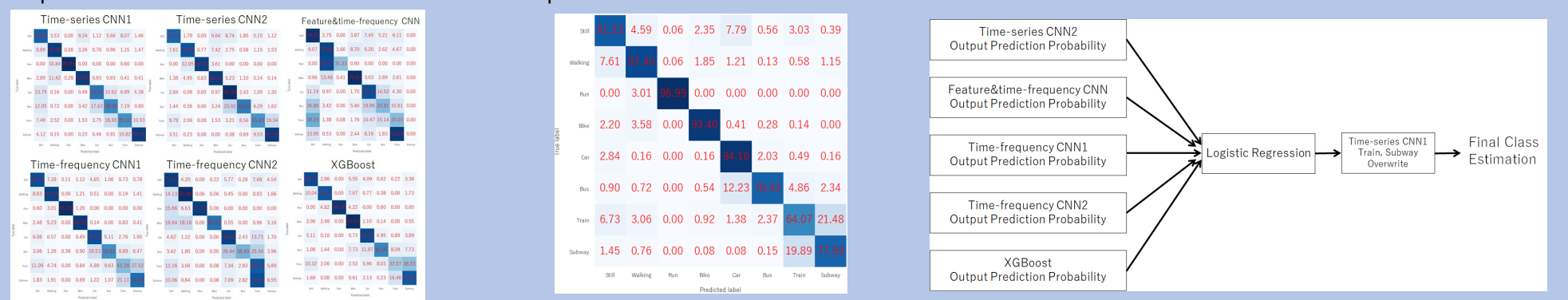


# Ensemble Learning for Human Activity Recognition

Ryoichi Sekiguchi, Kenji Abe, Takumi Yokoyama, Masayasu Kumano, Masaki Kawakatsu  
Tokyo Denki University

The use of ensemble learning, which combines the outputs of multiple classifiers to produce a single estimation result, improved the accuracy of activity recognition. The ensemble model consists of 1:Time-frequency CNN, 2:Feature&time-frequency CNN, 3:XGBoost and 4:Time-series CNN. The phone location of SHL test-set was estimated to be Hips. The best F-measure obtained for last 30% SHL validation-set was 84.8%.



### Preprocess1

Acc, Gyr, and Mag were converted into the world coordinate system (North-East-Down Coordinate) by obtaining the rotation matrix from the orientation.

$$R_{NB} = \begin{bmatrix} 1 - 2(q_y^2 + q_z^2) & 2(q_x q_y - q_w q_z) & 2(q_x q_z - q_w q_y) \\ 2(q_x q_y + q_w q_z) & 1 - 2(q_x^2 + q_z^2) & 2(q_y q_z - q_w q_x) \\ 2(q_x q_z - q_w q_y) & 2(q_y q_z + q_w q_x) & 1 - 2(q_x^2 + q_y^2) \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_N = R_{NB} \begin{bmatrix} x \\ y \\ z \end{bmatrix}_B$$

### Preprocess2

User and holding position of the test data were estimated. The following Time-Frequency CNN was used for estimation.

User Estimation (N=2)

True label \ Predicted label	user2	user3
user2	87.24	12.76
user3	9.62	90.38

Phone Location Estimation (N=4)

True label \ Predicted label	Bag	Hips	Torso	Hand
Bag	89.00	3.91	5.77	1.32
Hips	3.34	88.53	5.06	3.07
Torso	3.58	9.01	83.73	3.68
Hand	3.81	11.04	4.72	80.43

### MODEL1: Time-frequency CNN

The model was constructed using Acc, Mag and Gyr time-frequency spectrums

### MODEL2: Feature&time-frequency CNN

In LAcc, we made values that was sum of continuous two points of all 500 points in each frame for a axis-z and  $\sqrt{x^2 + y^2}$  (Input shape 499x2).  
In Mag, time-frequency spectrums for an axis-z and  $\sqrt{x^2 + y^2}$  time-window two seconds overlap 100 ms (Input shape 131x31x2).

### MODEL3: XGBoost

	LAcc XY	LAcc Z	Gyr Z	Mag Z
Mean, Variance, Skewness and Kurtosis	○	○		
Sum of FFT results every 5Hz		○	○	○
Maximum values of FFT results every 5Hz		○	○	○
Frequencies that takes the maximum value of the FFT result every 5Hz		○	○	○

### MODEL4: Time-series CNN

This model was constructed using Pre and LAcc in the world coordinate system.

### Ensemble Learning

Time-series CNN1 and Time-frequency CNN1 are relearned with SHL Validation-set to become Time-series CNN2 and Time-frequency CNN2. These four models (Time-series CNN2, Feature&time-frequency CNN, Time-frequency CNN1, Time-frequency CNN2, and XGBoost) provide output prediction probabilities to a Logistic Regression model, which then uses Time-series CNN1 for Train, Subway Overwrite to produce the Final Class Estimation.

Logistic regression model with 8 outputs of 5 models as input (Input shape 8x5=40). The estimation class for the logistic regression model has been partially overridden. The frames that the time-series CNN1 estimated to be train and subway were overwritten with it.