IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 5, 2021, accepted January 20, 2022, date of publication January 25, 2022, date of current version February 4, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3146346

Efficient Occupancy Detection System Based on Neutrosophic Weighted Sensors Data Fusion

NOHA S. FAYED[1], MOHAMMED M. ELMOGY[1], (Senior Member, IEEE), AHMED ATWAN[1,2], AND EMAN EL-DAYDAMONY[1]
[1]Information Technology Department, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt
[2]Faculty of Computing and Information Technology, Northern Border University, Rafha 1321, Saudi Arabia

Corresponding author: Mohammed M. Elmogy (melmogy@mans.edu.eg)

ABSTRACT Recently, a great interest has been dedicated to improving data fusion techniques for indoor occupancy detection. Indoor occupancy detection is extensively used in various applications, such as energy consumption control, surveillance systems, and disaster management. Using environmental sensors to collect data for detecting the occupancy state has the benefit of maintaining privacy. Also, it helps in improving monitoring systems and saving money due to energy consumption control. Nevertheless, sensor data is usually incomplete and noisy, which makes it uncertain and unreliable. These problems affect the detection accuracy. This paper proposes a comprehensive occupancy detection system that depends on a new fusion technique for fusing heterogeneous sensor data, which highly improves occupancy detection efficiency. Using Neutrosophy, the proposed technique handles sensor data uncertainty. Additionally, it improves reliability by fusing multiple sensors data. As it uses only one feature generated from fusing multiple sensors data, training and testing time is reduced. Consequently, the experimental results of applying the proposed fusion technique on a public benchmark dataset exhibit a significant enhancement in binary occupancy detection accuracy. The proposed technique enhanced the worst-case accuracy from 75.1 to 81.3%, 84.7 to 90.7%, 72 to 84.2%, 73.58 to 85.1%, and 65.9 to 78% using Linear Discriminant Analysis (LDA), K-Nearest Neighbors (K-NN), Naive Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF) classifiers, respectively. Using the other six performance metrics, the proposed technique results also outperform some state-of-the-art techniques.

INDEX TERMS Data fusion, heterogeneous sensors, neutrosophy, occupancy detection, uncertainty, unreliability.

I. INTRODUCTION

Indoor occupancy detection has a continually rising interest as an active research field due to its significant benefits in various critical applications. Examples of critical applications include building surveillance systems. Occupancy detection is used in building surveillance systems to provide additional human services, such as emergency response and decision support [1]. It also increases the security level of Intrusion Detection Systems (IDSs), which is an active research area, by detecting intruders and occupants' suspicious activities [2]. Another example is tracking space utilization, using occupancy detection techniques, which enables employees to locate colleagues and places to work in an address free environment. Lopez-de-Teruel *et al.* [3] proposed a localization system based on received signal strength for deploying fast occupancy services in a building.

Recently, many countries have been exerting every effort in response to climate change. Global warming is the main reason for this serious problem. Therefore, controlling energy consumption in buildings according to the occupancy state in these buildings is essential for decreasing global warming [4]. Energy consumption in buildings is about 40% of the global consumption of resources [5]. Hence, the allocation of spatiotemporal services [6] in smart buildings is another example of occupancy detection applications. Based on the occupancy state, services like air conditioning, heating, ventilating, and lighting systems can be automatically controlled, thus saving energy resources [7]. Energy consumption control is also an effective way to reduce dependency on fossil fuels and

The associate editor coordinating the review of this manuscript and approving it for publication was Wai-Keung Fung.

decrease CO_2 emissions. Enhancing indoor air quality and increasing occupants' comfort is another use of occupancy detection [8]. The utilizations mentioned above of occupancy detection are not the only uses that highlight its importance or rather necessity.

Risk assessment applications also use occupancy detection in many cases, such as dealing with criminal operations, environmental disasters, and indoor pollution [9]. During emergency planning (e.g., evacuation), knowing the occupancy state, the occupants' movements, and their positions can save many lives [10]. In such cases, the common real-time occupancy tracking techniques are based on camera data. There are two types of occupancy detection data sources: cameras and environmental sensors (e.g., CO_2, temperature, motion, humidity, and light).

A. CAMERA-BASED SYSTEMS

Cameras are a common data source for building occupancy estimation and detection systems because of their high precision. Using deep learning, Tien *et al.* [11] suggested a vision-based method for occupancy detection. The suggested method could detect multiple occupants as well as their activities. Liu *et al.* [12] suggested a system for detecting occupants inside a room and at the entrance. They proposed a two-stage static technique depending on shapes and appearances for occupancy detection in a room. As for detecting occupancy at the entrance, they suggested a motion-based technique. Also, they proposed a dynamic method based on a Bayesian network to fuse detection results at entrances with those inside rooms for more accuracy.

Zou *et al.* [13] presented an occupancy estimation framework based on a cascade classifier to detect human heads using cameras. In the first step, a pre-classifier was applied to concentrate on head windows. After that, a Convolutional Neural Network (CNN), the primary classifier, classified the head areas. Finally, for high estimation accuracy, a clustering analyzer was applied to fuse consecutive frames. Khalifa *et al.* [14] presented a new database for pedestrian detection. This database contains synchronized images captured from two types of cameras: a mobile car camera and a static road camera. They also suggested a novel framework for multi-view pedestrian detection based on the presented database.

Generally, camera-based occupancy detection and estimation systems can achieve high accuracy detection and estimation. Consequently, cameras are usually used to produce the ground-truth and labeled data for other occupancy detectors [15], [16]. Even though camera-based occupancy detection systems achieve high detection accuracy, they suffer from some problems, such as illumination conditions influence, high computational complexity, privacy concerns, and costly hardware required for advanced signal processing. Moreover, a line of sight is required for camera-based systems to minimize obstructions [17].

B. ENVIRONMENTAL-SENSOR-BASED SYSTEMS

Environmental sensors, including humidity, temperature, motion, CO_2, light, and pressure, are often existent in modern buildings, especially in lighting and Heating, Ventilation, and Air-Condition (HVAC) systems [18]. Considering that occupants' existence affects the indoor environment, environmental sensor measurements can be used as a good indication for occupancy. There exists a rich body of research on building occupancy detection systems based on environmental sensors. Some researchers used data from only one sensor type, such as dust concentrations [19], motion [20], and CO_2 data, which has shown good occupancy detection accuracy results, while others used multiple types [9], [21].

Environmental sensors are preferable to using cameras because sensor data processing requires fewer processing capabilities and smaller storage sizes. Besides, it maintains the privacy of individuals. However, uncertainty and unreliability are the main problems of using environmental sensor data because sensor data tends to be incomplete and noisy. These problems affect the detection accuracy and lead to a challenging area of research. For these reasons, this paper focuses on binary occupancy detection using only sensor data. Occupancy detection can be treated either as a binary classification problem or a multiclass one. The target of occupancy detection in the case of binary classification is predicting whether a specific place is occupied (1) or not (0). In the multiclass case, the target is the occupants' number [22], [23].

Most of the current occupancy detection researches concentrate on methods of classification [24], such as Support Vector Machines (SVMs) [15], [25], Neural Network (NN) [26], and Hidden Markov Models (HMM) [27], [28]. Dealing with the uncertainty and unreliability of data, however, is not given the required attention. Therefore, this paper proposes a new fusion technique to fuse heterogeneous sensor data based on Neutrosophic sets. Using Neutrosophy to represent a certain percentage of the data increases the data validity, providing better accuracy in detecting the occupancy state. Moreover, using a variety of sensor types increases data reliability. Fusing the training and testing phases' input features provides lower computational cost than using these features separately. These results were proved by applying the proposed technique on a public dataset for occupancy detection [29].

The remaining part of this paper is organized into five sections. Section 2 states contemporary fusion studies on sensor-based occupancy detection classified into three fusion levels, current limitations, and how we overcame these limitations in the proposed system. Section 3 introduces the proposed occupancy detection system's framework with a detailed explanation for the suggested fusion technique. Section 4 describes the conducted experiments. Section 5 discusses the experimental results of applying the suggested technique on a public dataset. Finally, Section 6 summarizes the proposed work conclusions and presents future work directions.

TABLE 1. A comparison of the three data fusion levels.

	F2F Fusion	F2D Fusion	D2D Fusion
Definition	-Features are combined to generate one feature.	-Features are combined using MLA to make a decision.	-Decisions obtained from multiple MLAs based on individual features are fused to make a final decision.
Pros	-It can benefit from features correlation. -One feature for learning is time efficient.	-It is simple: just apply MLA on the features.	-Decisions have the same representation. -It selects a suitable MLA for each feature.
Cons	-Fused feature must be in the same format.	Using many features: -is time inefficient. -may cause overfitting.	-It cannot benefit from features correlation. -Using multiple MLAs causes time inefficiency.

II. RELATED WORK

Contemporary environmental occupancy detection research can be classified based on used sensors into homogeneous sensors and heterogeneous sensors. In heterogeneous sensor-based occupancy, using a variety of sensor types for detection increases the reliability of data. Data fusion is also used to deal with heterogeneous sensor data, unlike homogeneous sensor data, which requires data aggregation. Thus, this section will be dedicated to classifying heterogeneous-sensors-based researches into three fusion levels: features-to-feature fusion (F2F), features-to-decision fusion (F2D), and decisions-to-decision fusion (D2D). In F2F, features extracted from sensors data are combined to generate one new feature. In F2D, the features are combined using a Machine Learning Algorithm (MLA) to make a decision. In D2D, decisions obtained from multiple MLAs based on individual features are fused to make a final decision. Each level has its pros and cons as summarized in Table 1. The following three subsections discuss the three fusion levels and their pros and cons in detail. Also, the related work researches were reviewed and classified, each according to the fusion level it represents.

A. FEATURES-TO-FEATURE (F2F) FUSION

Although using multi-sensor features can improve detection accuracy, it may lead to overfitting [29]. Also, using more than one feature increases time complexity. In the F2F fusion level, the features extracted from sensor data are combined to generate a new feature used as an input to a MLA for detecting the occupancy state. The main advantage of using this fusion level is that it can benefit from the correlation among multiple features. Also, only one feature will be used as input for the learning phase, therefore saving some computation time. However, the fused feature must be in the same format [30].

Chaney *et al.* [31] fused features derived from indoor CO_2, temperature, and electrical power sensors using a method that depends on Dempster-Shafer theory. After that, an HMM was used for predicting the occupancy. Also, the energy consumption was fused with the occupancy state for developing a metric to assess the likelihood of the occupant participating in a demand response at various times of the day. Based on cross-validation and ground truth information, the suggested approach could predict daytime occupancy and handle missing sensor data.

Christodoulou *et al.* [32] combined Fuzzy Cognitive Maps (FCM) with SVM. First, the correlation between sensor data variables (Temperature, Humidity, Humidity Ratio, Light, and CO_2) was attained through using FCM to generate a single variable. Then, SVM took the generated variable as an input to enhance prediction accuracy. Accuracies of 97.9 and 99.45 were achieved using two testing datasets [29].

Fayed *et al.* [33] proposed Neutrosophic Features Fusion (NFF) method to generate the fusion equation dynamically, using the correlation of Neutrosophic data. A neutrosophic feature was produced using CO_2, humidity, temperature, and light sensors readings from occupancy detection dataset [29]. Linear Discriminant Analysis (LDA), FUzzy GEnetic (FUGE), and Random Forest (RF) algorithms were used to prove that using the proposed method as a preprocessing step enhanced the worst-case accuracy. In the case of RF, the accuracy was enhanced from 57.51 to 88.16. Applying LDA, the accuracy enhanced from 75.13 to 88.01. On the other hand, using FUGE enhanced the accuracy from 57.93 to 84.55. Also, their method achieved accuracy up to 99.16 for the best-case accuracy using LDA.

