* Final Project Proposal – see <https://edstem.org/us/courses/30808/discussion/2497206> for requirements
* Canvas Group Project page: <https://gatech.instructure.com/courses/294044/assignments/1225504>
* Example final project proposal (scroll about halfway down): <https://edstem.org/us/courses/30808/discussion/2497206>
  + State
    - Team – list each team member
    - Topic
    - Project title
    - Project summary (motivation)
    - Related work (literature)
      * List of papers we think we should mention:
        + <https://ece.northeastern.edu/fac-ece/ioannidis/static/pdf/2020/J_Jian_RFDeepLearning_IoT_2020.pdf>
        + Each come with 2 papers of SOTA in RF Fingerprinting
    - What we will do: approach 4-5+ sentences
    - Datasets: provide link to dataset
      * https://genesys-lab.org/oracle

-----------------------------------------------------------------------------------------------------------------------------------------

Project Title:

*Deep Learning for Identification of RF Emitters based on Transmitter Artifacts*

Due date: March 17th, 2023, 8:00 AM ET

Team***Electromagnetics!*** *(*3 members)

|  |  |
| --- | --- |
| Name | GATECH Email |
| Greg W Zdor | [Gzdor3@gatech.edu](mailto:Gzdor3@gatech.edu) |
| Jordan W Barker | [Jbarker63@gatech.edu](mailto:Jbarker63@gatech.edu) |
| Seab McCarty | [smccarty30@gatech.edu](mailto:smccarty30@gatech.edu) |

**Section 1.0 Project Summary:** [@Mccarty, Sean P](mailto:smccarty30@gatech.edu)

MITRE Corporation’s 2021 white paper “6G and Artificial Intelligence & Machine Learning” <https://www.mitre.org/sites/default/files/2021-11/pr-21-0214-6g-and-artificial-intelligence-and-machine-learning.pdf> authored by Curtis Wilson, et. all, mentioned that one of the International Telecommunication Union (ITU)’s “Key Performance Indicators (KPI)” for 5G wireless systems is the ability to “serve up to one million” Internet of Things (IoT) “devices per square kilometer”. Another key enabler of “massive IoT” the authors of this white paper highlight is “low-complexity” and “low-power consumption devices”. However, a 2019 IEEE journal article “IoT Device Security Using RF Fingerprinting” by Douae Nuoichi, et. all., <https://ieeexplore.ieee.org/document/8714205> cautions that “the computational complexity of cryptographic protocols and scalability problems make almost all cryptography-based authentication protocols impractical for IoT” [https://ieeexplore.ieee.org/document/8714205](https://ieeexplore.ieee.org/document/8714205 .) For example, “Ghost-in-ZigBee: Energy Depletion Attack on ZIgBee-Based Wireless Networks” discusses the ramifications of an adversarial attack that aims to reduce a ZigBee device’s battery life via “luring a node to do superfluous security-related computations”.

Therefore, the need to enable the secure operation of large scale IoT device networks has motivated the development of RF fingerprinting (RFF) solutions. RFF is defined as the identification (classification) of a transmitter (TX) based off its emitted signal, and properties unique to that given transmitter and transmission, for example TX hardware artifacts. As an example of RFF in industry, an Association of Computing Machinery (ACM) WiseML 22’ Conference Spotlight Session “Systems View to Designing RF Fingerprinting for Real-World Operations” by Scott Kuzdeba, et. all., mentions that RF fingerprinting research has traditionally focused on “scaling up to large population sizes (e.g., 10,000 emitter populations)” <https://dl.acm.org/doi/10.1145/3522783.3529520>. Also, Captain Benjamin W. Ramsey, USAF, states in the introduction of his Air Force Institute of Technology (AFIT) PhD thesis “Improved Wireless Security Through Physical Layer Protocol Manipulation and Radio Frequency Fingerprinting” <https://scholar.afit.edu/cgi/viewcontent.cgi?article=1543&context=etd> that “low-cost” devices can achieve “>90% authentication accuracy” via “RF fingerprinting”.

The goal of our project is to compare the performance of three deep learning networks based on prior research. We selected the Northeastern University Institute for the Wireless Internet of Things “ORACLE: Optimized Radio Classification through Convolutional Neural Networks” dataset for this project <https://genesys-lab.org/oracle.> Our rationale for this decision was to appropriately scope the difficulty of our semester project given the need to balance our time between preparing for the upcoming quizzes and completing assignment #4 and our semester project.

