# Indoor Position Estimation Dataset Information Sheet

## GENERAL INFORMATION

### Title of Dataset

Indoor Position Estimation Using Machine Learning

### Author Information

Names: Jaspreet Kaur, Mahmoud A. Shawky, Michael S Mollel, Olaoluwa R Popoola, Muhammad Ali Imran, Qammer H Abbasi, and Hasan T Abbas

Institution: University of Glasgow

Address: James Watt School of Engineering, College of Science and Engineering, University of Glasgow, Glasgow, G12 8QQ

Emails:

[2585552k@student.gla.ac.uk](mailto:2585552k@student.gla.ac.uk), [2569186s@student.gla.ac.uk](mailto:2569186s@student.gla.ac.uk), [micheal.mollel@glasgow.ac.uk](mailto:micheal.mollel@glasgow.ac.uk), [Olaoluwa.Popoola@glasgow.ac.uk](mailto:Olaoluwa.Popoola@glasgow.ac.uk), [muhammad.imran@glasgow.ac.uk](mailto:muhammad.imran@glasgow.ac.uk), [qammer.abbasi@glasgow.ac.uk](mailto:qammer.abbasi@glasgow.ac.uk), [Hasan.Abbas@glasgow.ac.uk](mailto:Hasan.Abbas@glasgow.ac.uk)

### Date of data collection:

The data was collected in December 2022.

### Geographic location of data collection:

Mast Lab, James Watt South Building, Glasgow, G12 8QQ, United Kingdom

### Information about funding sources that supported the collection of the data:

NA

## SHARING/ACCESS INFORMATION

### Licenses/restrictions placed on the data:

NA

### Links to other publicly accessible locations of the data:

NA

### Was data derived from another source?

No

Recommended citation for this dataset: J. Kaur et al., "AI-enabled CSI fingerprinting for indoor localisation towards context-aware networking in 6G," 2023 IEEE Wireless Communications and Networking Conference (WCNC), Glasgow, United Kingdom, 2023, pp. 1-5, doi: 10.1109/WCNC55385.2023.10118652.

## DATA & FILE OVERVIEW

**Details of Data Folders and Files:**

Version: 1.0

Release Date: 1st March 2024

The dataset comprises data recorded from 9 positions within an indoor environment, under both Line of Sight (LOS) and Non-Line of Sight (NLOS) conditions, marked in a square grid and divided into blocks of varying sizes (1 meter, 0.75 meters, and 0.5 meters). Data is organised into 4 text files for each of the 9 positions, capturing amplitude and phase values from 2 receiving antennas, resulting in a total of 36 files. Each file contains 125 x 500 samples, translating to Channel State Information (CSI) matrices combining amplitude and phase information for both channels.

**Folder Structure and Content:**

The dataset is structured under the following path in the Windows file system:

~\USRPs\los\... and ~\USRPs\nlos\...

Under this directory, data is further categorized by LOS conditions and the distance between the transmitter and receiver:

los\0.5 meter distance\P1\ - This directory contains data collected under Line of Sight conditions with a 0.5-meter distance between the transmitter and receiver for position 1 (P1). Similar structures are available for positions P2 to P9 and for different distances as applicable.

Inside each position folder (e.g., P1), you will find:

amplitude\_channel\_[1-2].txt - Text files containing amplitude data from the two receiving antennas.

phase\_channel\_[1-2].txt - Text files containing phase data from the two receiving antennas.

Additional folders for other distances (0.75-meter distance, 1-meter distance) and NLOS conditions follow a similar structure, adjusted for the specific condition and distance.

