Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh

IEEE
Intelligent Systems IS'20

Comparative Study of Classifiers on Human Activity Recognition by Different Feature Engineering Techniques

Mahbuba Tasmin, Taoseef Ishtiak, Sharif Uddin Ruman,Arif ur Rahman Chowdhury Suhan, N.M. Shihab Islam,Sifat Jahan, Md. Sajid Ahmed, Md. Shahnawaz Zulminan, Abdur Raufus Saleheen, Dr. Rashedur M. Rahman

Contact Email: rashedur.rahman@northsouth.edu

Introduction

Enhanced Recognition Task

Final Result

Faster Computation

Low Resource Dependency

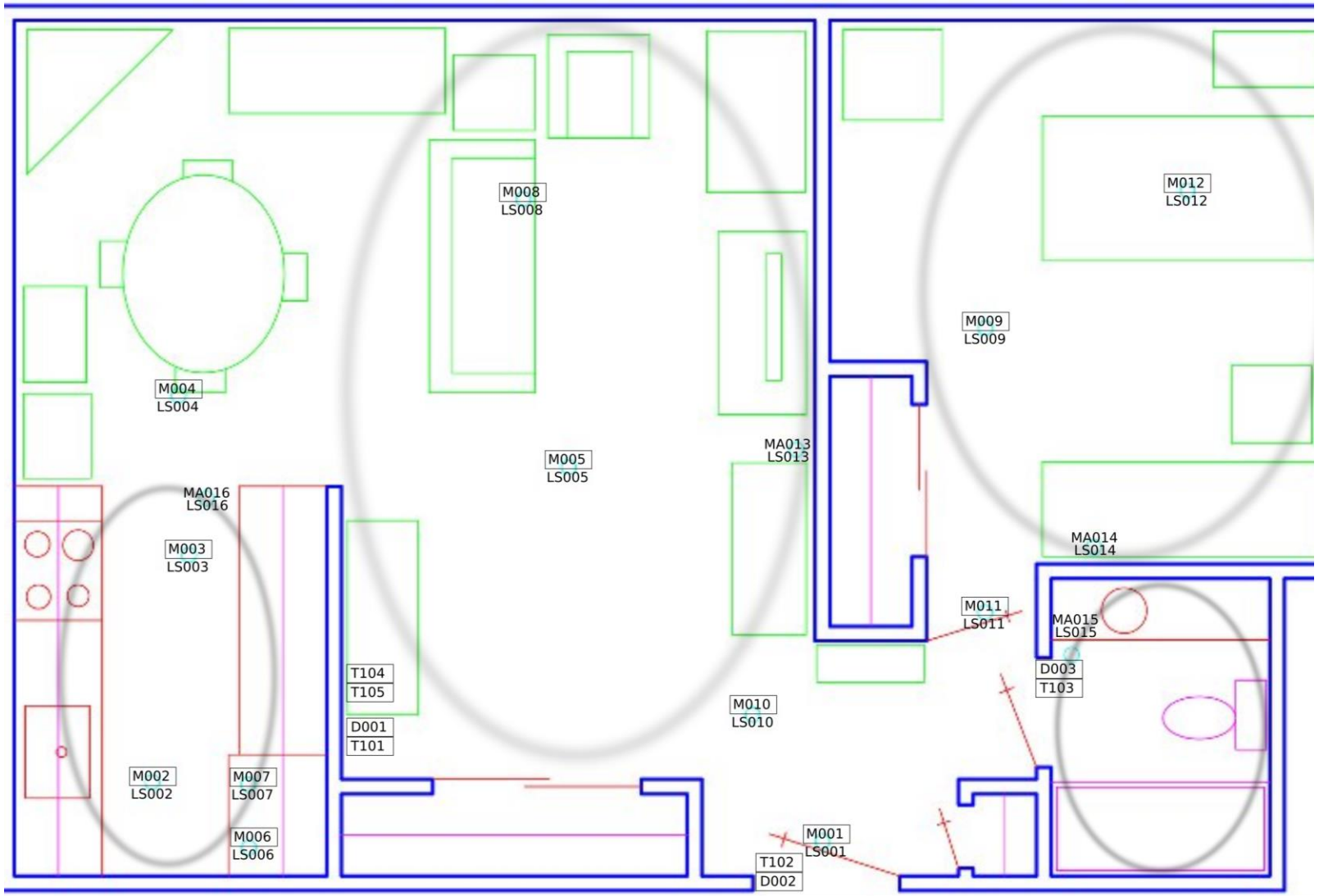
Research Expectation	Presenting the significance of a feature engineering pipeline on large HAR dataset for faster computation with improved accuracy and low-resource dependency infrastructure aimed for activity-recognition tasks through comparative analysis.

