Summary of ”Device-Free People Counting in IoT Environments: New Insights, Results, and Open Challenges”

MohammadReza AhmadiTeshnizi
Amir Marvasti Nejad

December 21, 2023

Focus of the Study

The paper titled ***"Device-Free People Counting in IoT Environments: New Insights, Results, and Open Challenges"*** focuses on the development of Internet of Things (**IoT**) solutions for detecting, tracking, counting, and identifying human activity without requiring individuals to carry any device or participate actively in the detection process. The ubiquity of wireless technologies, especially in **IoT** environments, has led to significant research and development in non-invasive detection of human movements through the analysis of radio-frequency signal propagation effects.

Generalities and Thematic Literature

The article offers an in-depth look at the evolution and current methods of device-free people counting in **IoT** environments. It discusses key concepts, historical developments, and various machine learning techniques used for counting people without personal devices, emphasizing advancements in feature engineering and model selection. This review provides insights into the field's progression and its state-of-the-art technologies.

Main Challenge

The main challenge in **"Device-Free People Counting in IoT Environments"** is enhancing the accuracy and efficiency of counting people in **IoT** spaces without personal devices. The article introduces a novel solution using machine learning to analyze Wi-Fi signal changes due to human presence, aiming to improve people counting in indoor settings. This research is pivotal for smart environment applications, focusing on the effectiveness of various machine learning models and feature engineering techniques for more precise and reliable counting methods.

Suggested Solution

The solution proposed in the article uses machine learning to analyze Wi-Fi signal changes for counting people in **IoT** environments, without requiring personal devices. It effectively counts people indoors by detecting signal fluctuations caused by human presence. This innovative approach leverages existing Wi-Fi networks and machine learning, offering an efficient, non-intrusive way to monitor occupancy in smart environments without additional sensors or devices.

Previous Methods

A comprehensive review of the previous methods is given in **Table I**, **Table II**, and **Table V**.

Research Niches

While most studies focus on new predictors for people counting, they often overlook the variety of machine learning models used. It suggests a broader exploration of these models and stresses the importance of understanding interactions between features. The section also highlights the need for effective feature selection techniques to evaluate the collective predictive power of new and existing features.

Simulator Used

The article doesn't specify a particular simulator for its machine learning model in people counting. It focuses more on applying these models to Wi-Fi signals rather than on simulation tools, suggesting reliance on real-world data or custom setups instead of a named simulation software.

Key Highlights

- **Channel State Information (CSI) in IoT:** The paper discusses the use of WiFi radio receivers as sensors in IoT environments for device-free human activity recognition. It emphasizes the importance of CSI, which captures the shape of the received signal or channel response in the frequency domain, making it suitable for activity recognition and crowd counting due to its sensitivity to multi-path propagation.
- **Feature Engineering and Machine Learning Models:** The paper elaborates on feature engineering and selection methodologies for counting device-free people. It reviews various feature engineering and machine learning (ML) models used in the literature, emphasizing crowd counting and tracking problems. The researchers highlight the lack of a universally outperforming set of features for all scenarios, suggesting the importance of selecting appropriate feature combinations for different environments.
- **Types of Features for People Counting:** The paper reviews different types of features used for device-free people counting, including time correlation, time statistics, spectral analyses, and probability distribution functions. It highlights the use of various statistical metrics and methods like Pearson correlation matrices, Principal Component Analysis (PCA), the coefficient of variation, and the Earth Mover's Distance (EMD).
- **ML Models for Device-Free People Counting:** The paper discusses a variety of ML models employed for people counting, including Gaussian Naive Bayes, Adaptive Boosting, k-Nearest Neighbors, Decision Trees, Extreme Learning Machine, Linear Discriminant Analysis, Logistic Regression, and Random Forest. The study uses these models in a nested cross-validation methodology to assess their performance in different scenarios.

- **Influence of Scenario Characteristics on Model Performance:** The study confirms that the physical characteristics of the environment significantly influence the performance of the model. It shows that feature selection (FS) techniques generally improve the model's performance, but the extent of this improvement depends on the specific characteristics of the scenario and the model used.
- **IoT Framework Design and Cyber-Physical Architecture:** The paper delves into the design of a device-free people counting IoT framework, which can be regarded as a cyber-physical system. It discusses various levels of such a system, including smart connection, data-to-information conversion, cyber level, cognition level, and configuration level.

