amount. But there is additionally a significant amount of new misclassification errors. This can mostly be observed in ensemble stacking since it adds a completely new layer of machine learning model, and 2-step localization.

The performance of combining localization algorithms depends on the chosen implementation. Weighted averaging does not impact the performance significantly, since it uses datasets that were previously labeled by localization models for weight estimation, and only a single hyperparameter has to be trained. The same reasoning applies to two-step localization to estimate the list size amount. Stakcing of classifiers does produce a significant overhead since it requires an additional machine model to be trained, and the time of training scales with the size of the dataset.



(a) Confusion matrix of the weighted average algorithm

(b) Confusion matrix of 2-step localization

(c) Confusion matrix of ensemble stacking

Figure 4: Confusion matrices of combined classifiers

7 Responsible Research

This section will tackle ethical concerns that can arise from the implementation of accurate room-level localization algorithms. Additionally, the reproducibility of the results of this paper will be discussed.

7.1 Ethical concerns

Indoor localization can have a beneficial societal impact by advancing indoor robotics and helping users locate themselves within complicated and large buildings. But there are certain ethical issues that can arise from this implementation.

Identification from audio files
Acoustic localization needs to record the echoes in the environment that are in 20kHz frequency. Since the microphone of the smartphone also picks up lower-frequency sounds, the voices of the persons around the smartphone will also be recorded. The persons, in this case, are recorded without consent and can be potentially identified by their voices. This can be seen as a privacy violation by some individuals and institutions[21]. To avoid this, only the processed spectrograms of the echoes are stored in the database, which cannot be used to identify the persons in the room.

Since the audio files are sent from the client to the server, the data can be intercepted by a malicious party before being processed. While conducting this experiment it was all done on a local network, so there was no threat of a malicious party, but secure communication protocols must be used if this application is ever used outside of a safe local network.

Unconsentual indoor user tracking
Geolocation tracking is used by many services to both enhance the user's experience, or profit from selling their user data [18]. Usually, the user consents to these practices by approving access to location data, and most of the time they are not informed about how their data is being used. This presents another privacy breach since the user does not directly consent to his location being used by other entities besides the application maintainers.

The application presented in this paper does not collect any identifiable user data. However, if indoor localization is adopted by large corporations, there is not much regulation that can stop them from unethically obtaining user location data.

7.2 Reproducability

This paper showcases both the implementation of the application thoroughly and guides the reader through the entire experimental setup.

In Section 4, each classifier is described in detail, and in Section 5 a direct implementation is presented. Additionally, the source code is available on GitHub [5].

The experimental setup is described in Section 6. It shows exactly how the data was gathered, and in Appendix B showcases the exact spots in which the data was collected.

8 Discussion

In this section, the results of this paper will be compared to other related work in this field of research.

Acoustic localization underperformed in comparison to RoomRecognize [24], even if the epochs were increased. This can be due to the fact that the dataset is too small to conduct proper deep learning, and the RoomRecognize data sample was gathered by using a robot, that more evenly collects the data samples within a room.

WiFi localization was able to reach a very good accuracy, even with a very small training set in comparison to other implementations [27; 16]. This can also be explained by the fact that a lot of locations within the building were quite spaced from each other and that the building did have a very large amount of WiFi access.

The 2-step localization results correspond to [25] since they used the same method to further increase the accuracy of their passive acoustic classifier. Weighted averaging also corresponds to research findings that suggest that aggregating information produces more accurate estimates [6; 17].

Ensemble stacking was able to sometimes outperform the individual localization methods, but not consistently. This can happen due to insufficient training set size. During this experiment, only 50 samples were supplied at most during training, while it's recommended to have at least 500 [9].

9 Conclusions and Future Work

This paper proposes a combination of already existing localization algorithms to further increase the accuracy of room classification. The main focus of localization algorithms in this research was active acoustic localization and WiFi localization. RoomRecognize [24] was used as the baseline for acoustic localization and Support Vector Machine that used WiFi Received Signal Strength Indicators as fingerprints were used for WiFi localization. Combining these 2 localization methods using weighted averaging, ensemble stacking, and 2-step localization, the room-level localization accuracy was increased by 5%, 0%, and 2% respectively. It was found that misclassification of combinations tends to overlap with individual faults, but there was a significant increase in new misclassification errors. Weighted averaging and two-step localization did not significantly impact the performance of the overall classifier, while ensemble stacking required an additional model to be trained.