B. FEATURES-TO-DECISION (F2D) FUSION

Regarding the F2D fusion level, the features are combined to make a decision. In other words, it is the phase of applying an MLA on the features to detect the occupancy state [34]. Although applying MLA directly to multi-sensor features is a simple operation and can improve the detection accuracy, it may lead to overfitting. Also, using multiple features for training is time-consuming. A lot of researches used this fusion level.

Lam *et al.* [27] applied HMM to features produced from acoustic and CO_2 sensors data. Due to the HMM model's ability to drop small sudden changes in occupancy levels during static intervals, HMM achieved a reasonable accuracy of 80% in detecting the occupants' number. Hailemariam *et al.* [35] used a decision tree to fuse CO_2, power use, motion, and sound sensors features. Using the root mean square error feature of a passive infrared motion sensor, a good accuracy of 97.9% was achieved in occupancy detection. The accuracy was increased to 98.4% by fusing multiple motion sensor features using a decision tree.

Yang *et al.* [36] predicted the occupants' number using the Radial Basis Function neural network with an accuracy of 87.62%. This NN used the radial basis function, which uses Euclidean distance, as the hidden layers activation function. Hence, using the radial basis function allowed converting low-dimensional inputs (linear inseparable) to high-dimensional inputs (separable). The sensors data used were humidity, light, sound, motion, CO_2, and temperature.

Ekwevugbe *et al.* [37] proposed an Adaptive Neuro-Fuzzy Inference System (ANFIS) based method for predicting the occupancy. This method was suggested to combine indoor climatic measures, indoor events, and energy consumption. The suggested method was expected to increase reliability. Ekwevugbe *et al.* [26] used Feed-Forward NN with back-propagation learning to fuse information, including case temperature, motion, CO_2, and sound level, for estimating the number of occupants. For feature selection, they used symmetrical uncertainty analysis. Also, they used a genetic-based search to optimize the sensor combination. They achieved an accuracy of 75% using the proposed method.

Using relative humidity, light, CO_2, temperature, infrared, sound, door switch, and motion sensors, Yang *et al.* [38] used various techniques to detect occupancy in different occupancy levels. The best accuracy range, [96.0% - 98.2%], was achieved using the decision-tree method. Their results showed that the proposed occupancy-based demand-response HVAC control could save 18% of electricity and 20% of gas compared to the conventional HVAC control.

Using the Auto-Regressive Hidden Markov Model (ARHMM), Ai *et al.* [39] estimated the occupants' number with an accuracy of 84%. The suggested method derived the ARHMM autoregressive part coefficients and analyzed wireless sensor network data. The analyzed data was from temperature, PIR, reed switches, airspeed, CO_2, and relative humidity sensors. The results showed that ARHMM was better than HMM in estimating the occupants' number when the occupancy level was changed frequently.

Candanedo and Feldheim [29] combined temperature, light, CO_2, and humidity sensors data using different statistical classification models: RF, Gradient Boosting Machines (GBM), LDA, and Classification and Regression Trees (CART). High accuracy of 97% was achieved by LDA using two sensors data: (light, CO_2), (temperature, light), (light, humidity), or (light, humidity ratio). These promising results were owing to the aforementioned combinations having good separation for occupancy status. Low accuracy of (68.63% and 32.68%) was obtained by RF using temperature, humidity, CO_2, and humidity ratio. These poor results were due to the high correlation between variables.

Hua *et al.* [40] fused temperature, lights energy, working time, and solar factor using the Support Vector Regression technique. They produced the training and testing data using the thermal software EnergyPlus. The error ratio of using the 5-feature model was 0.0264, while for the 4-feature model was 0.0532. Hence, the performance of the 5-feature model was better than the 4-feature model's one because the 4-feature model suffered from under-fitting.

Tutuncu *et al.* [41] applied seven various NN algorithms on humidity, temperature, CO_2, and light sensors data from the UCI dataset [29]. The seven NN algorithms were Batch Back Propagation (BBP), Levenberg-Marquardt, Conjugate Gradient Descent, Online Back Propagation, Limited Memory Quasi-Newton (LMQN), Quick Propagation, and Quasi-Newton. LMQN algorithm achieved the highest accuracy (99.06%), while the BBP algorithm achieved the lowest one (80.32%). Alghamdi [42] combined humidity rate, temperature, CO_2, and light sensors data from the UCI dataset [29] using Naïve Bayes (NB), Ada boosting, SVM, and K-Nearest Neighbors (K-NN). The accuracy achieved by NB and SVM was 94%. The best accuracy of 99% was achieved by Ada boosting and K-NN.

For binary occupancy classification, Kraipeerapun and Amornsamankul [7] used stacking for multiclass classification. They used two outputs NN and stacking, to fuse relative humidity, temperature, light, humidity ratio, and CO_2 from the UCI dataset [29]. The multiclass classification stacking outputs were combined to obtain the binary classification. For binary classification, the accuracy of the proposed stacking method was better than classical stacking. Average accuracy of 90.27% was achieved for the five input features.

For the unsupervised occupancy detection problem, Candanedo *et al.* [9] applied a suggested HMM-based method on only one or two features from temperature, CO_2, humidity ratio, light time, and humidity readings to infer occupancy schedules. The model was evaluated using a labeled dataset from UCI [29]. The suggested method was also applied to a case study for humidity ratio in building different rooms to infer occupancy schedules. There was no ground truth data, so the estimated occupancy schedules were validated with one building occupant. The best accuracy (90.24%) was achieved using the CO_2 data first order difference.

Pedersen *et al.* [43] suggested a rule-based method to determine the probability of occupancy. Two different sets of rules were suggested: one for PIR and noise sensors data and the other for relative humidity, air temperature, Volatile Organic Compound (VOC), and CO_2 concentration sensors data. Accuracy of 98% at most was reported for the first set. Masood *et al.* [16] presented two novel feature selection algorithms: Wrapper Rank-Extreme Learning Machine (WR-ELM) and Relative Information Gain-ELM (RIG-ELM). WR-ELM obtained its best accuracy using a combination of pressure, temperature, CO_2, and humidity sensors features. On the other hand, RIG-ELM needed only CO_2 features to obtain its best accuracy. Accuracies higher than 96% were reported.

Based on occupancy status, Kim and Moon [44] suggested a new thermal comfort control algorithm. The suggested algorithm included two parts: one for occupancy status detection and the other for switching on the devices. The suggested algorithm contained a multinomial logistic regression model. On the other hand, an integrated comfort algorithm was used to operate HVAC systems upon the outdoor environmental conditions to guarantee occupant's comfort without wasting energy. The suggested algorithm obtained an accuracy of 94.9% using PIR, CO_2 concentrations, lighting electricity consumption data, and door sensors.

Koklu and Tutuncu [45] used three classification algorithms (Decision Tree, RF, and Bagging) to combine temperature, light, CO_2, and humidity sensors data from UCI dataset [29]. The highest accuracy achieved was 99.368% using RF. Elkhoukhi *et al.* [46] proposed an online distributed machine learning framework to predict the occupancy state. They used a distributed version of the decision tree classifier, called Vertical Hoeffding Tree (VHT), to combine data from temperature, power consumption, CO_2, light, and humidity sensors. An accuracy of 95% was achieved using Occupancy Detection dataset [29], while an accuracy of 80% was achieved using deployed sensors data streams.

Giri *et al.* [47] combined temperature, light, CO_2, and humidity sensors data using different classification models: Classification Via Regression, RF, Multi-class Classification, Naive Bayes, Simple Logistic, and Decision Table. High accuracy of 99.0874% was achieved by Simple Logistic. Kampezidou *et al.* [48] proposed a Physics-Informed Pattern Recognition Machine (PI-PRM) method for occupancy detection using CO_2 and temperature sensors data. Using PI-PRM, which is a multi-layer perceptron NN, achieved an accuracy of 97%.

Wang *et al.* [49] presented a two-layer occupancy detection method. The first layer detects five human activities using temperature and PIR sensors data. The five activities were inside and outside door handle touch, tap and toilet usage, and motion near the door area. The second layer detect the occupancy state (1: entering, −1: leaving, and 0: no change) using RF, Decision Tree, K-NN, or SVM. Higher accuracy of 99% was achieved using RF.

C. DECISIONS-TO-DECISION (F2D) FUSION

At the D2D fusion level, occupancy states (decisions) obtained from multiple MLAs based on individual features are fused to decide the final occupancy state (the final decision). The main advantage of using the D2D fusion level is that the suitable MLA for each feature can be used. Besides, it is easy to fuse decisions that have the same representation. On the other hand, using this level has two disadvantages: failing to use the correlation among features and increasing the computation time due to multiple MLAs [30].

Chen *et al.* [15] presented a wrapper method depending on ELM for selecting the convenient features. They predicted an initial occupants' estimation using different models, which are SVM, K-NN, NN, LDA, ELM, and CART. Then, these estimations were fused using a particle filter algorithm. An accuracy of 93% was achieved by implementing the suggested fusion framework. Yang *et al.* [50] used the all-subsets regression model to select features from humidity, light, temperature, and CO_2 data. Then, they applied multiple ELM models to the selected features. After that, the resulting decisions were combined using a voting algorithm, Voting-based Weighted Extreme Learning Machine (WV-ELM). High detection accuracy of 97.32% was achieved by using light and CO_2 sensors data.

From the mentioned related works, which were summurized in Table 2, only Fayed *et al.* [33] dealt with the data uncertainty using the Neutrosophic approach, which can handle the problem of data uncertainty [26]. The other studies focused on handling the decision uncertainty using either the probability theory [31], [39] or the fuzzy set theory [32], [37]. Fayed *et al.* [33] fused Neutrosophic features using a dynamic equation. Despite achieving a good accuracy enhancement, producing the equation consumes O(n) time. Besides, the fusion equation requires alteration according to changes in the number or the type of sensors. Consequently, it limits the system's scalability. For this reason, the following section suggests a neutrosophic weighted fusion technique to achieve high accuracy while maintaining low time consumption. This technique is a F2F fusion technique, used to benefit from features correlation to handle data uncertainty in a time efficient manner as mentioned in F2F fusion section and summarized in Table 1.

III. THE EFFICIENT OCCUPANCY DETECTION SYSTEM

In this section, the stages of an efficient and comprehensive occupancy detection system using environmental sensors are discussed in detail. The proposed occupancy detection system is a binary occupancy detection system that uses environmental sensors as its source of information. To handle sensor data heterogeneity, a sensors data fusion technique is suggested. After that, the fused data is used by a classifier to make the final decision. Accordingly, a framework for an efficient occupancy detection system based on the Neutrosophic Weighted Fusion (NWF) method is proposed to handle the data uncertainty. As shown in Fig. 1, the proposed system framework consists of four stages: preprocessing, feature extraction, neutrosophic weighted fusion, and occupancy state detection stages. The main contribution of this paper is in the feature extraction and neutrosophic weighted fusion stages. In the following subsections, the four stages are described in details.

A. PREPROCESSING STAGE

Sensors data is noisy and redundant. It may also have different formats such as numeric data from simple sensors, binary or categorical data from switch-based sensors, video, images, and audio from complex sensors like cameras and microphones. Using and transmitting raw sensor data is costly and not effective. Thus, sensor data should be processed before using and transmitting it. The preprocessing stage's objective is to remove the noise and redundant data, decrease transmission cost, decrease storage requirements, and enhance usability [51]. According to its node capabilities, the preprocessing stage can be done locally on each sensor node or remotely on the sink/edge node. This stage consists of four steps, which are noise smoothing, missing values handling, sampling, and outlier removal [52].