RESEARCH CONTRIBUTION

Result Statement	Accuracy score of activity-recognition improves to 78.5% by Multi-layer perceptron model from 76.7% after feature engineering applied on raw dataset by Extra-tree classifier method.
Highlights	• Feature Engineering through Tree-based Selection Methods • Dimensionality Reduction of dataset through PCA • Performance Evaluation of baseline five classifier model

DATASET

Source	Collected from UCI Machine Learning Repository, ***Human Activity Recognition from Continuous Ambient Sensor Dataset*** (Published – September 20,2019)
Number of Activities	12
Number of Attributes	37
Attribute Characteristics	Integer/Real with missing data, 4M instances (approx)
Selected number of activities for this Study	'Cook':0, 'Eat':1, 'Phone':2, 'Read':3, 'Watch_TV':4



DATA COLLECTION SETUP

METHODOLOGY

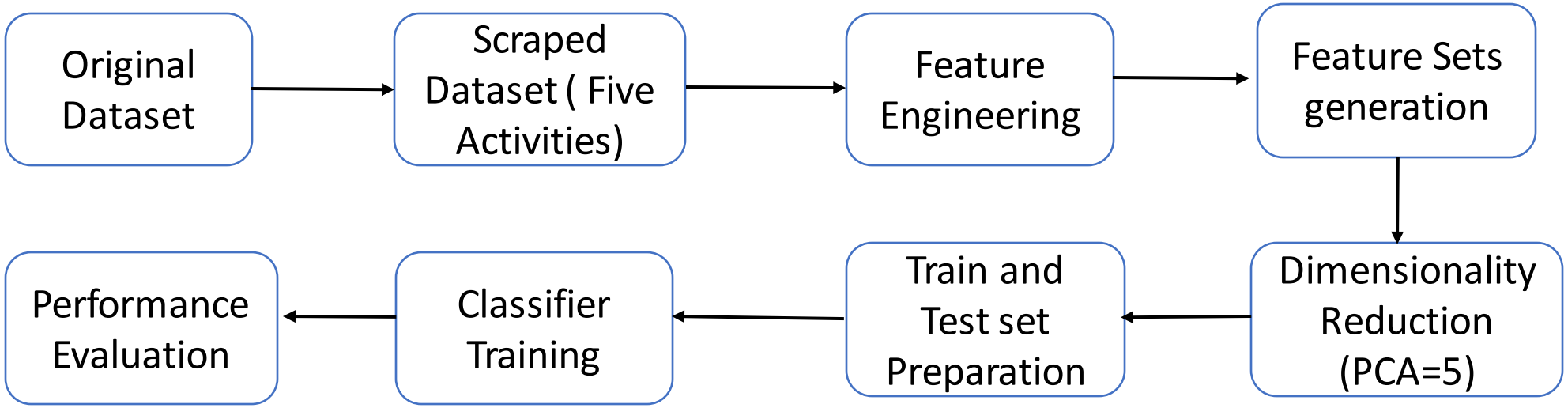


Figure 1: Workflow of the proposed Approach

FEATURE ENGINEERING

- Selects significant features from original 37 attributes.
- Produces four different data-sets from original dataset.