**Section 1.1 Related Work:** [@Mccarty, Sean P](mailto:smccarty30@gatech.edu)

Transmitted digital signals have a particular unique property, they are inherently complex-valued or in-phase quadrature (IQ). This property, coupled with inherent propagation properties inherent to electromagnetic waves, has led to the niche field of research of radio frequency (RF) machine learning (ML), and in this section relevant works in RFF specifically will be elucidated.

“A Comprehensive Survey on Radio Frequency (RF) Fingerprinting: Traditional Approaches, DeepLearning, and Open Challenges” <https://arxiv.org/pdf/2201.00680v1.pdf> by Anu Jagannath, Jithin Jagannath, & Prem Sag is a good introduction to RF Fingerprinting. The authors of this journal article identified three avenues that researchers have explored to solve this technical problem. First, they summarize “traditional approaches” as the combination of engineered features and machine learning algorithms (e.g., SVM & Random Forest). Second, researchers have evaluated the combination of manual feature engineering and neural networks. A limitation of this approach is that it does not take advantage of a neural network’s ability to learn a data representation that yields optimal algorithm performance. Third. Researchers have explored applying deep learning to RF signal data. This includes evaluating the performance of a 1-D Convolutional Neural Network (CNN)’s RF fingerprinting performance given unprocessed complex RF signal (i.e., I/Q) data. Current state of the art (SOTA) performance in RFF indicates at least for the dataset of consideration in this work, a 99% classification accuracy may be achieved per Sankhe et al. (page 8 <https://arxiv.org/pdf/1812.01124.pdf)> , when accounting for electromagnetic propagation specific augmentations and hardware augmentation in the training set. However, not accounting for these domain specific variations Sankhe et al. found performance to drop to 35% accuracy, demonstrating the importance of “careful impairment allocation” (page 8 <https://arxiv.org/pdf/1812.01124.pdf>).

**Section 1.2 Technical Approach:** [@Barker, Jordan](mailto:jbarker63@gatech.edu)

This work will analyze the effectiveness of using Convolutional Neural Networks (CNNs) to learn and identify different radios. Our team will research the potential of modern neural network architectures and experiment with multiple hyperparameter configurations to find the most effective neural network for radio frequency fingerprinting. In addition to exploring different CNN-based architectures and hyperparameter configurations, we will also investigate the impact of data augmentation techniques, such as random cropping, flipping, and rotation, on the model's performance. Our data will be split into training, validation, and test sets.

Inherent to over-the-air IQ data problems is modeling the electromagnetics propagation properties of the signal, and ensuring the classification model generalizes across wireless communication channels. Therefore, a focus point in this effort will be augmenting the train, validation, and testing data with RF-specific augmentations like applying a Rician fading path channel. Additionally, IQ data, a sequence of complex-valued samples, may be converted into a variety of higher-level feature representations, including the spectrogram, discrete Fourier transform plot, wavelets transform, constellation plot, and more. Each of these forms has its pros and cons, for example a spectrogram captures long time-varying features, for example burst hop pattern, while a constellation representation of IQ captures better IQ impairments such as DC offset and IQ imbalance, features directly relevant in RFF. Exploring which representation of IQ is optimal for RF fingerprinting will be assessed too.

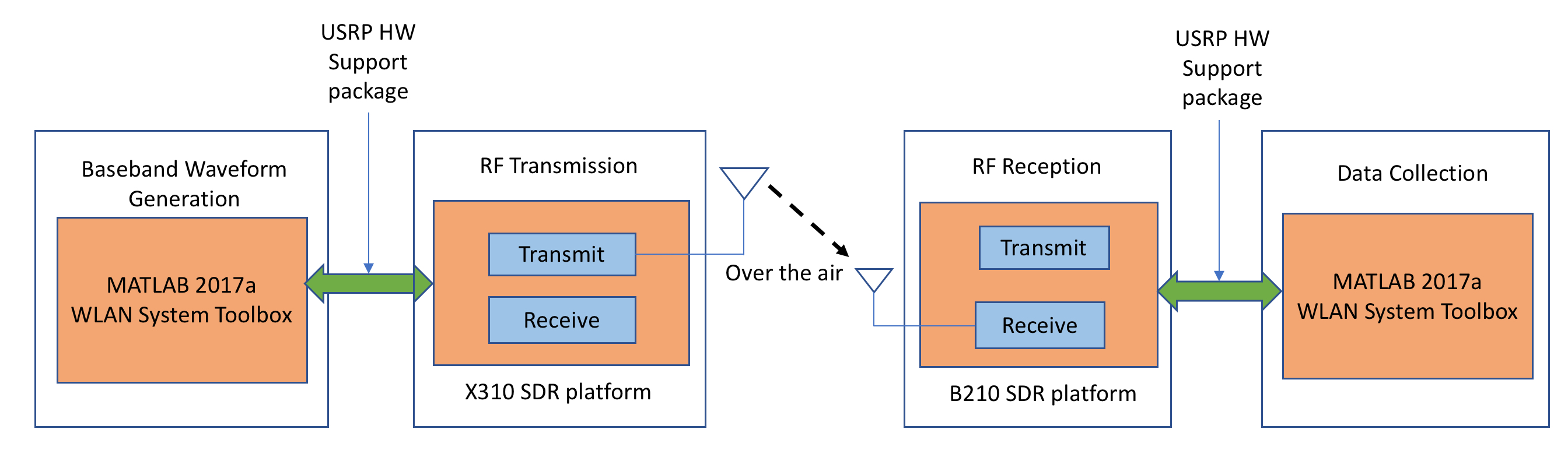
Overall, this technical approach aims to develop an accurate and efficient radio frequency fingerprinting model using deep learning techniques.