9.1 Future work

For future work, the dataset can be significantly increased and could account for more noise in the environment. Since it was expected that the dataset for this research will be small, noisy environments were avoided to only focus on assessing the effects of fusing localization algorithms. To properly assess the quality of these combination algorithms, the dataset should be taken in a more dynamic environment.

It was also shown that the classifier's top 3 predictions always contain the correct label, which suggests that adding another layer of classification can increase the room-level localization accuracy even further.

Because of the lack of time, a fusion of fingerprints was not explored in this research, but it has shown to have results on par with 2-step localization [25].

Instead of ensemble stacking, mixture of experts approach can be used. This method distributes the classification problem into subtasks, and then uses a gating model to pick the final output.

A University building WiFi access point amount

To verify that there is a sufficient density of WiFi access points, measurements of the wifi fingerprints were done. The

[5] https://github.com/kazemaksOG/IndoorNavigationRP

fingerprints were collected in the Delft University of Technology campus building Pulse. From the analysis, it is gathered that the mean is 22.32, the median is 23, the minimum value is 4, and the maximum value is 39. These results are visually represented in Figure 5.

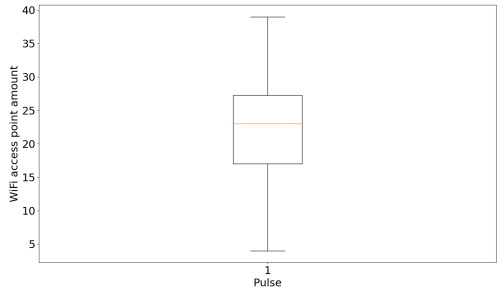


Figure 5: Boxplot showcasing the WiFi access point density within Pulse

B Localization locations

In Figure 6 all the locations where the fingerprints were collected can be seen. All of the locations were within the campus building Pulse.

C Extra confusion matrices

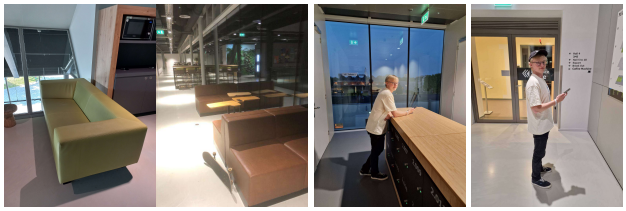
This appendix contains the rest of the matrices that were not analyzed in the results section. See Figure 7 and Figure 8

D Additional accuracies of classifiers

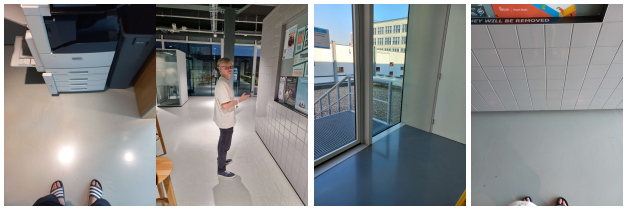
In Table 2 all the additional accuracies that were measured are presented.

Classifier	Split 1	Split 2	Split 3	Average
k-NN (WiFi)	88%	86%	88%	87%
k-NN (stacking)	91%	81%	94%	89%
SVM (stacking)	91%	80%	86%	86%

Table 2: Accuracies obtained from different splits



(a) Couches on the first floor (b) Coches on the second floor (c) Place above 1st floor bathrooms (d) Map on the way between first and second floor



(e) Printer room on the first floor behind Square (f) First floor kitchen, next to the posters (g) Place after going past the first floor bathroom (h) Second floor kitchen, next to the posters

(i) Place above 1st floor bathroom in front of locker 2029 (j) Spot in front of the locker 2079