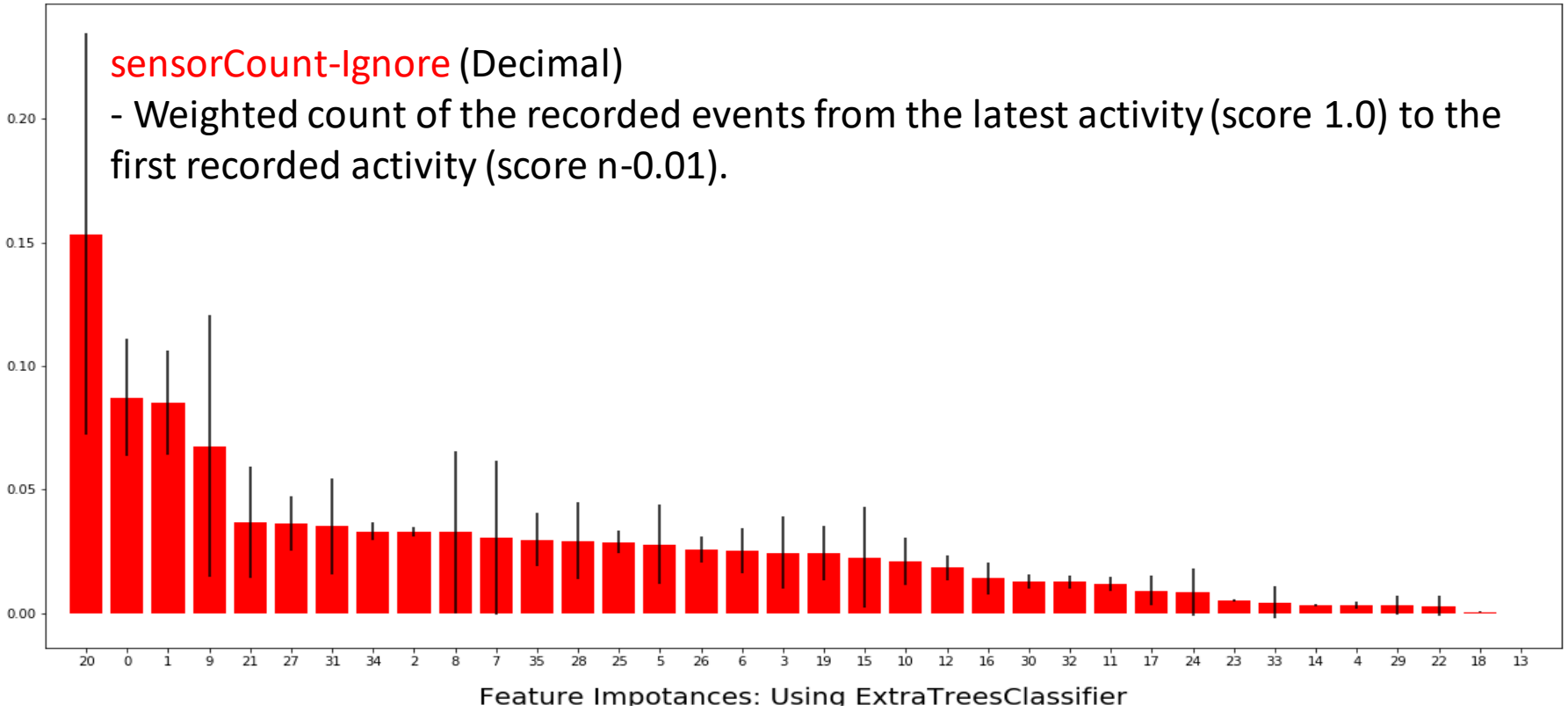
Four variations of Tree-based Feature Selection Approaches		
Variants	Name	
Feature Engineering through ML model	Tree-based Feature Selection	Feature Selection with Random Forest
Feature Rank Generation by Node-impurity Score	Extra-Tree Classifier	Random Forest Classifier

SELECTED FEATURE SETS THROUGH FEATURE-ENGINEERING

Attribute Name	Description
LastSensorEventHours (Integer)	These attributes refer to the last recorded time of activity.
LastSensorEventSeconds (Decimal)	
LastSensorDayOfWeek (Symbolic-value integer)	
windowDuration	It represents the duration of 30 event sliding window in seconds
LastSensorLocation (symbolic-valued integer)	Sensor location ID in the activity window.
LastMotionLocation (symbolic-valued integer)	It refers the location ID of last motion sensor used in the window
sensorCount-Kitchen (decimal)	For most recent event 1.0 is added, and for each previous event (n-0.01) is cumulatively added of the current recorded event. Records the weighted event count of tagged location.
sensorCount-LivingRoom (decimal)	