A screenshot of a computer

Description automatically generated

Figure 1: Data folder Structure

A screenshot of a computer

Description automatically generated

Figure 2: Data folder Structure

A screenshot of a computer

Description automatically generated

Figure 3: Data folder Structure

A screenshot of a computer

Description automatically generated

Figure 4: Data folder Structure

**Data Format Description:**

The raw data files are in text format with amplitude and phase values. The pre-processed CSV files structure the data into a matrix format, where each row represents a sample with its corresponding amplitude and phase values for both channels:

CSI = [[amplitude\_channel1, amplitude\_channel2], [phase\_channel1, phase\_channel2]]

**METHODOLOGICAL INFORMATION**

**Description of Methods Used for Collection/Generation of Data:**

Data was collected in an indoor lab setup with a specific floor plan divided into 9 blocks, under both LOS and NLOS conditions. A transmitting antenna was placed at predetermined positions (P1 to P9), and two receiving antennas were fixed at a distance of 0.035 meters apart. LabView 2021 was utilised for signal processing, including OFDM signal generation, 4-QAM modulation, and Least Square-based channel estimation. Data was recorded for the channel response of each receiving antenna and processed to obtain amplitude and phase information, as depicted in Figures 5 and 6.

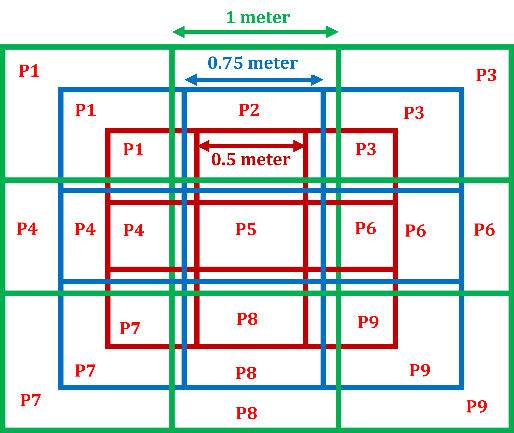
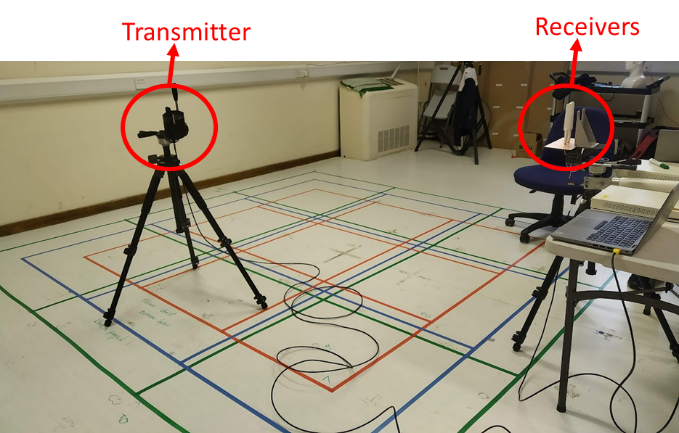
 

Figure 5: Setup

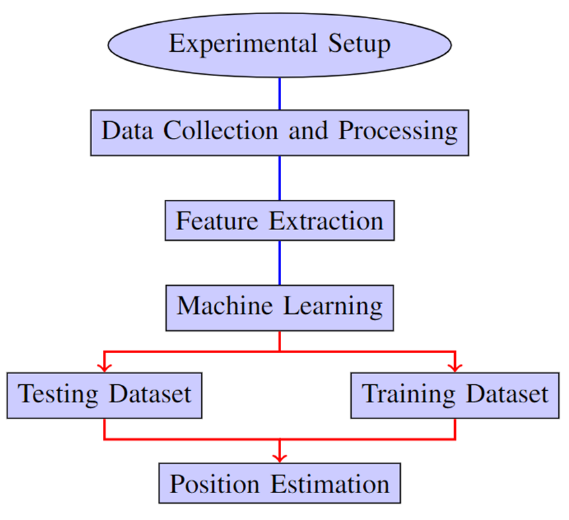


Figure 6: Flow chart of Process

**Methods for Processing the Data:**

Raw data was initially saved in text format, capturing amplitude and phase information under both LOS and NLOS conditions. This data was then translated into CSV format, organising the information into a structured form suitable for machine learning applications. Redundant values were eliminated during preprocessing.