Summary Table

Aspect	Details and Numbers
Focus of the Study	Development of IoT solutions for detecting, tracking, counting, and identifying human activity without active participation of individuals.
Key Technology	WiFi radio receivers used as sensors in IoT for device-free human activity recognition.
Important Feature	Channel State Information (CSI) - captures signal shape/response in frequency domain.
Feature Engineering	Various features and ML models reviewed. No universally outperforming set of features identified for all scenarios.
Types of Features for People Counting	Time correlation, time statistics, spectral analyses, probability distribution functions (e.g., Pearson correlation matrices, PCA, EMD).
Machine Learning Models Reviewed	Gaussian Naive Bayes, Adaptive Boosting, k-Nearest Neighbors, Decision Trees, Extreme Learning Machine, Linear Discriminant Analysis, Logistic Regression, Random Forest.
Cross-Validation Methodology	Nested cross-validation used for assessing ML model performance.
Influence of Scenario Characteristics	Physical characteristics of the environment significantly impact model performance. FS techniques generally improve performance.
IoT Framework Design Levels	Smart connection, data-to-information conversion, cyber level, cognition level, configuration level.
Dataset Used for Analysis	EHUCOUNT dataset from the Faculty of Engineering of the University of the Basque Country, capturing WiFi signal measurements in different indoor scenarios.
Measurement Campaign	WiFi signals transmitted and received using Anritsu equipment, with full CSI from 53 subcarriers of the OFDM signal.
Feature Selection (FS) Techniques	RF model built over the training dataset with all features included. Features with importance greater than the median are retained.
Performance Metrics Used	Accuracy and Kappa scores computed over 20 repetitions of the nested cross-validation method.
Significant Findings	Best Kappa scores above 0.4 indicating fairly good agreement between true and predicted number of people. Variability in scores among scenarios.

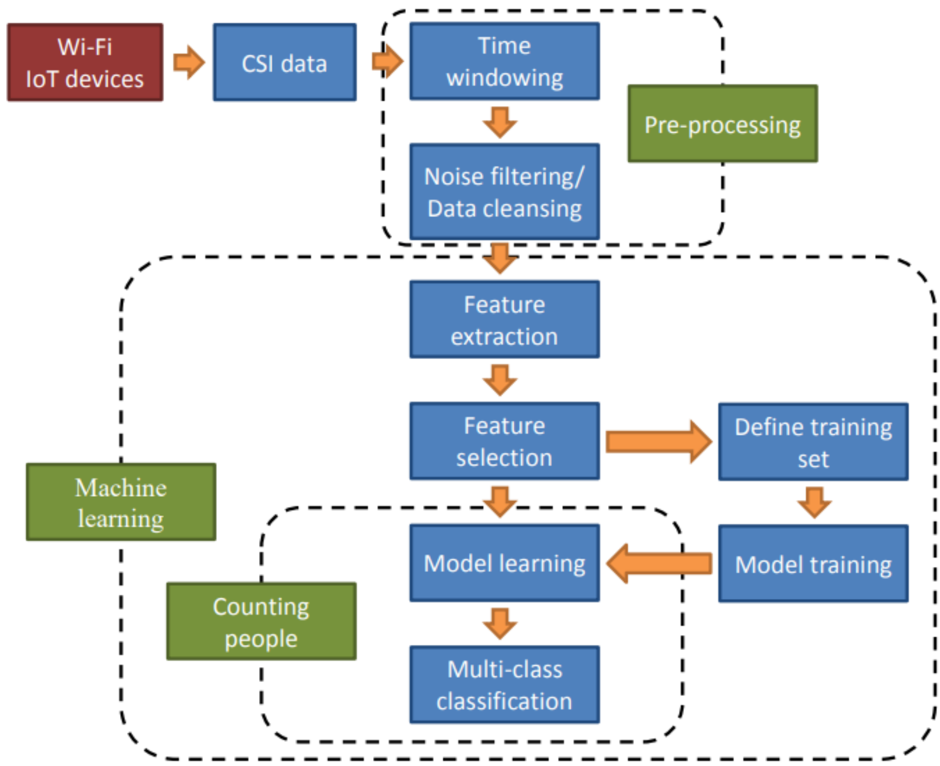

Fig. 1. Processing flow followed a free-device people counting WiFi device.