TABLE 2. A comparison of current sensor-based occupancy detection related works.

Author / Year	Fusion Level	Fusion Method	MLAs	Dataset	Sensors Data	Accuracy	Notes
Chaney *et al.* **/ 2016 [31]**	F2F	Dempster-Shafer	HMM	Private dataset	CO_2, Temperature, and Electrical Power	No detection accuracy only daytime occupancy profile mentioned	-Private data -Daytime occupancy profile -No accuracy mentioned
Christodoulou *et al.* **/ 2017 [32]**	F2F	FCM	SVM	Occupancy Detection [29]	Temperature, Humidity, Humidity Ratio, Light, and CO_2	97.9% and 99.45%	Requires domain experts to build connections between the concepts in FCM
Fayed *et al.* **/ 2019 [33]**	F2F	NFF	LDA, FUGE, and RF	Occupancy Detection [29]	Temperature, Humidity, Light, and CO_2	Up to 99.16%	Construct dynamic fusion equations based on the data correlation
Lam *et al.* **/ 2009 [27]**	F2D	HMM	HMM	Private dataset	Acoustic and CO_2	80%	-Occupancy Estimation -Private dataset
Hailemariam *et al.* **/ 2011 [35]**	F2D	Decision Tree	Decision Tree	Private dataset	CO_2, Power use, Motion, and Sound	97.9% and 98.4%	Private dataset
Yang *et al.* **/ 2012 [36]**	F2D	NN	NN	Private dataset	Humidity, Light, Sound, Motion, CO_2, and Temperature	87.62%	-Occupancy Estimation -Private dataset
Ekwevugbe *et al.* **/ 2012 [37]**	F2D	ANFIS	ANFIS	Private dataset	Indoor climatic measures, Indoor events, and Energy consumption	No detection accuracy mentioned	-Private dataset -No detection accuracy mentioned
Ekwevugbe *et al.* **/ 2013 [26]**	F2D	NN	NN	Private dataset	Temperature, Motion, CO_2, and Sound level	75%	-Occupancy Estimation -Private dataset
Yang *et al.* **/ 2014 [38]**	F2D	Decision Tree	Decision Tree	Private dataset	Humidity, Light, CO_2, Temperature, Infrared, Sound, Door switch, and Motion sensors	96.0% - 98.2%	-Occupancy Estimation -Private dataset
Ai *et al.* **/ 2014 [39]**	F2D	ARHMM	ARHMM	Private dataset	Temperature, PIR, Reed switches, Airspeed, CO_2, and Humidity	84%	-Occupancy Estimation -Private dataset
Candanedo *et al.* **/ 2016 [29]**	F2D	RF, GBM, LDA, and CART	RF, GBM, LDA, and CART	Occupancy Detection [29]	Temperature, Light, CO_2, and Humidity sensors	Up to 97%	
Hua *et al.* **/ 2016 [40]**	F2D	Support Vector Regression	Support Vector Regression	Private dataset	Temperature, Lights energy, and Solar factor	Error Ratio: 0.0264	Private dataset
Tutuncu *et al.* **/ 2016 [41]**	F2D	7-NN Algorithms	7-NN Algorithms	Occupancy Detection [29]	Humidity, Temperature, CO_2, and Light, Humidity Ratio	Up to 99.06%	
Alghamdi / 2016 [42]	F2D	NB, Ada boosting, SVM, and K-NN	NB, Ada boosting, SVM, and K-NN	Occupancy Detection [29]	Humidity, Temperature, CO_2, and Light	94% and 99%	
Kraipeerapun and Amornsamankul / 2017 [7]	F2D	NN and Stacking	NN and Stacking	Occupancy Detection [29]	Humidity, Temperature, Light, Humidity Ratio, and CO_2	90.27%	
Candanedo *et al.* **/ 2017 [9]**	F2D	HMM	HMM	Occupancy Detection [29]	Temperature, CO_2, Humidity Ratio, Light time, and Humidity	90.24%	
Pedersen *et al.* **/ 2017 [43]**	F2D	Rule-based method	Rule-based method	Private dataset	Humidity, Temperature, VOC, CO_2, PIR, and Noise	98%	Private dataset
Masood *et al.* **/ 2017 [16]**	F2D	WR-ELM and RIG-ELM	WR-ELM and RIG-ELM	Private dataset	Pressure, Temperature, CO_2, and Humidity	96%	-Occupancy Estimation -Private dataset
Kim and Moon / 2018 [44]	F2D	Multinomial Logistic Regression	Multinomial Logistic Regression	Private dataset	PIR, CO_2, Lighting Electricity consumption, and Door sensors	94.9%	Private dataset
Koklu and Tutuncu / 2019 [45]	F2D	RF, Decision Tree, and Bagging	RF, Decision Tree, and Bagging	Occupancy Detection [29]	Humidity, Light, Temperature, and CO_2	Up to 99.368%	
Elkhoukhi *et al.* **/ 2020 [46]**	F2D	VHT	VHT	Occupancy Detection [29] and Deployed sensors data streams	Temperature, Power consumption, Light, CO_2, and Humidity	95% and 80%	Private dataset
Giri *et al.* **/ 2021 [47]**	F2D	Naive Bayes, Classification Via Regression, RF, Simple Logistic, Multi-class Classification, Decision Table	Naive Bayes, Classification Via Regression, RF, Simple Logistic, Multi-class Classification, Decision Table	Private dataset	Humidity, Light, Temperature, and CO_2	Up to 99.0874%	Private dataset
Kampezidou *et al.* **/ 2021 [48]**	F2D	PI-PRM	PI-PRM	Private dataset	CO_2 and Temperature	97%	Private dataset
Wang *et al.* **/ 2021 [49]**	F2D	RF, Decision Tree, K-NN, and SVM	RF, Decision Tree, K-NN, and SVM	Private dataset	Temperature and PIR	Up to 99%	Private dataset
Chen *et al.* **/ 2016 [15]**	D2D	Particle Filter	SVM, K-NN, NN, LDA, ELM, and CART	Private dataset	CO_2, Humidity, Temperature, and Pressure levels	93%	Private dataset
Yang *et al.* **/ 2021 [50]**	D2D	WV-ELM	ELM	Private dataset	Humidity, Light, Temperature, and CO2	97.32%	Private dataset

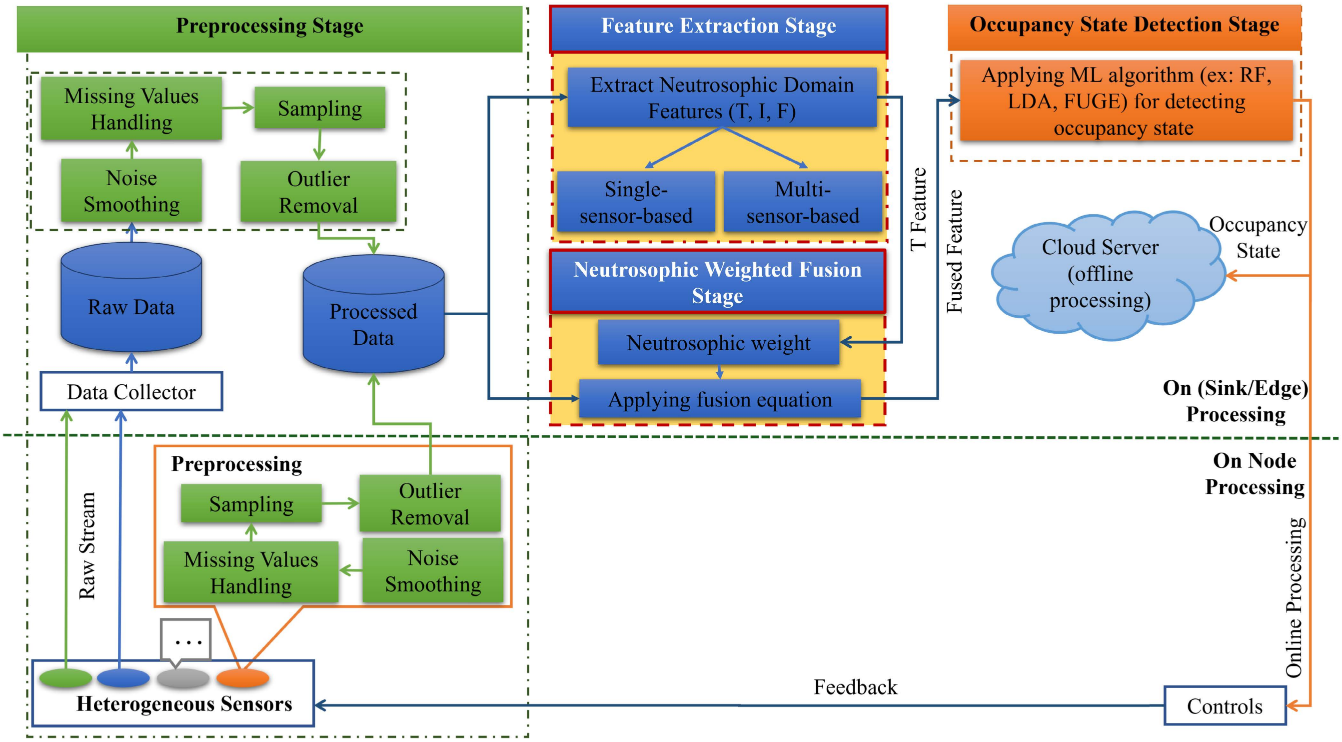


FIGURE 1. The proposed framework of an occupancy detection system based on NWF method.

- **Noise smoothing:** This step aims to remove random transient noise without affecting the original data.
- **Missing Values Handling:** Sensor data is subject to missing values because of electrical circuitry uncertainties. Data with missing values is harder to process. So, it could be handled by replacement or discarding before using it.
- **Sampling:** Sensor data is usually sensed at a high rate, but the actual life situations of interest do not change at that rate. Thus, sensor data at a lower rate can be more appropriate.
- **Outlier Removal:** Sensor data should be tested for the existence of outliers. After that, the detected outliers are replaced using mechanisms like missing value replacement.

After these steps, the preprocessed sensor's data is clean and ready to be used for the feature extraction stage.

B. FEATURE EXTRACTION STAGE

Feature extraction is about generating new features that are more informative and non-redundant for subsequent fusion steps [53]. A classifier cannot give reasonable results without features having discriminant power. Examples of simple features, which are suitable for real-time extraction, are mean, median, mode, standard deviation, etc. In the feature extraction stage, this paper suggests using sensors data in its Neutrosohpic Domain representation. Neutrosophy is defined as a philosophy branch that combines logic, probability/statistics, and set theory with philosophical knowledge to handle the uncertainty problem. Data in neutrosophic logic is represented by Truth (T), Indeterminacy (I), and False (F) as a 3D space (T, I, F). Each dimension is in the range of [0, 1] [54]. The T dimension is suggested to be used after that as an input feature to the proposed neutrosophic weighted fusion stage in order to handle the uncertainty problem.