Table 1: Parameter and Definition

|  |  |
| --- | --- |
| Parameter | Definition |
| Frequency | 3.75 GHz |
| Base Station Antenna | 5G/4G Terminal Mount Monopole Antenna |
| Base Station Height | 1.5 m |
| Receiver Height | 1.5 m |
| Receiver Antenna | 5G/4G Terminal Mount Monopole Antenna |
| Receiver-to-receiver distance | varying |
| Computer specs and software | an Intel(R) Core  (TM) i7-9700 CPU @ 3.00GHz, 32GB RAM PC, Wolfram Mathematica 13.2 and LabVIEW 2021 |
| Sampling Rate | 1/3.75 GHz, 125 OFDM data samples |

**Machine Learning Models Used:**

Several machine learning models were evaluated for their effectiveness in position estimation under LOS and NLOS conditions, including Decision Tree, SVM, Neural Networks, gradient-boosted trees, Logistic Regression, Nearest Neighbors, Random Forest, Class Distribution, Naive Bayes, and Markov models. SVM and Neural Networks were identified as providing superior location estimation accuracy, as demonstrated in Figure 7.

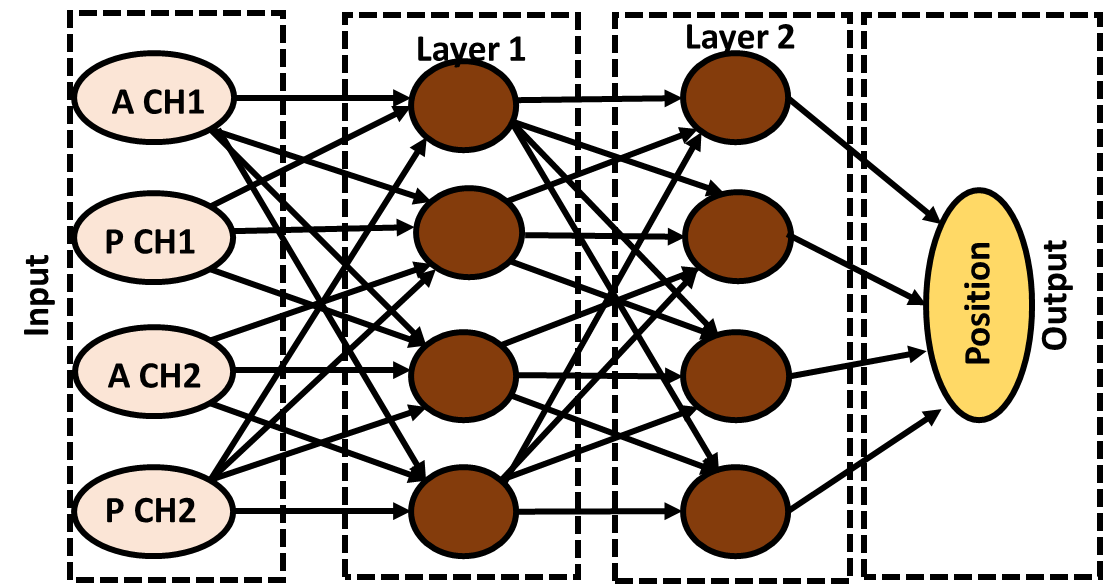


Figure 7: NN Model

**CONTRIBUTION AND USAGE**

This dataset is intended for research purposes, specifically for the development and testing of machine-learning models aimed at indoor position estimation based on channel response data under both LOS and NLOS conditions. Researchers are encouraged to use this dataset to compare the effectiveness of various machine learning algorithms in accurately predicting positions within indoor environments under varying conditions.

For any inquiries or feedback regarding the dataset, please contact:

Jaspreet Kaur

Ph.D. Scholar

James Watt School of Engineering

[2585552k@student.gla.ac.uk](mailto:2585552k@student.gla.ac.uk), [j.kaur.1@research.gla.ac.uk](mailto:j.kaur.1@research.gla.ac.uk)