Figure 1: Figure 1

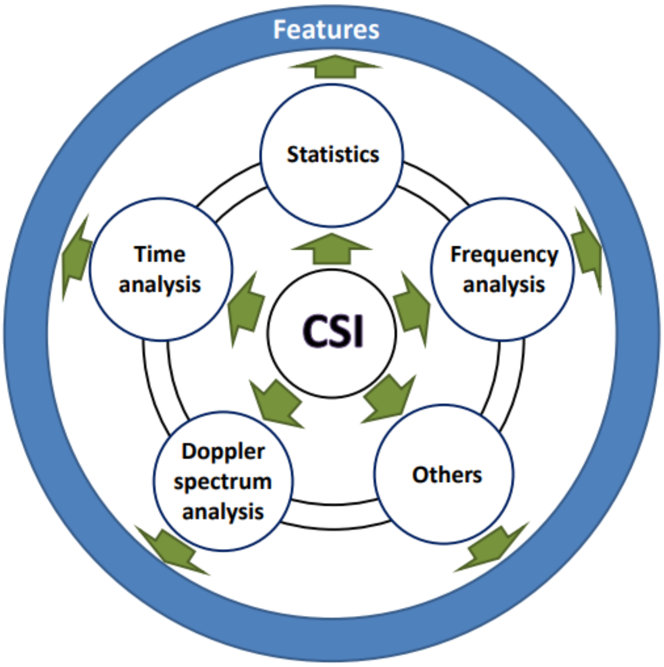

Fig. 2. Domains on which the feature engineering process for device-free people counting can be performed.

Figure 2: Figure 2

TABLE I

LITERATURE ON FEATURE ENGINEERING FOR DEVICE-FREE PEOPLE COUNTING CLASSIFIED BY DOMAIN IN WHICH FEATURES ARE COMPUTED (A [✓] INDICATES THAT THE LABELED FEATURE IS USED IN THE SIMULATIONS PRESENTED IN THIS WORK).

Domain	**Features**	**CSI used in every reference**	
		Amplitude	**Phase**
Time correlation	Eigenvalues	[16],[21],[22],[23],[✓]	[16],[22],[23],[✓]
	Average	[25]	–
Time statistics	Mean	[17],[25],[28],[29],[30],[32],[✓]	[17],[24],[30],[32]
	Std deviation	[18],[25],[30],[32]	[24],[30],[32]
	Variance	[26],[27],[28],[29],[32],[✓]	[32]
	Median	[18],[29],[30]	[30]
	Skewness	[28],[29],[30],[✓]	[30]
	Kurtosis	[25],[28],[29],[30],[✓]	[30]
	Mean-crossing rate	[28],[29]	–
	Inter-quartile range	[18],[29],[✓]	–
Probability Density Functions	Earth Mover's distance	[33],[34]	–
	Kullback-Leibler divergence	[35]	–
Doppler spectrum	Mean	[36],[✓]	[36],[✓]
	Std deviation	[36],[✓]	[36],[✓]
	Central moments	[36],[✓]	[36],[✓]
	Skewness	[36],[✓]	[36],[✓]
	Kurtosis	[36],[✓]	[36],[✓]
	Decay factor	[36]	[36]
	Spectral spread	[36]	[36]
	Largest peak	[28],[29]	–
Frequency analysis	Energy	[28],[29],[30],[32],[36],[✓]	–
	Max	[18],[28],[29],[30]	–
	Min	[28],[29],[30]	–
Others	Rician K-factor	[25]	–
	Entropy	[18],[28],[29],[30],[36]	[30],[36]
	1st derivative	[18],[36]	[36]
	Flatness	[36]	[36]
	Roll-off factor	[36]	[36]
	Euclidean distance	[17],[32]	[17],[32]
	Average step time	[30]	[30]
	Variance step time	[30]	[30]

Figure 3: Figure 3

TABLE II
ML MODELS UTILIZED IN THE LITERATURE

Supervised	Classification	Support Vector Machines (SVM)	[16],[18],[22],[26],[✓]
		Random Forest (RF)	[18],[29],[✓]
		Logistic Regression	[29],[✓]
		k Nearest Neighbor (kNN)	[28],[✓]
		Linear Discriminant Analysis (LDA)	[32],[✓]
		Naive Bayes (NB)	[36],[✓]
		Gaussian Mixture Model (GMM)	[30]
Unsupervised	Clustering	Density based (DBSCAN)	[21]
	Detection	Confidence Interval	[24]
		Threshold	[22],[23],[25]
		Hidden Markov Model	[26]
	Fitting	Earth Mover's Distance	[33],[34]
		Kullback-Leibler Divergence	[35]
		Grey-Verhulst Model	[27]