The neutrosophic features are generated using two proposed methods, *sensor-based* and *multi-sensor-based*, deduced from the method applied in [55]. Assume a sensors dataset $X = \{s_1, s_2 \ldots, s_n\}$. In the first method, *sensor-based*, transforming sensor readings set S_j to the neutrosophic domain is based only on its reading data. Equation (1) is the representation of S_j in the neutrosophic domain, where $1 \leq j \leq n$.

$$ND_{sj}(i) = \{T_{sj}(i), I_{sj}(i), F_{sj}(i)\} \tag{1}$$

where ND_{sj} is the neutrosophic representation for S_j data, and i is the i^{th} observation index in X. $T_{sj}(i)$, $I_{sj}(i)$, and $F_{sj}(i)$ represent Truth (T), Indeterminacy (I), and False (F) dimensions, respectively. ND is a dataset that contains the neutrosophic representation for n sensors data vectors. The Truth (T) membership values are derived using (2).

$$T_{sj}(i) = \begin{cases} \dfrac{\bar{S}_j(i) - Min(\bar{S}_j)}{Max(\bar{S}_j) - Min(\bar{S}_j)}, & \text{if } Max(\bar{S}_j) > Min(\bar{S}_j) \\ 0, & \text{if } Max(\bar{S}_j) = Min(\bar{S}_j) \end{cases} \tag{2}$$

where $T_{sj}(i)$ is the Truth dimension for S_j data, and i is the i^{th} observation index in X and the local mean, $\bar{S}_j(i)$, is computed using (3).

$$\bar{S_j}(i) = \frac{1}{W} \sum_{m=i-W/2}^{i+W/2} S_j(m) \tag{3}$$

where W is the window size and it could be assigned an even value in the range [2:2(n-1)]. $S_j(m)$ denotes the m^{th} measurement in the S_j vector. Max ($\bar{S_j}$) and Min ($\bar{S_j}$) denote the maximum and the minimum values in the local mean vector, $\bar{S_j}(i)$, respectively. The False (F) membership values are derived using (4).

$$F_{sj}(i) = 1 - T_{sj}(i) \tag{4}$$

Equation (5) is used to derive Indeterminacy (I) membership values.

$$I_{sj}(i) = \begin{cases} \dfrac{\delta_j(i) - Min(\delta)}{Max(\delta) - Min(\delta)}, & \text{if } Max(\delta) > Min(\delta) \\ 0, & \text{if } Max(\delta) = Min(\delta) \end{cases} \tag{5}$$

where $\delta_j(i)$ is the absolute value of the difference between an observation value and its local mean value and it is calculated as shown in (6). Max (δ) and Min (δ) denote the maximum and the minimum values in δ_j vector, respectively.

$$\delta_j(i) = abs(S_j(i) - \bar{S_j}(i)) \tag{6}$$

Algorithm 1 summarizes the proposed *sensor-based* Features Extraction process.

In the second method, *multi-sensor-based*, a 2D matrix whose columns are sensors readings vectors, is used for transforming all n sensors data vectors to the neutrosophic domain at the same time. Using this method, each sensor data can be affected by the other sensors during the transformation process. As in the real-world, one observation may affect another observation. For example, a place temperature may be affected by the lighting. According to each sensor readings range, the constraint of using the *multi-sensor-based* method is to arrange sensor vectors, either in descending or in ascending order. This arrangement helps in preventing high sensor readings from canceling the low readings sensors effect.

In (7), ND (i,j) is a neutrosophic dataset that contains the neutrosophic representation for n sensors data vectors.

$$ND(i,j) = \{T(i,j), I(i,j), F(i,j)\} \tag{7}$$

where i is the i^{th} observation index in X, and the index j refers to S_j vector. T(i,j), I(i,j), and F(i,j) represent Truth (T), Indeterminacy (I), and False (F) dimensions, respectively. The Truth (T) membership values are derived using (8).

$$T(i,j) = \begin{cases} \dfrac{\bar{X}(i,j) - Min(\bar{X})}{Max(\bar{X}) - Min(\bar{X})}, & \text{if } Max(\bar{X}) > Min(\bar{X}) \\ 0, & \text{if } Max(\bar{X}) = Min(\bar{X}) \end{cases} \tag{8}$$

where i is the i^{th} observation index and j refers to the S_j vector. The local mean, $\bar{X}(i,j)$, is computed, as shown in (9).

$$\bar{X}(i,j) = \frac{1}{W \times W} \sum_{m=i-W/2}^{i+W/2} \sum_{n=j-W/2}^{j+W/2} X(m,n) \tag{9}$$

Algorithm 1 Sensor-Based Features Extraction

Input: X; a sensors dataset where X = $\{s_1, s_2 \ldots, s_n\}$
Output: ND; dataset contains neutrosophic representation for X.

for all S_j in X **do**
  1.Compute the local mean $\bar{S_j}$
  for $i = 1$ **to** length (S_j) **do**
    $\bar{S_j}(i) = 0$
    for $m = (i - W/2)$ **to** $(i + W/2)$ **do**
      $\bar{S_j}(i) = \bar{S_j}(i) + S_j(m)$
    end for
    $\bar{S_j}(i) = \bar{S_j}(i)/W$
  end for
  2.Compute the Truth T_{Sj}
  lm_min = min ($\bar{S_j}$)
  lm_max = max ($\bar{S_j}$)
  for $i = 1$ **to** length (S_j) **do**
    if lm_max > lm_min **then**
      T_{Sj}(i) = ($\bar{S_j}$(i)- lm_min) / (lm_max - lm_min)
    else
      T_{Sj} (i) = 0
    end if
  end for
  3.Compute the False F_{Sj}
  for $i = 1$ **to** length (S_j) **do**
    $F_{Sj}(i) = 1 - T_{Sj}(i)$
  end for
  4.Compute the Indeterminacy I_{Sj}
  for $i = 1$ **to** length (S_j) **do**
    $\delta_j(i) = abs(S_j(i) - \bar{S_j}(i))$
  end for
  $\delta min = min(\delta)$
  $\delta max = max(\delta)$
  for $i = 1$ **to** length (S_j) **do**
    if $\delta max > \delta min$ **then**
      I_{Sj}(i) = $(\delta_j(i) - \delta min)/(\delta max - \delta min)$
    else
      I_{Sj} (i) = 0
    end if
  end for
  5.Construct ND_{Sj}
  $ND_{Sj} = \{T_{Sj}, I_{Sj}, F_{Sj}\}$
end for
ND = $\{ND_{S1}, ND_{S2} \ldots, ND_{Sn}\}$

where W × W is the window size and X(m,n) is the observation at (m,n) location in X. Max ($\bar{X}$) and Min ($\bar{X}$) denote the maximum and the minimum values in the local mean matrix $\bar{X}$, respectively. The False (F) membership values are derived using (10).

$$F(i,j) = 1 - T(i,j) \tag{10}$$

Equation (11) is used to derive Indeterminacy (I) membership values.

$$I(i,j) = \begin{cases} \dfrac{\delta(i,j) - Min(\delta)}{Max(\delta) - Min(\delta)}, & \text{if } Max(\delta) > Min(\delta) \\ 0, & \text{if } Max(\delta) = Min(\delta) \end{cases} \tag{11}$$

where δ(i,j) is the absolute value of the difference between an observation value and its local mean value and it is calculated as shown in (12). Max (δ) and Min (δ) denote the maximum and the minimum values in δ vector, respectively.

$$\delta(i,j) = abs(X(i,j) - \bar{X}(i,j)) \tag{12}$$

Algorithm 2 summarizes the proposed *Multi-Sensor-Based* Features Extraction process.

C. NEUTROSOPHIC WEIGHTED FUSION STAGE

Data fusion combines data from various sources to achieve more efficient and accurate inferences than what was achieved by using a single source [34]. There are two types of information to be fused: features or decisions. One of the simplest and most used fusion methods is Linear Weighted Fusion (LWF). In LWF, the sensor information is combined linearly using sum or product operators. To fuse sensor information, a normalized weight is assigned to each sensor's information.

Common normalized weights computation methods are decimal scaling, min-max, tanh-estimators, and z score. Although min-max, decimal scaling, and z score methods are easy to compute, they are affected by outliers. In contrast, the tanh method is effective, but its parameters are estimated using training [30]. Not to mention that all of these methods do not consider the uncertainty of data. Therefore, the suggested Neutrosophic Weighted Fusion (NWF) method is used, in the fusion stage, to generate a single fused feature from multiple sensors data. Then, the fused feature is used as input for the occupancy state detection stage. Using only one feature as input for the learning phase saves some computation time. The fused feature also depends on multi-sensor features, which can improve the detection accuracy without leading to the overfitting problem. Besides, the NWF method uses neutrosophic weights as a percentage of certainty for sensors data to handle sensors data uncertainty, which in turn increases the detection accuracy. So, using NWF makes the occupancy detection system more efficient.

The proposed NWF method uses Neutrosophy to determine and adjust weights working as a percentage of the sensor's data certainty. NWF uses the T dimension from the previous stage as a weight for the original sensor data to handle the uncertainty problem. For occupancy detection, using neutrosophic weight for fusion was not used formerly. The Truth-based Weight (TW) of a sensor data is computed using two methods, *sensor-based* and *multi-sensor-based*, mentioned in the previous stage. In the first method, *sensor-based* weight is calculated using (13).

$$TW_{sj}(i) = T_{sj}(i) \tag{13}$$

where $TW_{sj}(i)$ is the neutrosophic weights vector for the S_j data, and i is the i^{th} observation index in X.

The weights matrix that contains the weights vectors for the n sensors is denoted TW. In the second method, a *multi-sensor-based* weight is a 2D weights matrix with the weights vectors corresponding to the sensors' vectors. By using this method, each sensor data can be affected by the other sensors during the process of weight computing. Equation (14) is used

Algorithm 2 Multi-Sensor-Based Features Extraction

Input: X; a sensors dataset where $X = \{s_1,\ s_2 \ldots, s_n\}$
Output: ND; dataset contains neutrosophic representation for X.

```
1.Compute the local mean X̄
for i = 1 to nrow (X) do
  for j = 1 to ncol (X) do
    X̄(i, j) = 0
    for m = (i − W/2) to (i + W/2) do
      for n = (j − W/2) to (j + W/2) do
        if (m > 0 and m ≤ nrow(X)) and (n > 0 and
        n ≤ ncol(X)) then
          X̄(i, j) = X̄(i, j) + X(m, n)
        end if
      end for
    end for
    X̄(i, j) = X̄(i, j)/W × W
  end for
end for
2.Compute the Truth T
lm_min = min (X̄)
lm_max = max (X̄)
for i = 1 to nrow (X) do
  for j = 1 to ncol (X) do
    if lm_max > lm_min then
      T(i,j) = (X̄(i,j)- lm_min) / (lm_max - lm_min)
    else
      T(i,j) = 0
    end if
  end for
end for
3.Compute the False F
for i = 1 to nrow (X) do
  for j = 1 to ncol (X) do
    F(i,j) = 1-T(i,j)
  end for
end for
4.Compute the Indeterminacy I
for i = 1 to nrow (X) do
  for j = 1 to ncol (X) do
    δ(i, j) = abs(X(i, j) − X̄(i, j))
  end for
end for
δmin = min(δ)
δmax = max(δ)
for i = 1 to nrow (X) do
  for j = 1 to ncol (X) do
    if δmax > δmin then
      I(i,j) = (δ(i, j) − δmin)/(δmax − δmin)
    else
      I(i,j) = 0
    end if
  end for
end for
5.Construct ND
ND = {T, I, F}
```

to produce the weights matrix.

$$TW(i,j) = T(i,j) \tag{14}$$

where TW is the weights matrix. i is the i^{th} observation index, and j refers to the S_j vector. After computing the sensor's data weights, they are used in the fusion equation (15) to increase the training and testing data's certainty.

$$F(i) = \sum_{j=i}^{n} TW(i,j) \times X(i,j) \tag{15}$$

where F is the fused feature, i refers to its i^{th} observation in the dataset and X(i,j) is the measurement at (i, j) location. Sensors vectors are considered alternatives because each sensor vector can be used separately for occupancy detection. That is why the sum operator is used for fusing the data. Algorithm 3 shows the proposed Neutrosophic Weighted Fusion process.