**Section 1.3 Dataset:** [@Zdor, Greg W](mailto:gzdor3@gatech.edu)

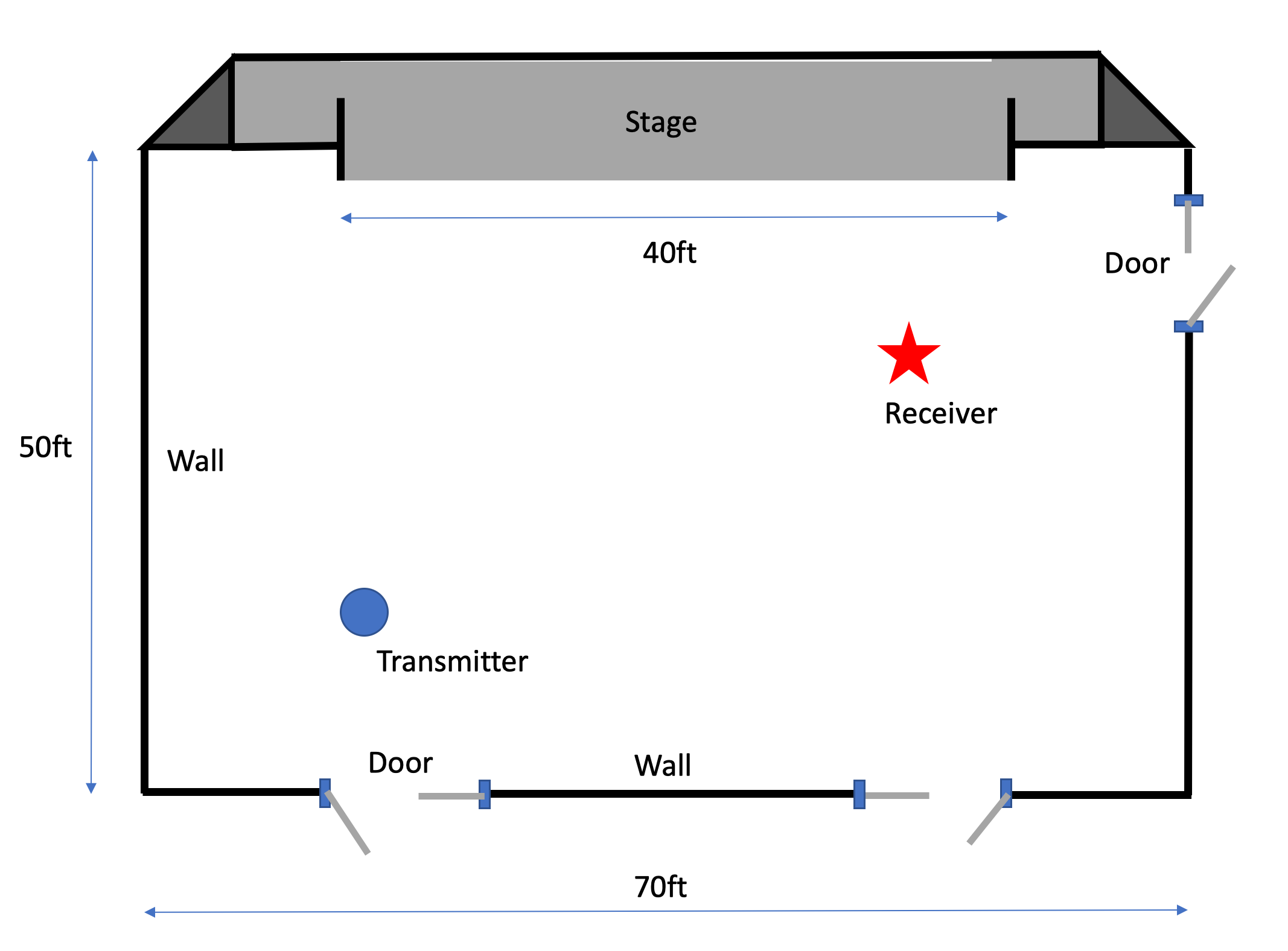
The dataset of use in this work comes from Northeastern University’s GENESYS Laboratory and was released as part of their 2018 IEEE INFOCOM paper [“ORACLE: Optimized Radio clAssification through Convolutional NeuraL nEtworks”](https://arxiv.org/pdf/1812.01124.pdf). This classification dataset consists of in-phase quadrature (IQ) base banded, digitized, raw, over-the-air received samples of Wi-Fi bursts, namely IEEE802.11a protocol. This dataset used 16 different transmitters (hence 16 unique classes) and one B210 receiver. While assessing whether an RFF model has learned a particular receiver or has generalized across receivers is of interest in RFF, this work will focus on strictly on learning transmitter differences, since this dataset does not contain multiple receivers to evaluate generalization across. Refer to following summary for dataset details and download links.