Figure 4: Figure 4

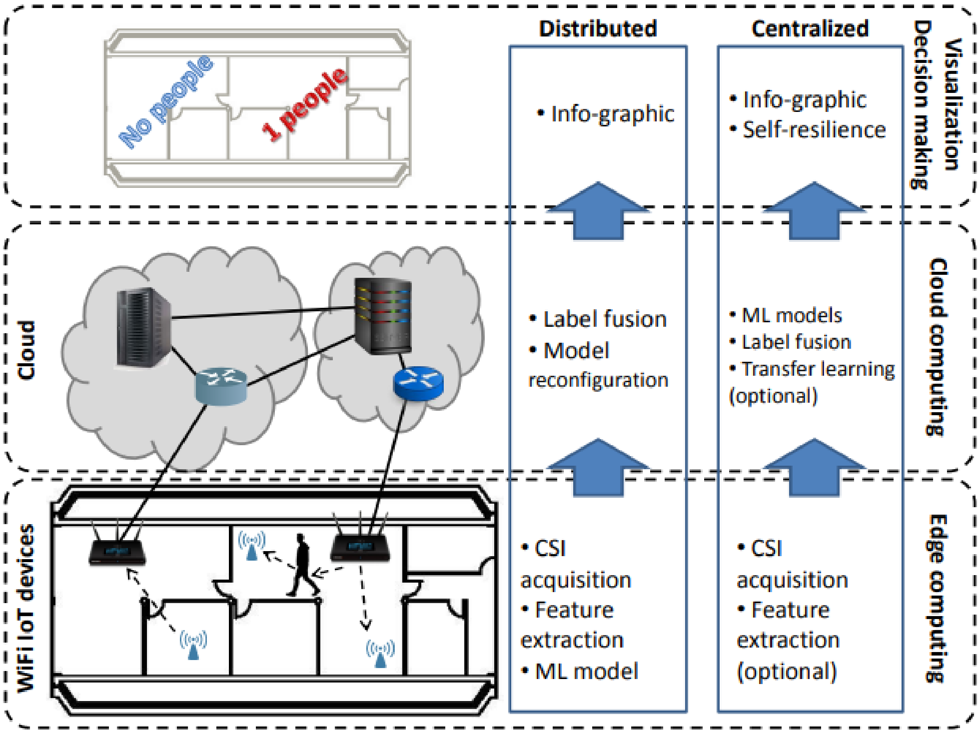

Fig. 3. Schematic diagram of the IoT architecture on which a device-free counting functionality can be deployed.

