# Data

Our dataset consisted of the *Culex tarsalis* mosquito waveform library from Cooper et. al 2023, which contains 62 recordings. From these recordings, which consist primarily of non-probing behavior, we extracted 94 probes that totaled 11 hours in length. Important to note is the high degree of imbalance in the dataset, with L and M waveforms making up 10 of the 11 hours of available data.

# Evaluation

All models were evaluated using 5-fold cross validation. To create each fold we first divided the recordings into 5 separate folds and then extracted the probes from them to prevent data leakage between probes from the same recording. Reported results are the average of the 5 folds for both accuracy and F1 score.

# Results

**Bolded** numbers refer to the best in their respective category

## Overall

| Model | Accuracy (%) | Macro F1 Score |
| --- | --- | --- |
| Random Forest | 77 | 0.39 |
| TCN | 79 | 0.37 |
| Segmentation Transformer | 74 | 0.39 |
| UNet (block) | **84** | 0.66 |
| **UNet (attention)** | 82 | **0.69** |

## F1 Score by Waveform Type

| Model | J | K | L | M | N | W |
| --- | --- | --- | --- | --- | --- | --- |
| Random Forest | 0.47 | 0.19 | 0.78 | 0.80 | 0.05 | 0.00 |
| TCN | 0.40 | 0.06 | 0.80 | 0.89 | 0.04 | 0.00 |
| Segmentation Transformer | 0.70 | 0.11 | 0.77 | 0.77 | 0.00 | 0.00 |
| UNet (block) | 0.87 | 0.60 | **0.85** | **0.91** | 0.14 | 0.59 |
| UNet (attention) | **0.88** | **0.67** | 0.83 | 0.87 | **0.32** | **0.62** |