Algorithm 3 Neutrosophic Weighted Fusion

Input: X; a sensors dataset where X = $\{s_1, s_2 \ldots, s_n\}$, **ND;** dataset contains neutrosophic representation for X.
Output: F; the fused feature

```
1.Compute the neutrosophic weights vector TW
for i = 1 to length (S_j) do
   for j = 1 to n do
      TW (i, j) = ND_T (i,j)
   end for
end for
2.Compute the fused feature F
for i = 1 to length (S_j) do
   F(i) = 0
   for j = 1 to n do
      F(i) = F(i)+ (TW(i, j) × X(i, j))
   end for
end for
```

D. OCCUPANCY STATE DETECTION STAGE

In this stage, a classification algorithm such as K-NN, LDA, NB, RF, or SVM is used to detect the occupancy state. The input to this stage is the fused feature resulted from the fusion stage. After detecting the occupancy state, the sink/edge node controls sensors or actuators based on the occupancy state or sends the occupancy state to the cloud for further processing and decision-making. The cloud performance and security issues should be taken into consideration [56], [57].

IV. EXPERIMENTAL RESULTS

This section is divided into five subsections. The first subsection describes a public occupancy detection dataset used to evaluate the proposed fusion technique. The second subsection states the hardware and software specifications used for the experiments. The third subsection specifies the performance metrics, and the fourth one presents the experimental results. Finally, the fifth subsection discusses the experimental results.

TABLE 3. The description of occupancy detection dataset [29].

Dataset	Total No. of observations	Empty observations(0)	Occupied observations(1)
Training	8143	79%	21%
Testing 1	2665	64%	36%
Testing 2	9752	79%	21%

A. DATASET DESCRIPTION

The proposed NWF technique is applied to a public dataset, which is occupancy detection [29]. The following subsection give more details about this dataset. The reasons for choosing this dataset are as follows:

- It is a public dataset from a well-known data repository (UCI Machine Learning Repository) [58].
- The dataset contains three sets one for training and two for testing the classification models. The testing data with different environmental conditions (one when the door is closed as training data and the other when the door is open) which is convenient for showing the effect of our approach on dealing with data uncertainty.
- The number of the used sensor types (4 types) is acceptable for fusion.
- The number of observations is suitable (Training: 8143, Testing1: 2665, Testing2: 9752).
- The data and its processing codes are available in [59], so the results comparison could be direct.
- The processing and evaluation was done using the open-source program R.

1) OCCUPANCY DETECTION DATASET

This dataset is experimental data from the UCI Machine Learning Repository [58]. The sensor readings were recorded every 14 seconds from sensors deployed in an office room with at most two occupants, then the mean of readings was calculated for each minute. The dataset is used for binary occupancy detection using environmental temperature, humidity, light, and CO_2 sensors. Every minute, a picture was taken to obtain the labels (occupancy state). The dataset contains three subsets; Training (for training), Testing1, and Testing2 (for testing). Table 3 summarizes the three datasets' details.

Most of the measurements in Training and Testing1 were taken with the door closed, while those of Testing2 were taken with the door opened. Each dataset contains readings of sensors: temperature (T) in Celsius, humidity (H) in percentage %, light (L) in Lux, and CO_2 in ppm labeled with the occupancy state (0: Empty, 1: occupied). Also, it contains a timestamp and derived humidity ratio (HR) in $kg_{vapour-water}/kg_{air}$, which is calculated using (16).

$$HR(i) = 0.622 \times \frac{p_w(i)}{p - p_w(i)} \tag{16}$$

where p = 101.325 kPa (standard atmospheric pressure) and p_w is calculated using (17).

$$H(i) = \frac{p_w(i)}{p_{ws}(i)} \tag{17}$$

where $p_{ws}(i)$ (the saturation pressure over liquid water in Pa) is calculated using (18).

$$ln(p_{ws}(i)) = \frac{C_1}{T_K(i)} + C_2 + C_3 \times T_K(i) + C_4 \times {T_K}^2(i) + C_5 \times {T_K}^3(i) + C_6 \times ln(T_K(i)) \tag{18}$$

where $C_1 = 5.8002206\text{E} + 03$, $C_2 = 1.3914993\text{E} + 00$, $C_3 = -4.86402396\text{E-}02$, $C_4 = 4.1764768\text{E - }05$, $C_5 = -1.44592093\text{E - }08$, and $C_6 = 6.5459673\text{E} + 00$. $T_K(i)$ is the absolute temperature and is calculated using (19).

$$T_K(i) = T(i) + 273.15 \tag{19}$$

where T(i) is the temperature in the Celsius unit.

B. HARDWARE AND SOFTWARE SPECIFICATIONS

A laptop with the following specifications was used to carry out the experiments:

- **Processor:** x64-based processor, Intel(R) Core (TM) i7-2640M, CPU at 2.80 GHz.
- **RAM:** 6 GB.
- **Operating System:** Windows 10 Education, 64-bit.

The open-source program R was used for implementing the Feature Extraction and Neutrosophic Weighted Fusion Stages. It is also used to evaluate and compare the occupancy detection based on the suggested technique with other state-of-the-art techniques.

C. PERFORMANCE METRICS

To evaluate the proposed fusion technique (NWF) performance, the following seven metrics were used:

- **Accuracy (ACC):** ACC is used to evaluate the classification model capability. In other words, it is the percentage of correct results (TP or TN). Generally, ACC is calculate using (20) [60].

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

  where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives.
- **Balanced ACC:** is the accuracy in case of imbalanced data. It is calculated using (21) [61].

$$BalancedACC = \frac{SPE + SEN}{2} \tag{21}$$

  where SPE is Specificity and SNE is Sensitivity. Both ACC and Balanced ACC were used as performance metrics in this paper. ACC was used to compare the proposed technique results with the results in [33] and with other state-of-the-art techniques, while Balanced ACC was used because the data is imbalanced.

TABLE 4. Variables combinations.

Case No.	Fused Features	Case No.	Fused Features
1	T, CO2, H, and L	7	T and CO2
2	T, H, and L	8	T and H
3	CO2, H, and L	9	H and L
4	T, CO2, and L	10	CO2 and L
5	T, CO2, and H	11	CO2 and H
6	T and L		

- **Specificity (SPE):** is the percentage of true negatives representing the classifier's ability to classify negative class patterns. It is calculated using (22) [62].

$$SPE = \frac{TN}{Neg} \tag{22}$$

  where Neg is the number of negative class patterns.
- **Sensitivity (SEN) or (Recall):** is the rate of true positives representing the classifier's ability to classify positive class patterns. The method can be precise without being sensitive, or it can be susceptible without being specific. SEN can be calculated using (23) [63].

$$SEN = \frac{TP}{Pos} \tag{23}$$

  where Pos is the number of positive class patterns.
- **Precision or Positive Predictive Value (PPV):** is the percentage of true positive cases related to all the predicted positive patterns. It can be calculated using (24) [62].

$$PPV = \frac{TP}{TP + FP} \tag{24}$$

- **F1-score (F-measure):** is a good indication for incorrectly recognized patterns than ACC. It is calculated using (25) [64].

$$F1 = 2 \times \frac{PPV \times SEN}{PPV + SEN} \tag{25}$$

- **Area Under the Curve (AUC):** it measures the ability to distinguish between classes. Also, it summarizes the Receiver Characteristic Operator (ROC) curve. ROC curve is a plot for TP rate against FP rate. A higher AUC is desirable because it means a better classifier performance at distinguishing between classes. AUC is calculated using (26) [64].

$$AUC = 0.5 \times (SEN + SPE) \tag{26}$$

D. RESULTS

The time consumption and the previously mentioned performance metrics for the proposed technique are analyzed in this section. They are also compared with other sensor data cases using K-NN, LDA, NB, RF, and SVM as classification algorithms. The six cases for comparison are a case