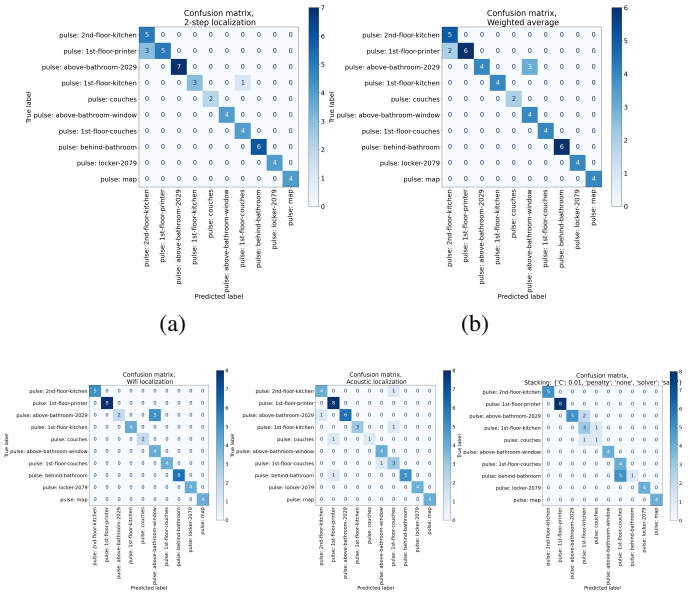
Figure 6: Locations where the fingerprints were collected and localization was tested

References

[1] Practical robust localization over large-scale 802.11 wireless networks. 2004.

[2] Tagel Aboneh, Abebe Rorissa, and Ramasamy Srinivasagan. Stacking-based ensemble learning method for multi-spectral image classification. *Technologies*, 10(1), 2022.

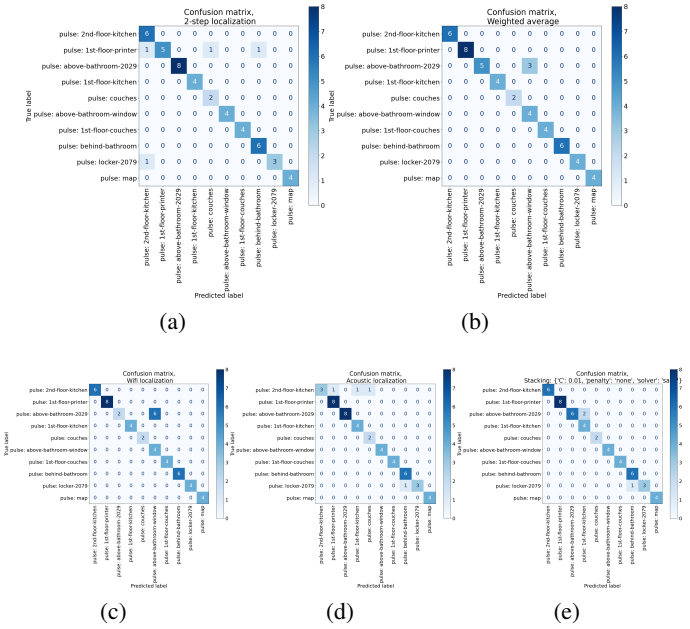
[3] Marco Altini, Davide Brunelli, Elisabetta Farella, and Luca Benini. Bluetooth indoor localization with multiple neural networks. In *IEEE 5th International Symposium on Wireless Pervasive Computing 2010*, pages 295–300, 2010.



(a) (b)

(c) (d) (e)

Figure 7: Confusion matrices for split 2



(a) (b)

(c) (d) (e)

Figure 8: Confusion matrices for split 3

[4] Martin Azizyan and Romit Roy Choudhury. Surroundsense: Mobile phone localization using ambient sound and light. *SIGMOBILE Mob. Comput. Commun. Rev.*, 13(1):69–72, jun 2009.

[5] P. Bahl and V.N. Padmanabhan. Radar: an in-building rf-based user location and tracking system. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, volume 2, pages 775–784 vol.2, 2000.

[6] J. M. Bates and C. W. J. Granger. The combination of forecasts. *OR*, 20(4):451–468, 1969.

[7] Les Blythe and Eva Cheng. Geomagnetic indoor positioning without hardware - the reality of a flawed system, Mar 2022.

[8] Philipp Bolliger. Redpin - adaptive, zero-configuration indoor localization through user collaboration. In *Proceedings of the First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-Less Environments*, MELT '08, page 55–60, New York, NY, USA, 2008. Association for Computing Machinery.