* **Dataset details:** 
  + Dataset producers: <https://genesys-lab.org/>
  + Overview: <https://genesys-lab.org/oracle> *(Download links on this page)*
  + Associated paper link: <https://arxiv.org/pdf/1812.01124.pdf>
  + Data amounts
    - 16 classes (16 different USRP X310 SDRs )
    - 20 million samples / class
    - Sampling rate: 5 MSPS
    - Center frequency: 2.45 GHz
  + Data format:
    - Raw data in binary file format (read as np.complex128)
    - Metadata (labels) in standard [SigMF](https://github.com/sigmf/SigMF/blob/sigmf-v1.x/sigmf-spec.md) format

**Figure 1**. illustrates the dataset generation flow used by the creators of this set, where the WLAN System Toolbox in MATLAB was used to generate the Wi-Fi waveform. Note generated frames contained random payload but had identical address fields ensuring equivalency of data across classes. Image reference: <https://genesys-lab.org/oracle>



**Figure 2**. shows that path distance was varied in increments of 6 feet, from 2 to 62 feet for the transmitter (TX) to receiver (RX) distance. This distance has a direct effect on the received signal strength. Image reference: <https://genesys-lab.org/oracle>



**Section 1.4 References:**   
  
[1] A. Jagannath, J. Jagannath, P. S. Pattanshetty V. Kumar [A Comprehensive Survey on Radio Frequency (RF) Fingerprinting: Traditional Approaches, Deep Learning, and Open Challenges](https://arxiv.org/abs/2201.00680) (2022)

[2] K. Sankhe, M. Belgiovine, F. Zhou, S. Riyaz, S. Ioannidis, and K. Chowdhury, “Oracle: Optimized radio classification through convolutional neural networks,” in IEEE INFOCOM 2019-IEEE Conference on Computer Communications. IEEE, (2019)

[3] A. Hermawan, R. R. Ginanjar, D. Kim, J. Lee, “[CNN-Based Automatic Modulation Classification for Beyond 5G Communications,”](https://www.semanticscholar.org/paper/b8803f1df34dd352376ad25d8d87aae4fd4e1d73)IEEE Communications Letters, (2020)

[4] S. Hanna, S. Karunaratne, D. Cabric, “[Open Set Wireless Transmitter Authorization: Deep Learning Approaches and Dataset Considerations](https://ieeexplore.ieee.org/document/9286564)” IEEE Transactions on Cognitive Communications and Networking (2021)

[5] G. Yin, J. Zhang, G. Shen, Y. Chen “[FewSense, Towards a Scalable and Cross-Domain Wi-Fi Sensing System Using Few-Shot Learning](https://ieeexplore.ieee.org/document/9947336),” IEEE Transactions on Mobile Computing (2022)

[6] G. Shen; J. Zhang, A. Marshall, J. R. Cavallaro “[Towards Scalable and Channel-Robust Radio Frequency Fingerprint Identification for LoRa,](https://ieeexplore.ieee.org/abstract/document/9715147)” IEEE Transactions on Information Forensics and Security (2022)

[7] M. Cekic, S. Gopalakrishnan, U. Madhow “[Wireless Fingerprinting via Deep Learning: The Impact of Confounding Factors](https://ieeexplore.ieee.org/document/9723393),” 55th Asilomar Conference on Signals, Systems, and Computers (2021)