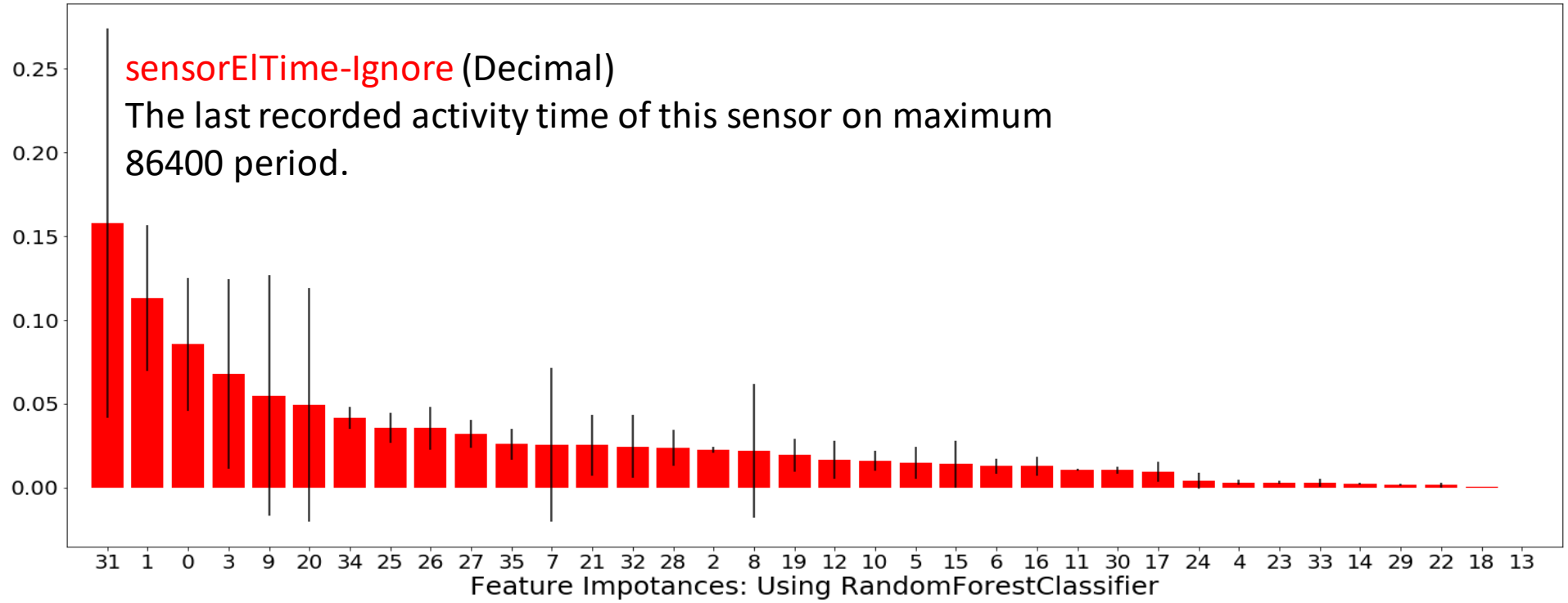
Feature Rank Generation

Extra-Tree Selection



Feature Rank Generation

Random Forest Selection



FINAL FEATURESET GENERATION

- Reduces the number of instances from previously feature-engineered dataset.
- Unsupervised dimensionality reduction of feature set through Principal Component Analysis (PCA=5).
- Projected variance ratio tested against K(K=3) nearest neighbor classifier on the 5-dimensional projected points.