[9] Mohamad Adam Bujang, Nor Ashikin Sa'at, TMITAB Sidik, and Lim Ching Joo. Sample size guidelines for logistic regression from observational studies with large population: Emphasis on the accuracy between statistics and parameters based on real life clinical data. *Malaysian Journal of Medical Sciences*, 25(4):122–130, 2018.

[10] Yin Chen, Dimitrios Lymberopoulos, Jie Liu, and Bodhi Priyantha. Fm-based indoor localization. 06 2012.

[11] Nabil Drawil, Haitham Amar, and Otman Basir. Gps localization accuracy classification: A context-based approach. *Intelligent Transportation Systems, IEEE Transactions on*, 14:262–273, 03 2013.

[12] Weipeng Guan, Shihuan Chen, Shangsheng Wen, Zequn Tan, Hongzhan Song, and Wenyuan Hou. High-accuracy robot indoor localization scheme based on robot operating system using visible light positioning. *IEEE Photonics Journal*, 12(2):1–16, 2020.

[13] Suining He and Kang G. Shin. Geomagnetism for smartphone-based indoor localization: Challenges, advances, and comparisons. *ACM Comput. Surv.*, 50(6), dec 2017.

[14] Wenyao Ho, A. Smailagic, D.P. Siewiorek, and C. Faloutsos. An adaptive two-phase approach to wifi location sensing. In *Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOMW'06)*, pages 5 pp.–456, 2006.

[15] Jie Lian, Jiadong Lou, Li Chen, and Xu Yuan. Echospot: Spotting your locations via acoustic sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(3), sep 2021.

[16] Hsiuping Lin, Ying Zhang, Martin Griss, and Ilya Landa. Wasp: An enhanced indoor locationing algorithm for a congested wi-fi environment. In Richard Fuller and Xenofon D. Koutsoukos, editors, *Mobile Entity Localization and Tracking in GPS-less Environnments*, pages 183–196, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[17] Spyros Makridakis and Robert L. Winkler. Averages of forecasts: Some empirical results. *Management Science*, 29(9):987–996, 1983.

[18] The Markup. There's a multibillion-dollar market for your phone's location data. Online, September 2021.

[19] Francisco Morgado, Pedro Martins, and Filipe Caldeira. Beacons positioning detection, a novel approach. *Procedia Computer Science*, 151:23–30, 2019. The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019) / Affiliated Workshops.

[20] Pavitha Nooji and Shounak Sugave. A novel multi-stage stacked ensemble classifier using heterogeneous base learners. *International Journal of Engineering Trends and Technology*, 71:65–71, 04 2023.

[21] ORECX. Call recording laws around the world. Online, 2018.

[22] Manos Orfanos, Harris Perakis, Vassilis Gikas, Günther Retscher, Thanassis Mpimis, Ioanna Spyropoulou, and Vasileia Papathanasopoulou. Testing and evaluation of wi-fi rtt ranging technology for personal mobility applications. *Sensors*, 23(5), 2023.

[23] Mirco Rossi, Julia Seiter, Oliver Amft, Seraina Buchmeier, and Gerhard Tröster. Roomsense: An indoor positioning system for smartphones using active sound probing. In *Proceedings of the 4th Augmented Human International Conference*, AH '13, page 89–95, New York, NY, USA, 2013. Association for Computing Machinery.

[24] Qun Song, Chaojie Gu, and Rui Tan. Deep room recognition using inaudible echos, 2018.

[25] Peter A Dinda Stephen P Tarzia and et al. Indoor localization without infrastructure using the acoustic background spectrum, 2011.

[26] Hao Xia, Xiaogang Wang, Yanyou Qiao, Jun Jian, and Yuanfei Chang. Using multiple barometers to detect the floor location of smart phones with built-in barometric sensors for indoor positioning. *Sensors*, 15:7857–7877, 04 2015.

[27] Moustafa Youssef and Ashok Agrawala. The horus wlan location determination system. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*, MobiSys '05, page 205–218, New York, NY, USA, 2005. Association for Computing Machinery.

[28] Buğra Şimşek and Osman Nuri Güneş. Indoor floor change detection using barometer. In *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2019.

[29] Maroš Šeleng. Wifi round trip time for indoor navigation [online]. Master's thesis, Masaryk University, Faculty of InformaticsBrno, 2019 [cit. 2023-06-05]. SUPERVISOR: PhD Bruno Rossi.