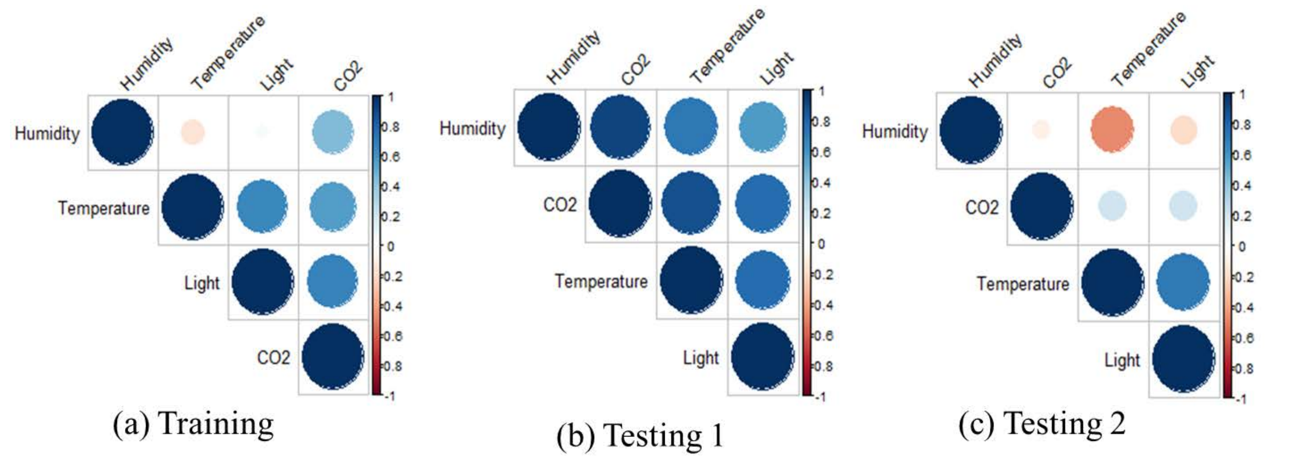


FIGURE 2. **Original dataset correlation plots.**

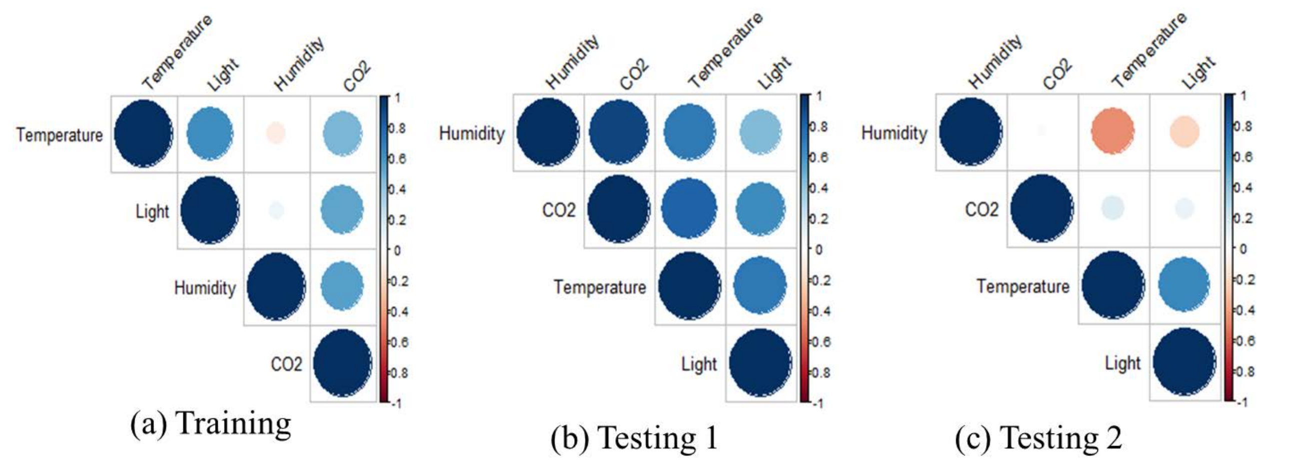


FIGURE 3. **SW dataset correlation plots.**

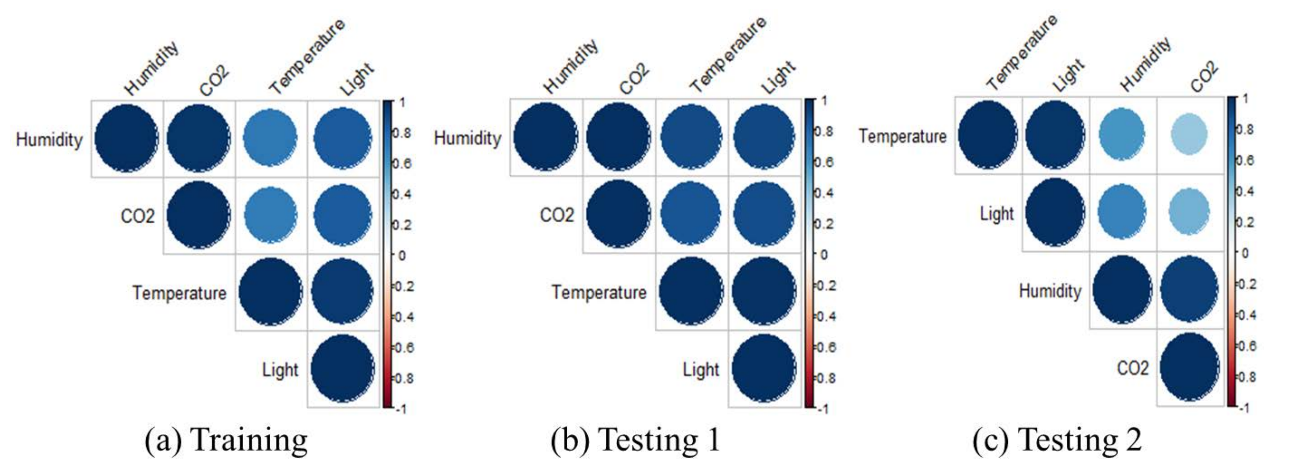


FIGURE 4. **MSW dataset correlation plots.**

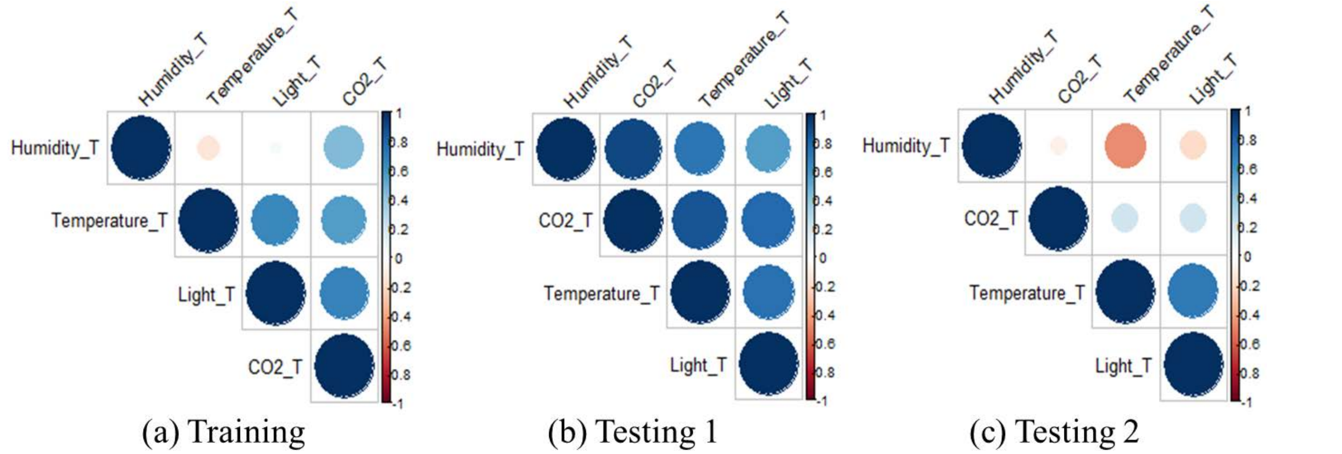


FIGURE 5. NS dataset correlation plots.

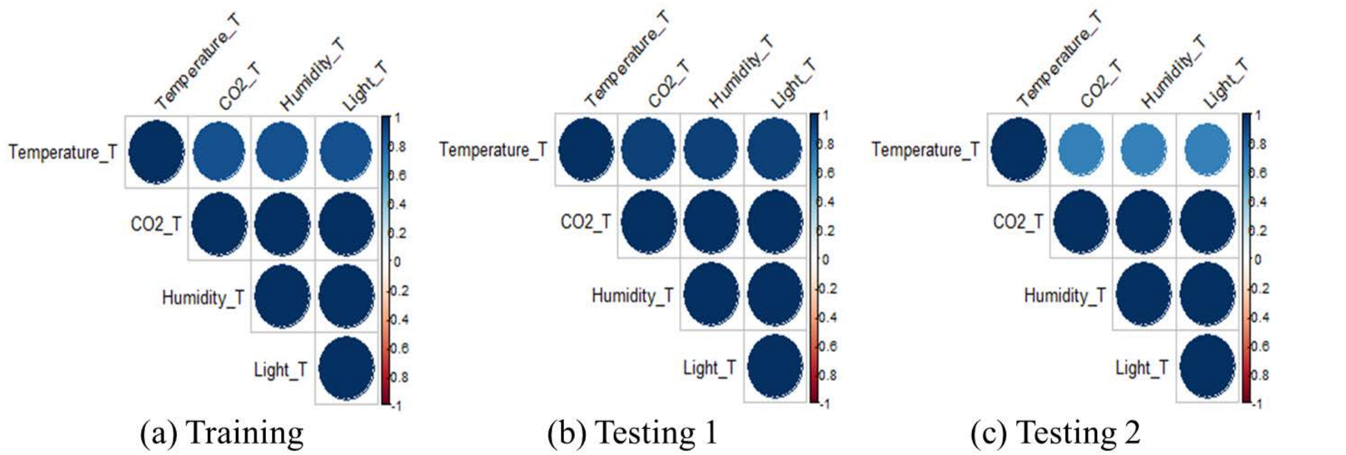


FIGURE 6. NS_all data set correlation plots.

for F2D fusion and five F2F fusion cases. The data used for each case are as follows:

- **No Fusion (NF) [29]:** original sensors data vectors without fusion (F2D fusion).
- **Unweighted Fusion (UWF):** a fused data vector generated through fusing sensors data weighted by 1 using the sum operator which is a well-known fusion method (F2F fusion).
- **Sensor-based Weighted Fusion (SWF):** a fused sensor-based weighted data vector generated using the NWF method (F2F fusion).
- **Multi-Sensor-based Weighted Fusion (MSWF):** a fused multi-sensor-based weighted data vector generated via NWF method (F2F fusion).
- **NS [33]:** a fused data vector generated using the dynamic fusion equations mentioned in [33]. The features fused are the truth of each sensor data (F2F fusion).
- **NS_all [33]:** a fused data vector generated using the dynamic fusion equations mentioned in [33]. The features fused are the truth of each sensor data affected by other sensors' data (F2F fusion).

To evaluate the proposed technique, various classification models are used. LDA and RF were used in [29], [33]. So, they were used for comparing the proposed technique results with the results in [29], [33]. Also, K-NN, NB, and SVM are used to compare with other state-of-the-art techniques. Besides, the five algorithms are from different MLAs categories and using them provides a chance to study the effect of the proposed fusion methods on the results of different MLAs categories. RF is an ensemble algorithm while LDA is a dimensionality reduction algorithm. K-NN and SVM are instance-based algorithms while NB is a Bayesian algorithm.

Table 4 shows the four sensor variables possible combinations. HR vectors were not included in the experiments since H and HR can be used alternatively. The case number is represented in the first column, while the second one is the features fused in each case. The time parameters, Week Status

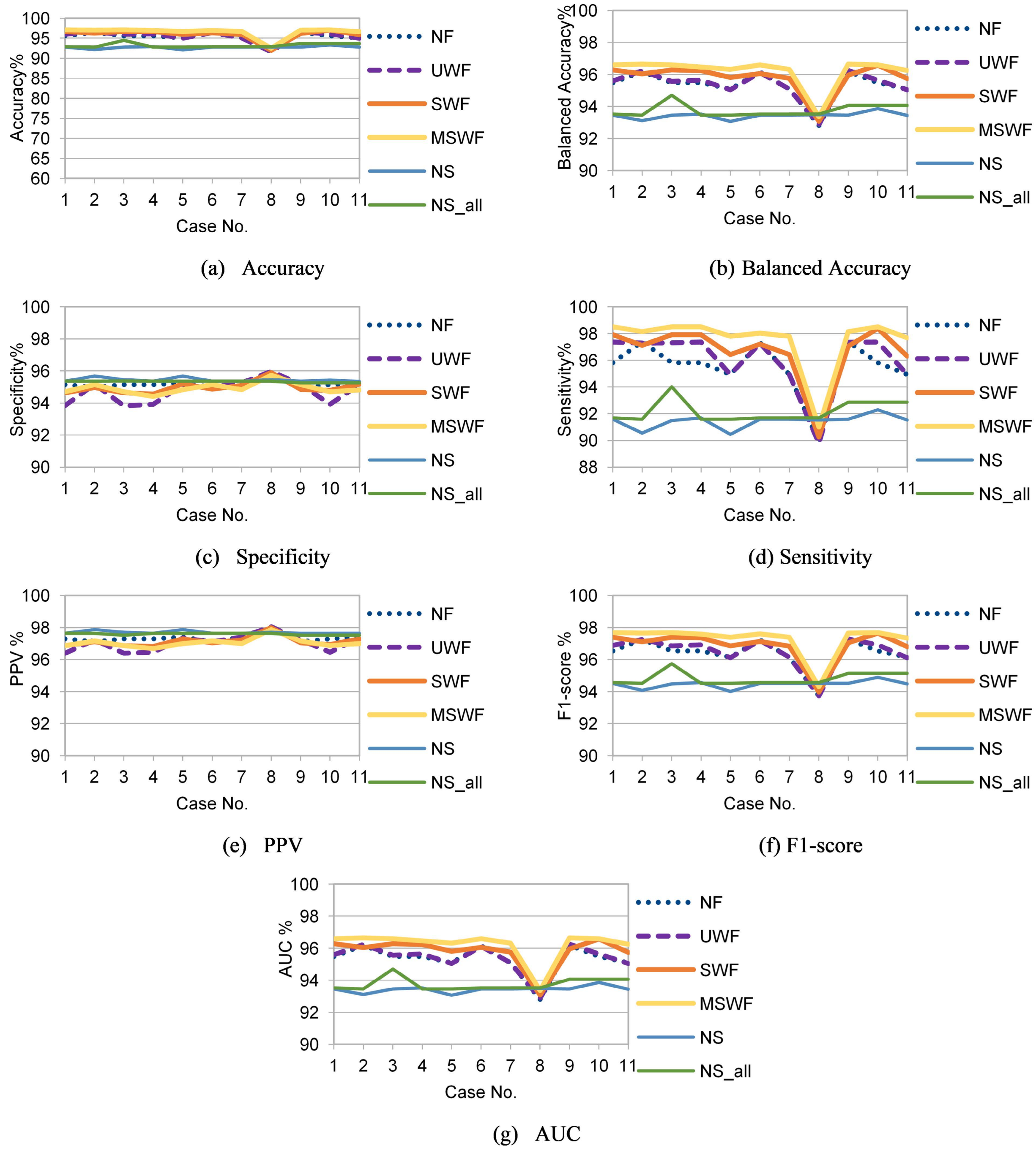


FIGURE 7. **The performance metric of applying K-NN on for closed door case.**

(WS), and Number of Seconds from Midnight (NSM) were used in the experiments since it helps enhance the accuracy of the classification algorithms [29]. The selected window size for producing the weight matrix is four because it provided better accuracy and consumed reasonable time than using 2 or 6 as the window size. Also, a unified seed was set to a random number (1234) before any of the experiments was initiated to ensure that the same results are reproduced when other researchers repeat these experiments. Using multi-sensor-based weight, the original dataset vectors were arranged in an ascending order based on sensor readings range before producing the weight matrix to prevent high sensors readings from canceling the effect of other low sensors readings.