Variance Ratio Evaluation against KNN

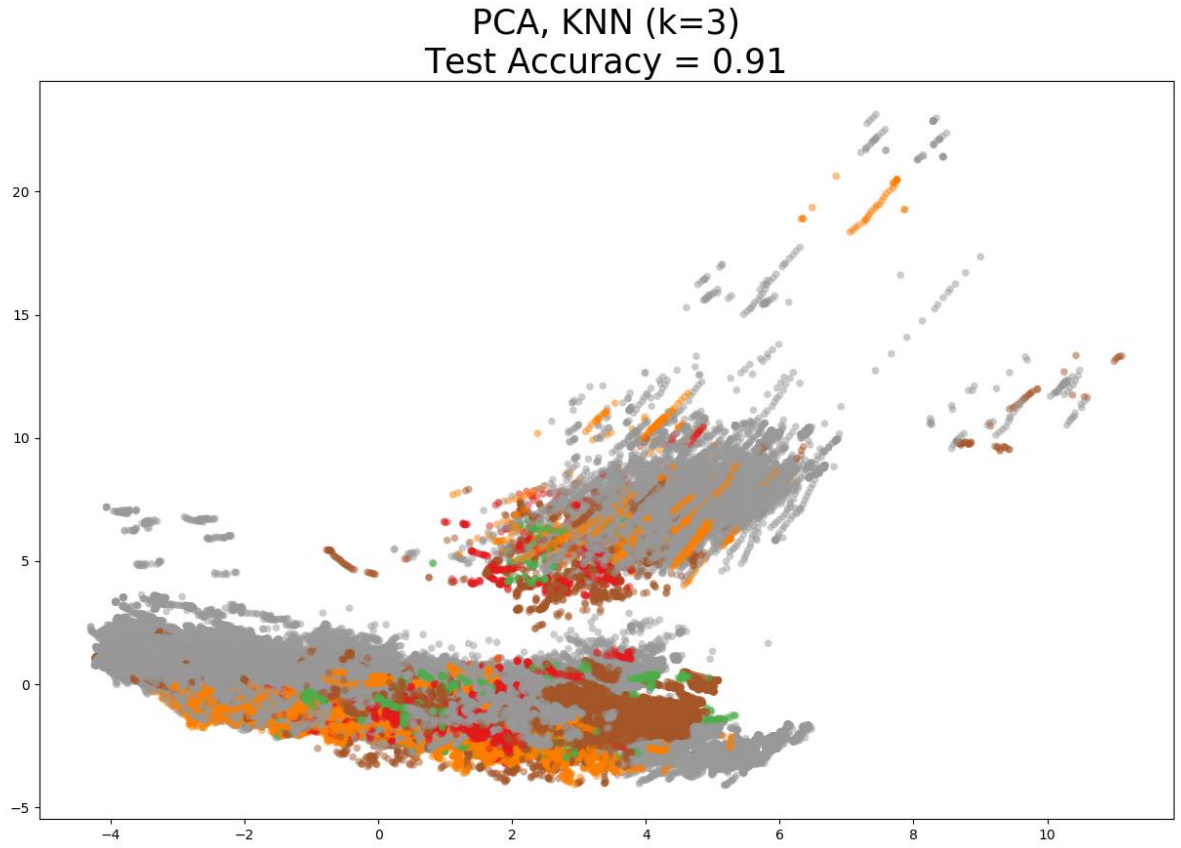


Figure II: Variance Test Accuracy for PCA-5 Dimensionality Reduction

- Reduced subspace of dataset – 5-point projected space.
- KNN verifies the variance ratio.
- 91% variance of data-instances covered.

BASELINE CLASSIFIER MODELS

- K Nearest Neighbors
- Decision Tree
- Random Forest
- Gaussian Naive Bayes
- MLP Classifier using Backpropagation

PERFORMANCE EVALUATION

- Five classifier models evaluated for each five feature sets(feature engineered=4, original=1)
- Feature engineering significance inferred from variance of accuracy score from the evaluation sets.

Classifier	Original Set of 37 Features	Model Obtained Selected Features with Tree-based Feature Selection	Model Obtained Selected Features with Random Forest Classifier	Top 21 Extracted Features by Extra Tree Classifier	Top 21 Extracted Features by Random Forest Classifier
Nearest Neighbor	74.2	75.4	72.9	75.7	75.3
Decision Tree	75.3	76.9	75.9	76.3	75.5
Random Forest	74.4	76.4	75.7	75.9	74.5
Naive Bayes	74.9	76.5	76.4	76.9	76.3
Neural Net	76.7	78.3	77.3	78.5	78.1