Figure 5: Figure 5

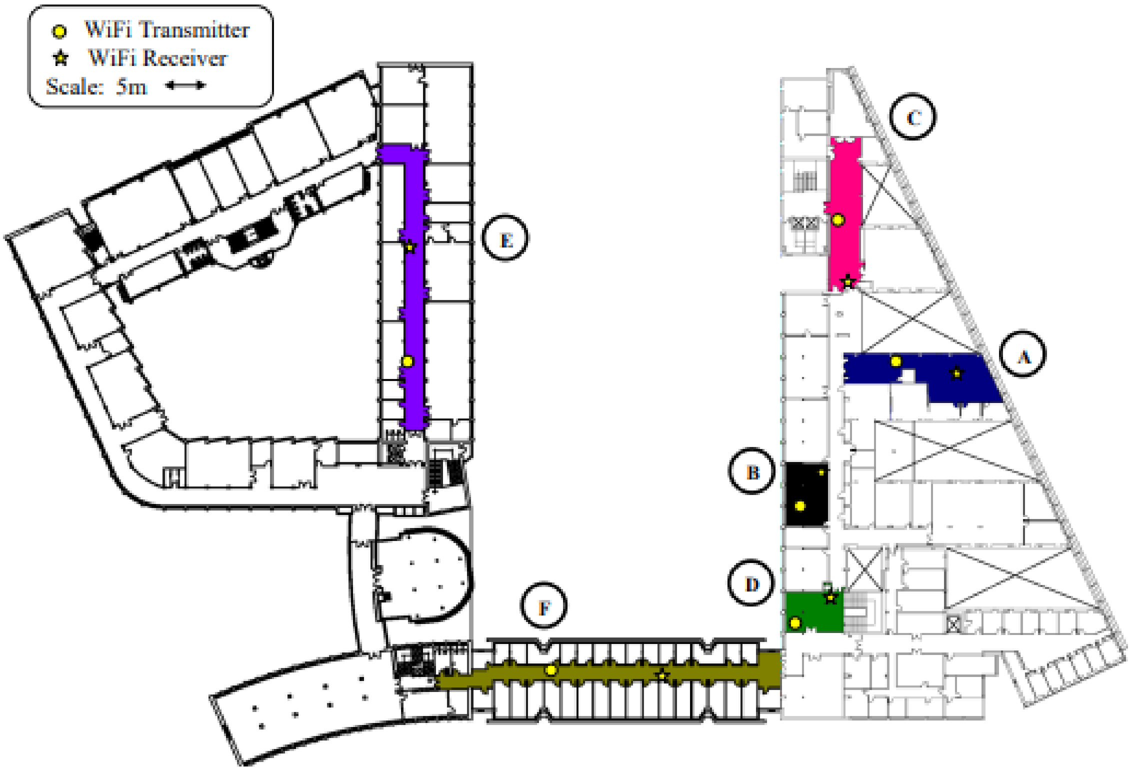

Fig. 4. Experimental scenarios (A to F) and testbed deployments embedded in the EHUCOUNT dataset presented in this work.

TABLE III
DESCRIPTION AND PARAMETERS OF THE SCENARIOS IN EHUCOUNT

Scenario	People (#)	P_t (dBm)	f_C (GHz)
A (Office)	[0,1,2,3]	-10	2.462
B (Lab)	[0,1,2,3,4,5]	-10	2.462
C (Corridor)	[0,1,2,3,4]	0	2.437
D (Hall+Stairs)	[0,1,2,3,4,5]	0	2.437
E (Corridor)	[0,1,2,3,4]	2	2.422
F (Corridor)	[0,1,2,3,4,5]	1	2.452

Figure 6: Figure 6

TABLE IV
FEATURES FOR THE ML MODEL CONSTRUCTION

Features	Indexes
Time-windowed amplitude-based eigenvalues	1-50
Time-windowed phase-based eigenvalues	51-100
Time-windowed mean of CSI amplitudes	101-153
Time-windowed variance of CSI amplitudes	154-206
Time-windowed skewness of CSI amplitudes	207-259
Time-windowed kurtosis of CSI amplitudes	260-312
Time-windowed inter-quartile range of CSI amplitudes	313-365
Doppler spectrum mean	366
Doppler spectrum variance	367
Doppler spectrum centroid	368
2nd order moment of Doppler spectrum	369
Doppler spectrum skewness	370
Doppler spectrum kurtosis	371
Time-frequency-windowed mean-std deviation ratio of CSI energy	372
Time-frequency-windowed skewness of CSI energy	373
Time-frequency-windowed kurtosis of CSI energy	374