For five cases of the dataset, Figs. 2-6 show the correlation plots. The first case is for the original dataset [29],

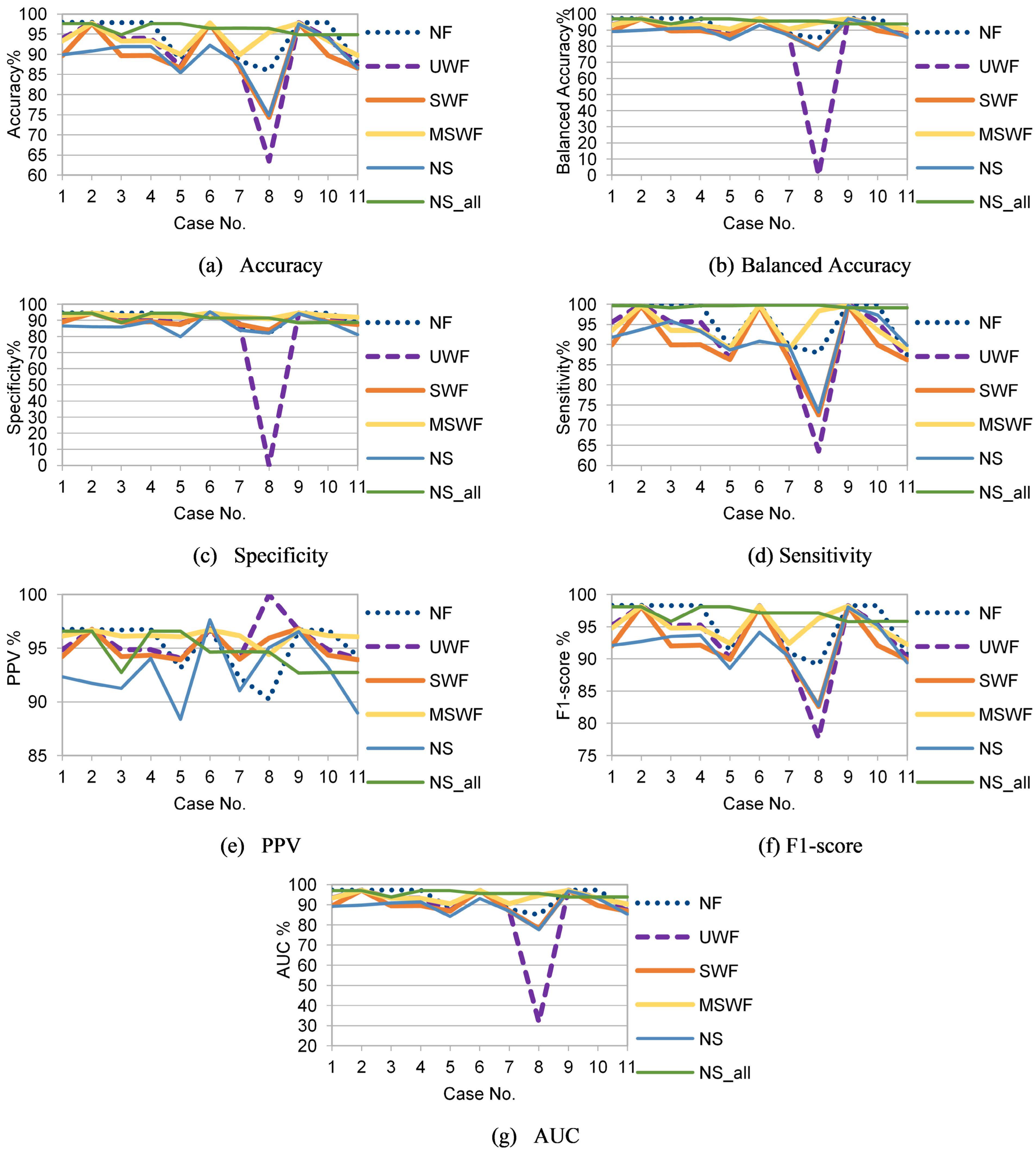


FIGURE 8. The performance metric of applying LDA on for closed door case.

Fig. 2. The second one is for the Sensor-based Weighted (SW) dataset, Fig. 3. The third one is for the Multi-Sensor-based Weighted (MSW) dataset, Fig. 4. The fourth case is for the NS dataset [33], Fig. 5. Finally, the fifth one is for the NS_all dataset [33], Fig. 6. Two points were observed from these correlation plots:

1) original, SW, and NS correlation plots are different for the training and testing data.
2) MSW and NS_all correlation plots are similar for the training and testing data.

Although there are changes in the training and testing data values, only MSW and NS_all correlation plots show a similar correlation. That is because, for MSW, a sensor weight matrix values do not depend only on its measurements but also on the other sensors'. For NS_all, its values are the truth values, which were also calculated using other

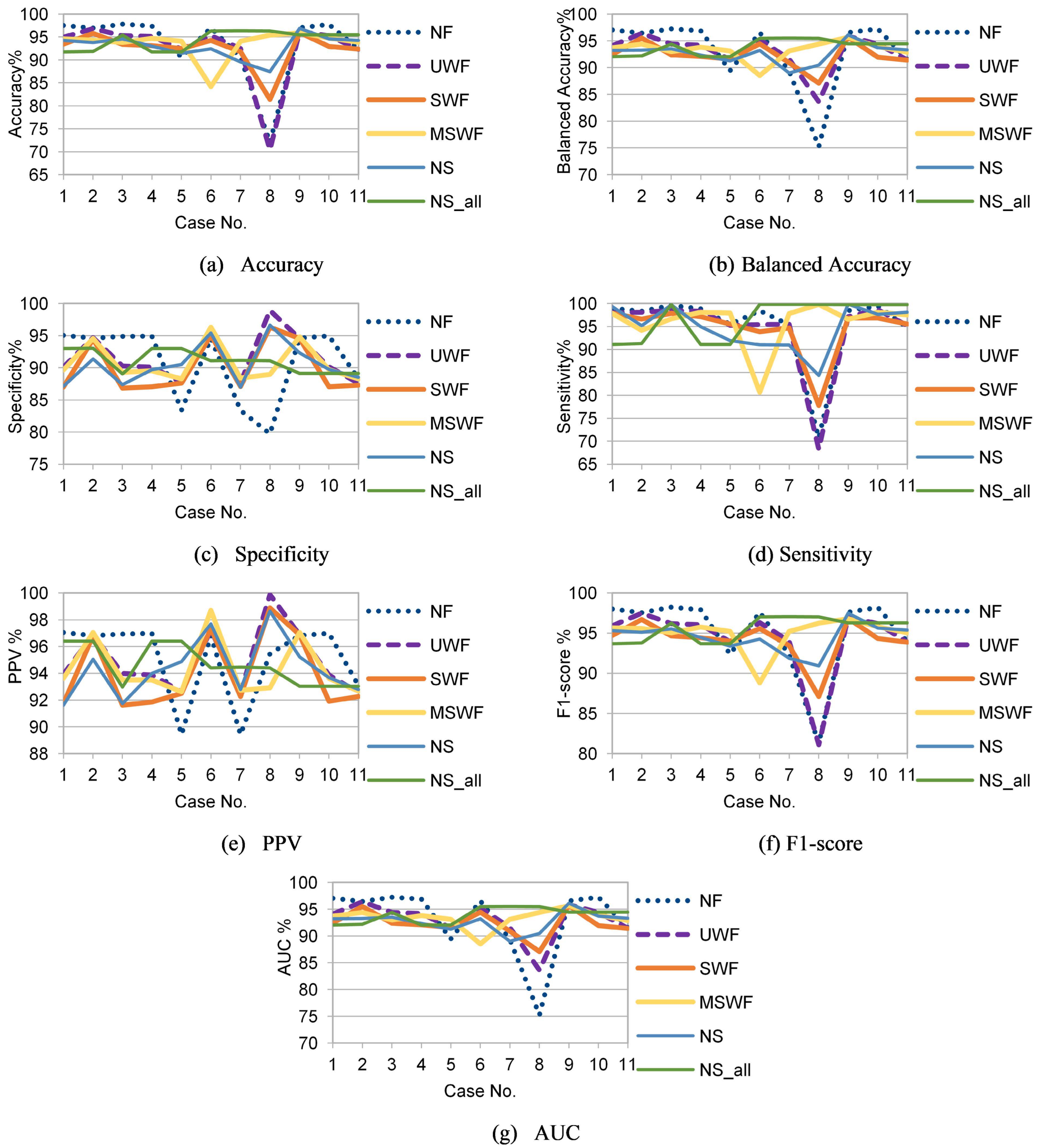
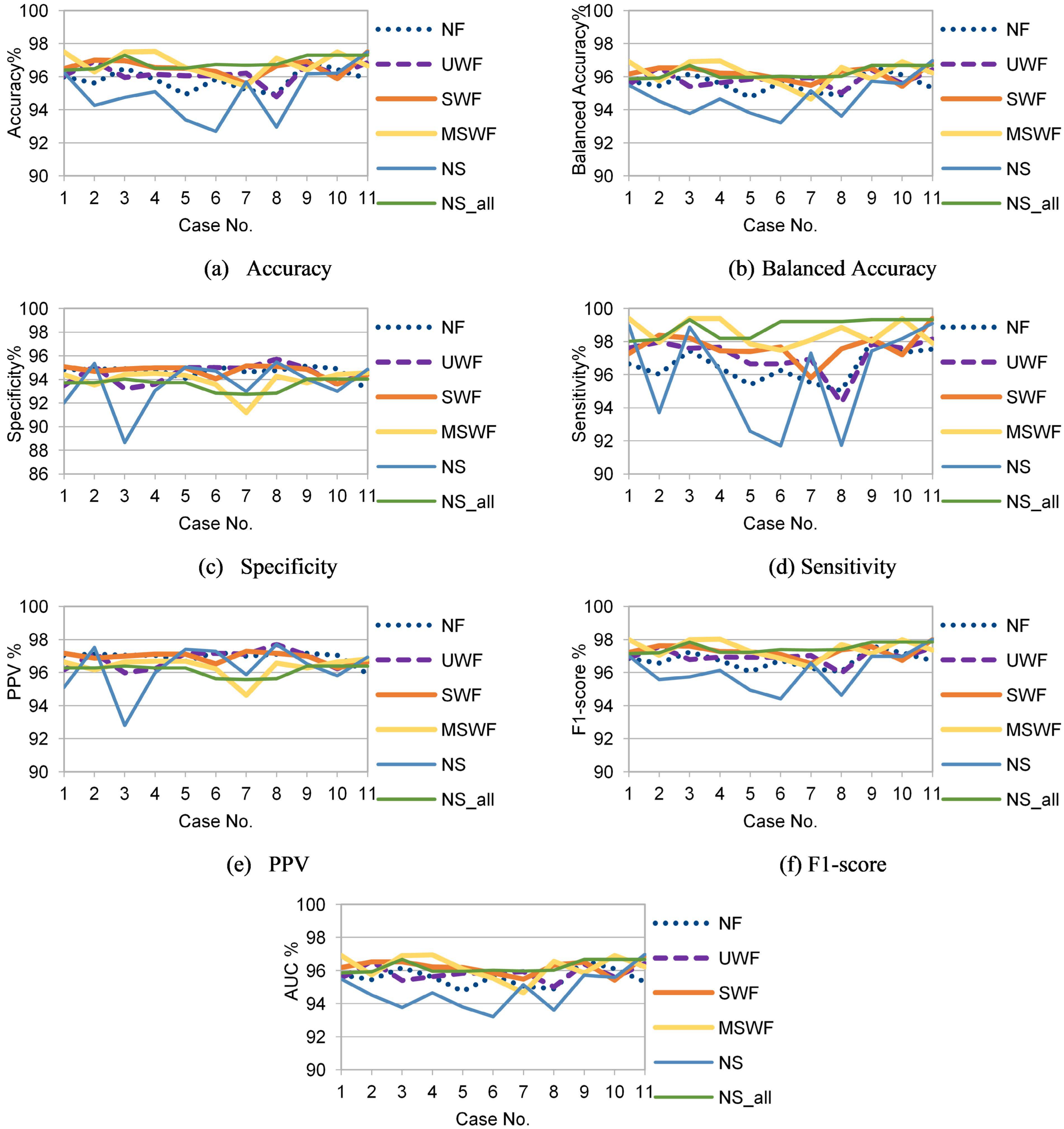