TABLE I: COMPARISON OF PERCENTAGE IN ACCURACY SCORES OF CLASSIFIERS ON FEATURE SETS

Recognition Score

Better Accuracy Scores:
Cook, Read, Watch TV

Lower Accuracy Scores:
Eat, Phone

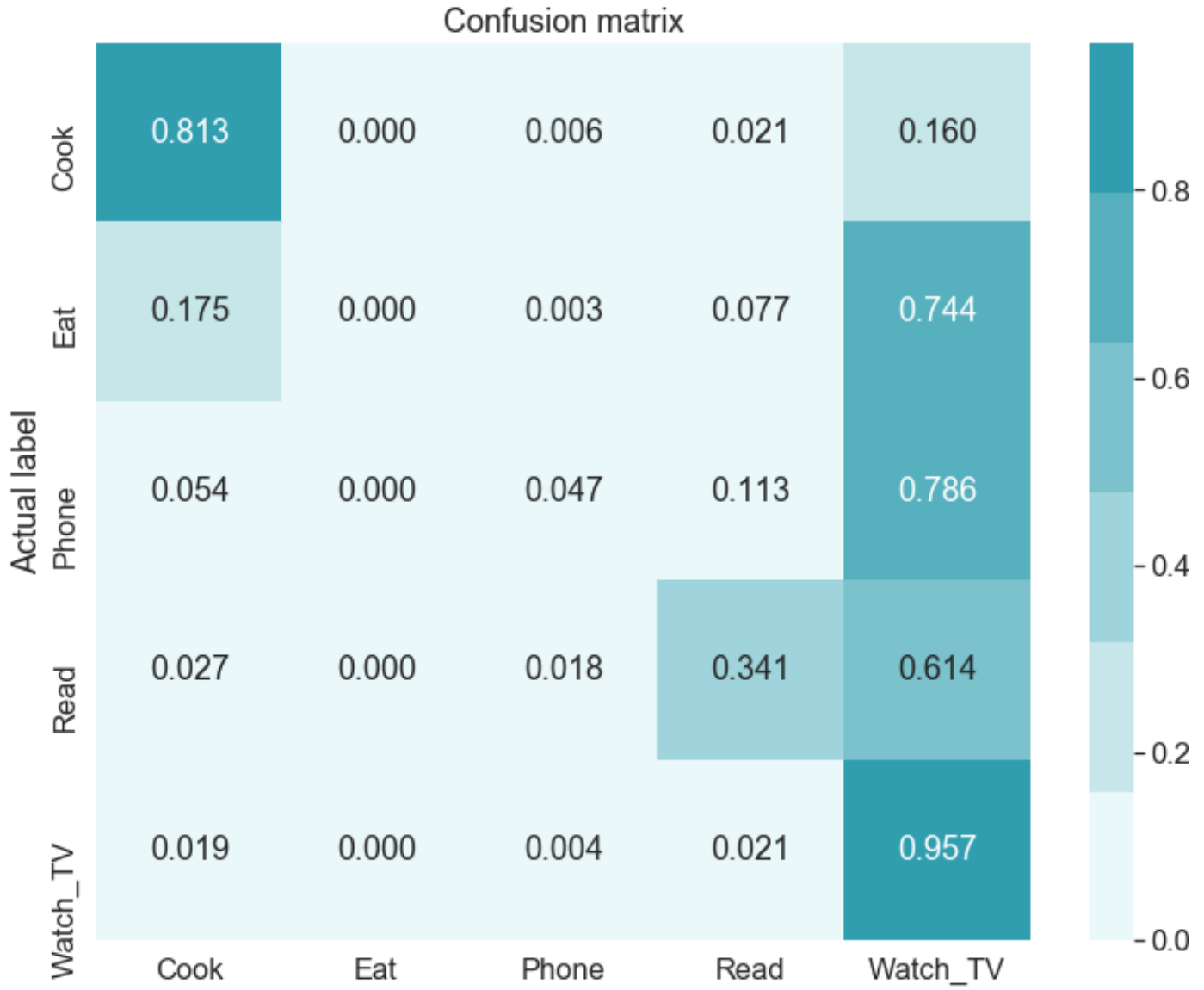


FIGURE III: Evaluation Metric of Neural Network on Extra Tree Classifier Feature set.

DISCUSSION

- The proposed work has suggested how ***tree-based classifier and methods are effective for important feature selection*** to make the training process faster.
- The proposed methodology also validates the fact that model training accuracy can be increased through elimination of redundant features and reduction of noisy instances by ***Dimensionality Reduction***.
- Makes up ***a Computationally inexpensive classification scheme*** based on baseline machine-learning models.

Future Work

Scope of Application	Example
Time-series Data	Faster data-analysis and low-cost structural support for ML model analysis.
Activity Surveillance	Ambient Assisted Living
	Health Sector (Isolation/Quarantine)
	Indoor Human-Activity Recognition

Thank you!

Questions?