Figure 7: Figure 7

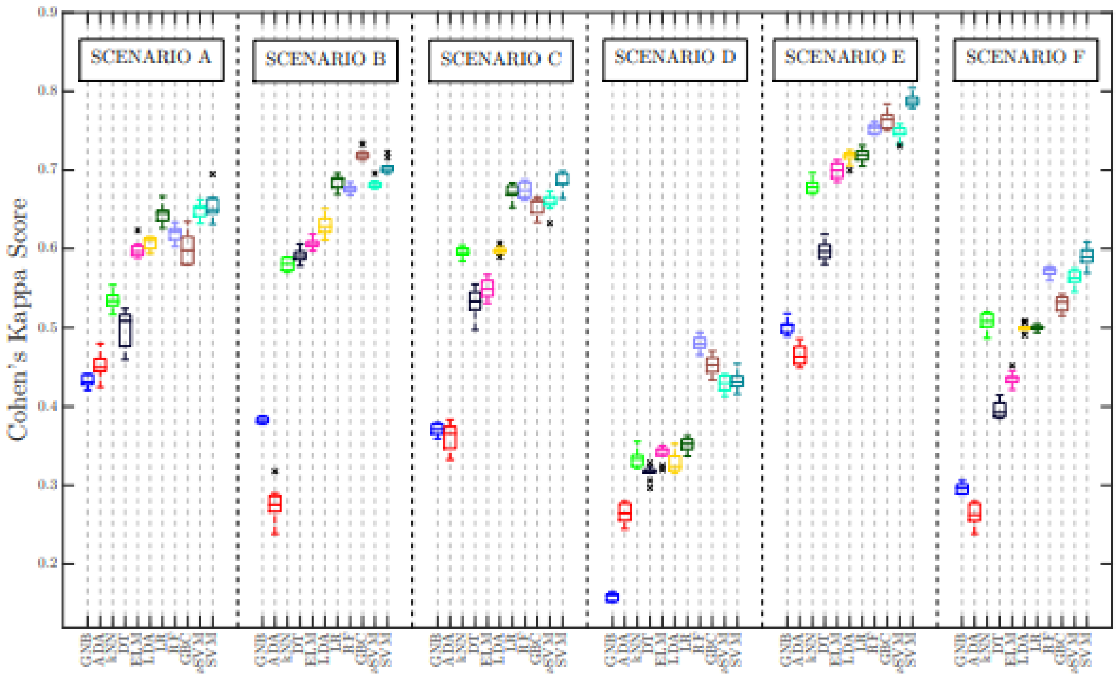

Fig. 5. Boxplot of the Kappa scores obtained by every model for all scenarios in the EHUCOUNT dataset, with FS. Score values falling outside the range $[Q_1 - 1.5 \cdot \text{IQR}, Q_3 + 1.5 \cdot \text{IQR}]$ (outliers) are highlighted as ×, with Q_1, Q_3 and IQR denoting the lower quartile, the upper quartile and the interquartile range, respectively.

Figure 8: Figure 8

TABLE V

ACCURACY AND COHEN'S KAPPA COEFFICIENT (FIRST AND SECOND ROW OF EVERY CELL, RESPECTIVELY), WITHOUT AND WITH FS. RESULTS ARE EXPRESSED AS MEAN±STD COMPUTED OVER NESTED CV WITH $M_{inner} = M_{outer} = 4$ FOLDS AND 20 REPETITIONS

