# Capacity of Smart Thermostat Data to Identify Thermal Discomfort

Katherine Bassett, Northeastern University

IBUILD GRADUATE RESEARCH FELLOWSHIP

## Occupant Thermostat Overrides Result in Unreliable Load Flexibility

- Increased electrification, extreme weather, and renewable energy penetration accelerate the need for electric load flexibility.
- Current residential thermostat load flexibility programs are unreliable owing to the unpredictability of occupant temperature overrides.
- Identifying when an occupant is uncomfortable can improve the reliability of thermostat override predictions.
- Real-time comfort identification could lead to personal and dynamic load flexibility controls that allow for continuous load control.

## Current Research

- Real-time comfort identification models require input data that are readily available, such as data from smart thermostats.
- Understanding the extent to which this data can identify thermal comfort is required to understand the practicality of this strategy.
- Using data collected in an ongoing DOE-funded study run by Dr. Michael Kane, raw and engineered smart thermostat features were tested for their Thermal Preference Vote (TPV) separability.
- This work explored if the TPV separability of some features would be improved if a *z*-score autonormalization technique was used as opposed to standard *z*-score normalization. The results for each feature can be seen in **Figure 1**.
- Autonormalization is the process of *z*-score normalizing a home's data relative to its own mean and standard deviation rather than against the mean and standard deviation of preaggregated data.

## Results and Implications

- Autonormalization of some features improve the 1D separability of TPV while diminishing the separability of other features.
- Autonormalized Humidex has the overall highest TPV separability and results in the best performing support vector machine (SVM) classifier when it is used as the only training feature (**Figure 2**).
- Although the initial performance of SVMs was fairly low, a significant portion of the limitations could be the result of insufficient feature combinations that can support the separability of TPV classes.
- Future exploration into developing new features and understanding what explanatory dimensions are missing could aid in improved performance of SVM and other TPV classifiers.

**Separability of Thermal Preference Vote Class**

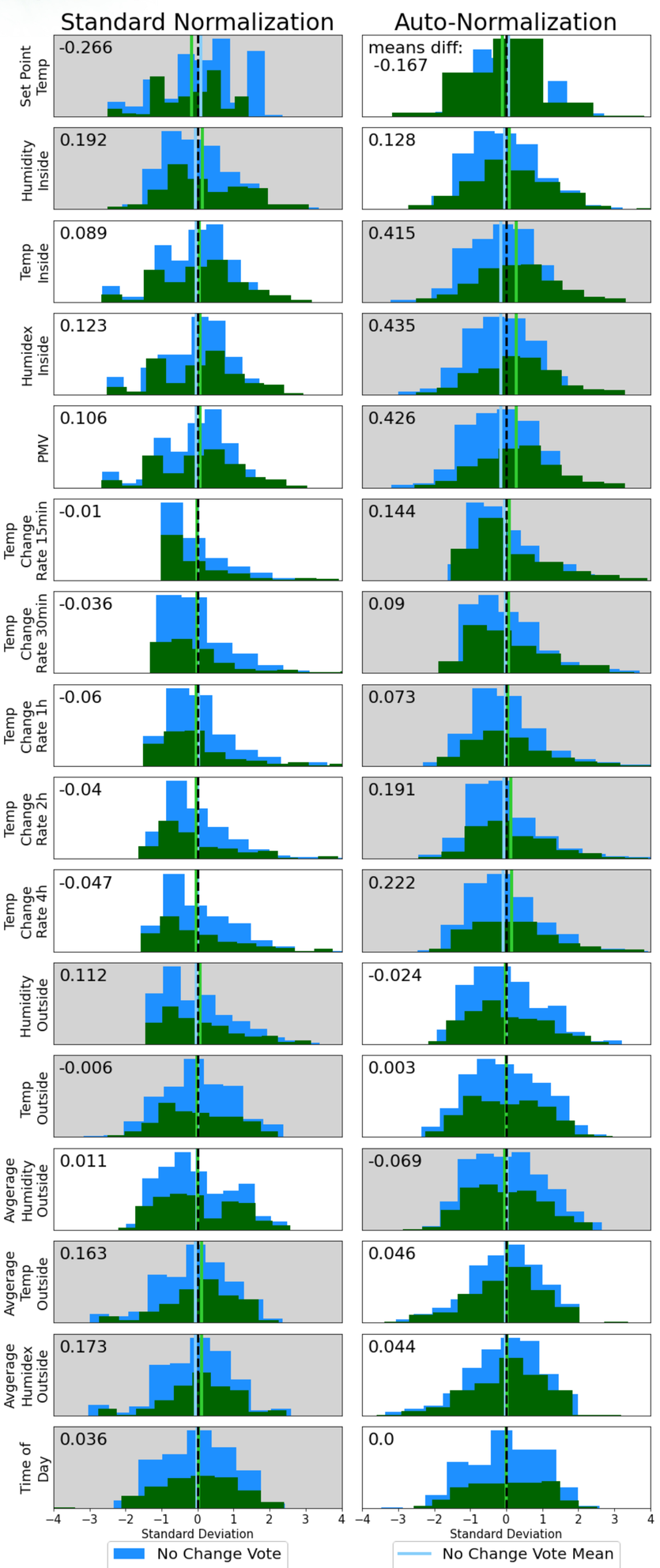


**Figure 1:** Each histogram represents the count of TPVs across a normalized *x*-axis. For each variable, the grey plot is more separable.

**Performance of SVMs Trained with Different Normalization Strategies**

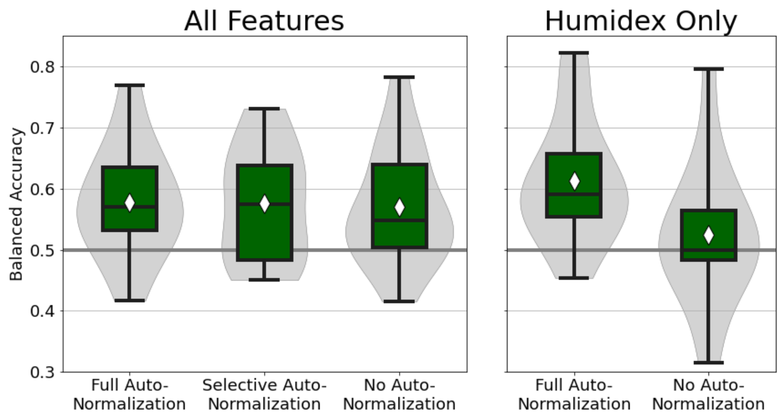


**Figure 2:** SVMs trained only on the autonormalized Humidex feature outperformed any combination of features.