Model	Scenarios											
	A		B		C		D		E		F	
	Without FS	With FS	Without FS	With FS	Without FS	With FS	Without FS	With FS	Without FS	With FS	Without FS	With FS
GNB	0.57 ± 0.00 0.43 ± 0.00	0.58 ± 0.01 0.43 ± 0.01	0.46 ± 0.00 0.35 ± 0.00	0.48 ± 0.02 0.38 ± 0.01	0.48 ± 0.00 0.35 ± 0.01	0.49 ± 0.00 0.37 ± 0.01	0.28 ± 0.00 0.14 ± 0.01	0.31 ± 0.01 0.16 ± 0.01	0.61 ± 0.00 0.51 ± 0.01	0.60 ± 0.01 0.50 ± 0.01	0.40 ± 0.00 0.28 ± 0.00	0.41 ± 0.00 0.30 ± 0.01
ADA	0.60 ± 0.01 0.46 ± 0.02	0.58 ± 0.01 0.45 ± 0.01	0.40 ± 0.00 0.27 ± 0.01	0.40 ± 0.01 0.28 ± 0.02	0.50 ± 0.01 0.37 ± 0.02	0.50 ± 0.01 0.36 ± 0.02	0.38 ± 0.01 0.26 ± 0.01	0.38 ± 0.01 0.27 ± 0.01	0.56 ± 0.02 0.45 ± 0.03	0.57 ± 0.02 0.47 ± 0.01	0.37 ± 0.01 0.25 ± 0.01	0.36 ± 0.01 0.26 ± 0.01
kNN	0.62 ± 0.01 0.50 ± 0.01	0.65 ± 0.01 0.53 ± 0.01	0.60 ± 0.01 0.52 ± 0.01	0.66 ± 0.00 0.58 ± 0.01	0.62 ± 0.01 0.52 ± 0.01	0.68 ± 0.00 0.60 ± 0.01	0.43 ± 0.01 0.31 ± 0.01	0.44 ± 0.01 0.33 ± 0.01	0.68 ± 0.01 0.60 ± 0.01	0.74 ± 0.02 0.68 ± 0.01	0.53 ± 0.00 0.44 ± 0.00	0.59 ± 0.01 0.51 ± 0.01
DT	0.61 ± 0.01 0.48 ± 0.02	0.60 ± 0.01 0.50 ± 0.02	0.65 ± 0.00 0.58 ± 0.01	0.68 ± 0.01 0.59 ± 0.01	0.62 ± 0.01 0.52 ± 0.01	0.64 ± 0.02 0.53 ± 0.02	0.42 ± 0.01 0.31 ± 0.01	0.42 ± 0.01 0.32 ± 0.01	0.67 ± 0.01 0.58 ± 0.01	0.67 ± 0.00 0.60 ± 0.01	0.48 ± 0.01 0.37 ± 0.01	0.50 ± 0.01 0.40 ± 0.01
ELM	0.66 ± 0.02 0.56 ± 0.01	0.71 ± 0.01 0.60 ± 0.01	0.66 ± 0.01 0.60 ± 0.01	0.68 ± 0.01 0.61 ± 0.01	0.58 ± 0.01 0.47 ± 0.01	0.64 ± 0.00 0.55 ± 0.01	0.39 ± 0.02 0.27 ± 0.01	0.44 ± 0.01 0.34 ± 0.01	0.72 ± 0.01 0.65 ± 0.01	0.75 ± 0.01 0.70 ± 0.02	0.48 ± 0.00 0.38 ± 0.00	0.53 ± 0.01 0.44 ± 0.01
LDA	0.67 ± 0.01 0.56 ± 0.01	0.72 ± 0.01 0.61 ± 0.01	0.68 ± 0.01 0.62 ± 0.01	0.70 ± 0.01 0.63 ± 0.01	0.65 ± 0.01 0.56 ± 0.01	0.68 ± 0.01 0.60 ± 0.01	0.41 ± 0.00 0.29 ± 0.01	0.44 ± 0.01 0.33 ± 0.01	0.73 ± 0.01 0.66 ± 0.01	0.77 ± 0.02 0.72 ± 0.01	0.55 ± 0.01 0.46 ± 0.01	0.58 ± 0.01 0.50 ± 0.01
LR	0.68 ± 0.01 0.57 ± 0.01	0.73 ± 0.01 **0.64 ± 0.01**	0.69 ± 0.01 0.62 ± 0.01	0.74 ± 0.01 0.69 ± 0.01	0.63 ± 0.02 0.54 ± 0.01	**0.74 ± 0.01** 0.67 ± 0.01	0.40 ± 0.01 0.28 ± 0.01	0.46 ± 0.01 0.35 ± 0.01	0.72 ± 0.01 0.65 ± 0.01	0.77 ± 0.01 0.72 ± 0.01	0.53 ± 0.01 0.44 ± 0.01	0.58 ± 0.01 0.50 ± 0.01
RF	**0.71 ± 0.01** **0.61 ± 0.01**	0.70 ± 0.00 0.62 ± 0.01	0.72 ± 0.01 0.66 ± 0.01	0.73 ± 0.00 0.68 ± 0.01	**0.71 ± 0.00** **0.65 ± 0.01**	**0.73 ± 0.01** 0.67 ± 0.01	**0.55 ± 0.01** **0.46 ± 0.01**	**0.57 ± 0.01** **0.48 ± 0.01**	0.79 ± 0.00 0.74 ± 0.01	0.80 ± 0.01 0.75 ± 0.01	**0.61 ± 0.01** **0.54 ± 0.01**	0.64 ± 0.01 0.57 ± 0.01
GBC	**0.71 ± 0.01** **0.61 ± 0.01**	0.70 ± 0.01 0.60 ± 0.02	**0.76 ± 0.00** **0.71 ± 0.01**	**0.76 ± 0.01** **0.72 ± 0.01**	**0.71 ± 0.01** **0.64 ± 0.01**	0.72 ± 0.01 0.65 ± 0.01	0.53 ± 0.01 0.44 ± 0.01	0.54 ± 0.00 0.45 ± 0.01	**0.81 ± 0.01** **0.76 ± 0.01**	0.81 ± 0.01 0.76 ± 0.01	0.58 ± 0.01 0.50 ± 0.00	0.60 ± 0.01 0.53 ± 0.01
νSVM	0.69 ± 0.01 0.58 ± 0.01	**0.75 ± 0.01** **0.65 ± 0.01**	0.71 ± 0.01 0.66 ± 0.01	**0.75 ± 0.01** 0.68 ± 0.01	0.64 ± 0.01 0.55 ± 0.01	0.72 ± 0.00 0.66 ± 0.01	0.48 ± 0.00 0.38 ± 0.01	0.54 ± 0.01 0.43 ± 0.01	0.78 ± 0.01 0.73 ± 0.01	0.80 ± 0.01 0.75 ± 0.01	0.58 ± 0.01 0.50 ± 0.01	**0.65 ± 0.02** 0.56 ± 0.01
SVM	0.69 ± 0.01 0.59 ± 0.01	**0.76 ± 0.01** **0.65 ± 0.02**	0.71 ± 0.01 0.66 ± 0.01	**0.75 ± 0.01** 0.70 ± 0.01	0.64 ± 0.01 0.55 ± 0.01	0.72 ± 0.02 **0.69 ± 0.01**	0.48 ± 0.00 0.38 ± 0.01	0.54 ± 0.00 0.43 ± 0.01	**0.80 ± 0.01** **0.75 ± 0.01**	**0.85 ± 0.01** **0.79 ± 0.01**	0.59 ± 0.01 0.51 ± 0.01	**0.66 ± 0.01** **0.59 ± 0.01**

Figure 9: Figure 9

TABLE VI

NORMALIZED CONFUSION MATRIX CORRESPONDING TO SCENARIO D WITH SVM AND FS

Scenario D, SVM		Predicted class					
		0	1	2	3	4	5
True class	0	80.8%	8.1%	2.7%	2.3%	3.5%	2.6%
	1	11.0%	60.0%	13.2%	12.0%	2.1%	1.7%
	2	6.6%	19.0%	38.1%	17.4%	10.2%	8.7%
	3	6.2%	16.3%	15.1%	43.5%	11.2%	7.7%
	4	3.1%	4.4%	12.2%	13.7%	54.5%	12.1%
	5	2.7%	5.0%	15.7%	16.0%	18.2%	42.4%

Figure 10: Figure 10

TABLE VII
NORMALIZED CONFUSION MATRIX CORRESPONDING TO SCENARIO E WITH SVM AND FS

Scenario E, SVM		Predicted class				
		0	1	2	3	4
True class	0	99.6%	0.4%	0.0%	0.0%	0.0%
	1	0.8%	90.6%	4.6%	1.6%	2.4%
	2	0.6%	6.9%	83.4%	2.5%	6.6%
	3	0.7%	3.8%	5.1%	69.6%	20.8%
	4	0.1%	1.5%	8.2%	16.9%	73.3%

Figure 11: Figure 11

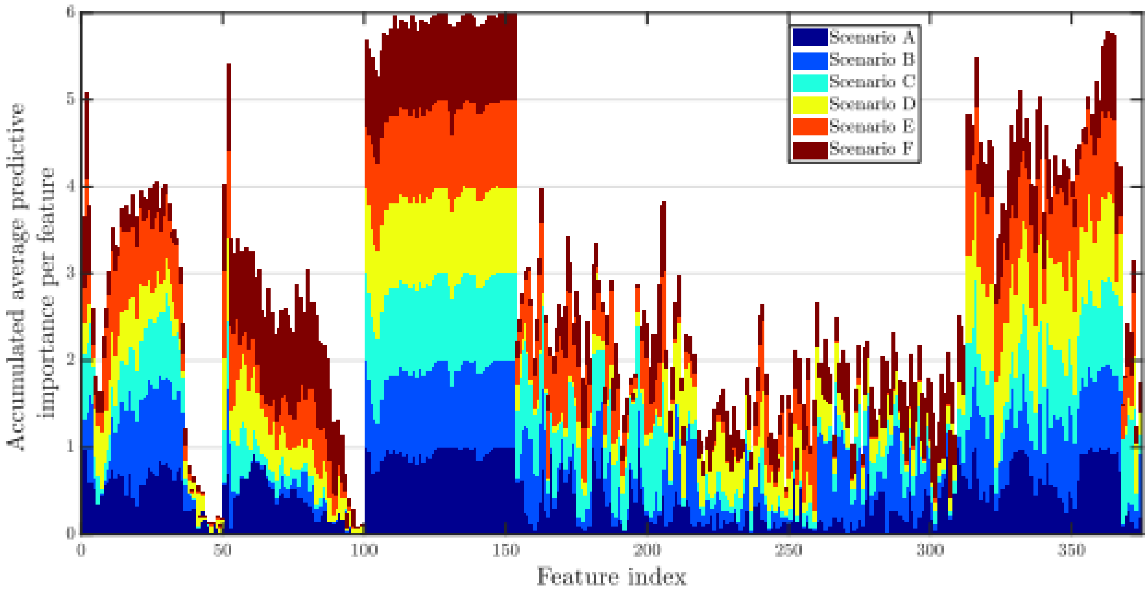

Fig. 6. Accumulative feature importance for the scenarios under study.

Figure 12: Figure 12