FIGURE 9. **The performance metric of applying NB on for closed door case.**

sensors' measurements. These points help in interpreting NWF results. Also, the dependency of occupancy state on the fused feature was tested using Pearson's product-moment correlation. The result of the test was p-value < 2.2e-16 with confidence level of 95%, so the occupancy state does depend on the fused feature. Since the testing data are for different conditions, the results are divided into three subsections. The first was for the occupancy detection results when most of the measurements were taken with the door closed, Testing1. The second was for testing results of the occupancy detection when most of the measurements were taken with the door opened, Testing2. The third is for comparing the time consumption for different data cases and different classification techniques.



(a) Accuracy (b) Balanced Accuracy (c) Specificity (d) Sensitivity (e) PPV (f) F1-score (g) AUC

FIGURE 10. **The performance metric of applying RF on for closed door case.**

1) TESTING RESULTS FOR THE CLOSED DOOR

Figs. 7-11 present the performance metrics of applying K-NN, LDA, NB, RF, and SVM classification models on the six data cases. These data are generated using the Testing1 dataset, which is similar to the Training dataset, as the readings of both datasets were taken mostly when the door closed. In Fig. 7, using NWF with multi-sensor-based weights (MSWF) and NWF with single-sensor-based weights (SWF) significantly enhanced the performance metrics compared to using the other four data cases in case of applying the K-NN model. SWF and MSWF achieved accuracy up to 96.70 and 97.07 and balanced accuracy up to 96.58 and 96.65, respectively. They also achieved SPE up to 95.89 and 95.73 and SEN up to 98.38 and 98.50, respectively. Besides, PPV up to 97.99 and 97.87 and F1-score up to 97.65 and 97.68 are achieved, respectively.

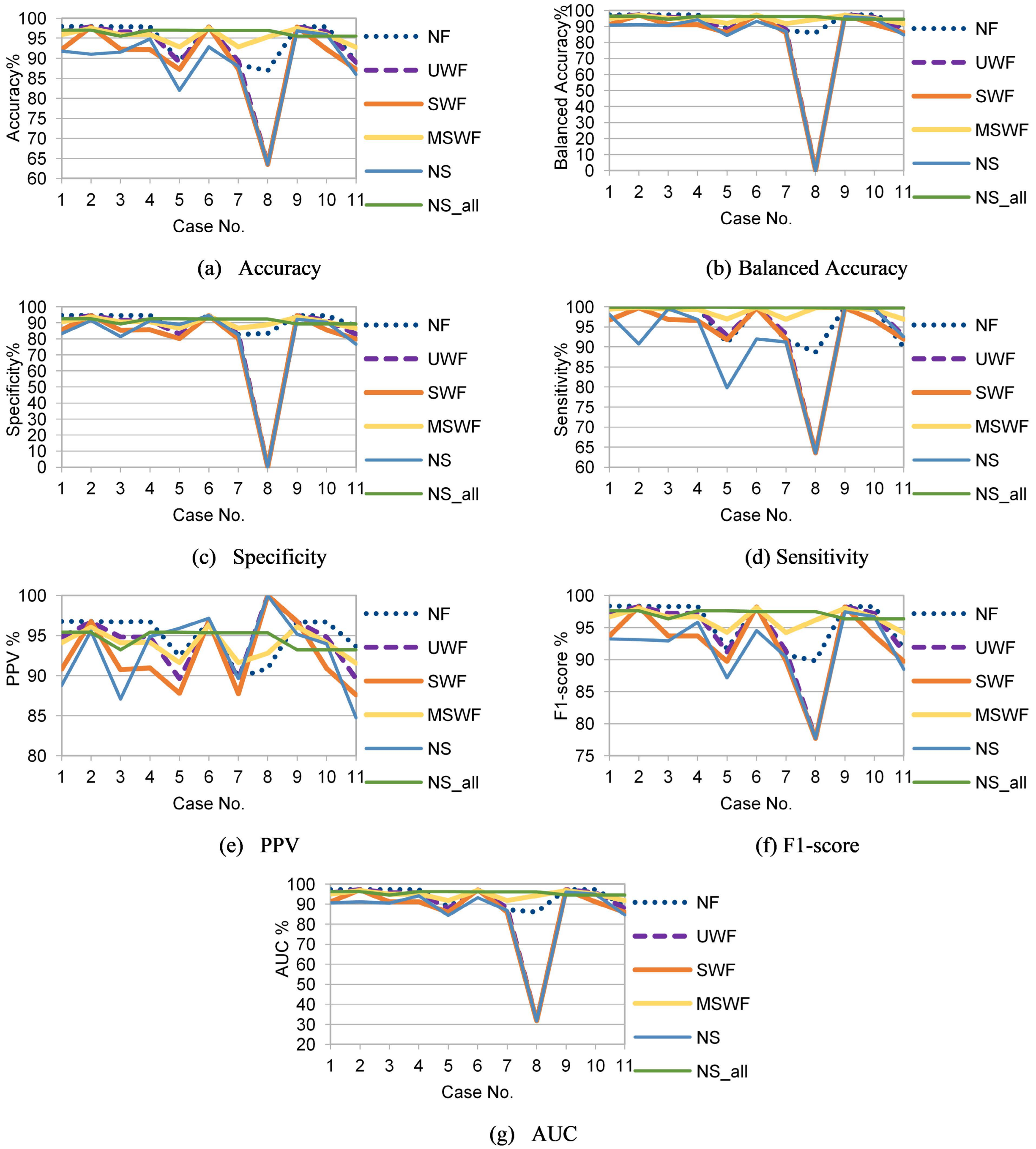


FIGURE 11. The performance metric of applying SVM on for closed door case.

Moreover, AUC up to 96.58 and 96.65 are achieved, respectively.

In the case of applying the LDA model, SWF and MSWF achieved accuracy up to 97.71 and 97.79 and balanced accuracy up to 97.14 and 97.18, respectively (see Fig. 8). They also achieved SPE up to 94.70 and 94.54 and SEN up to 99.57 and 99.82, respectively. PPV up to 96.81 and 96.69 and F1-score up to 98.17 and 98.23 are achieved, respectively. AUC up to 97.14 and 97.18 are achieved, respectively.

According to the results of applying the NB model, using SWF and MSWF achieved accuracy up to 95.98 and 95.91 and balanced accuracy up to 95.68 and 95.67, respectively (see Fig. 9). They also achieved SPE up to 96.30 and 96.30 and SEN up to 97.86 and 99.81, respectively. Besides, PPV up to 97.34 and 98.70 and F1-score up to 96.84 and

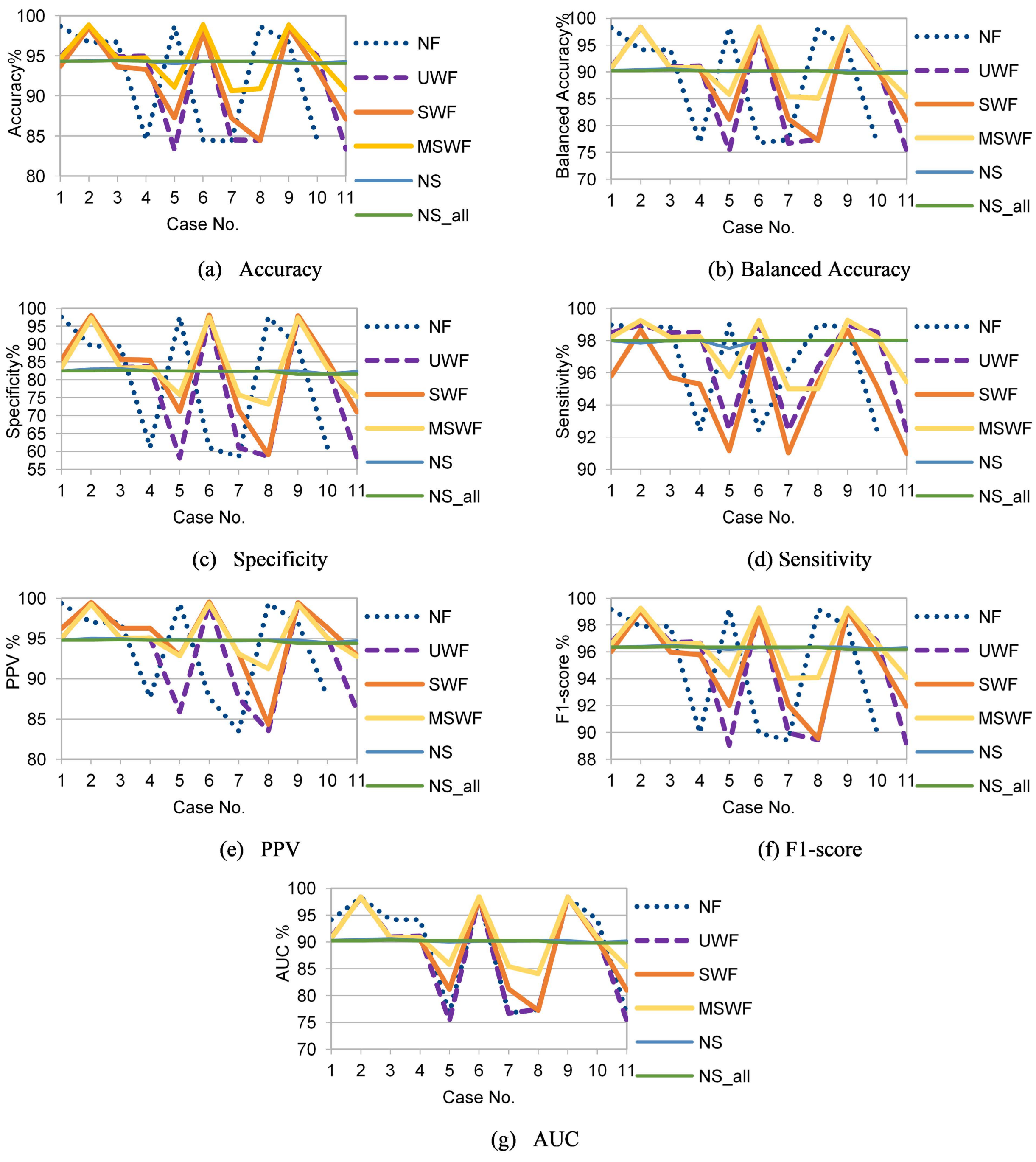
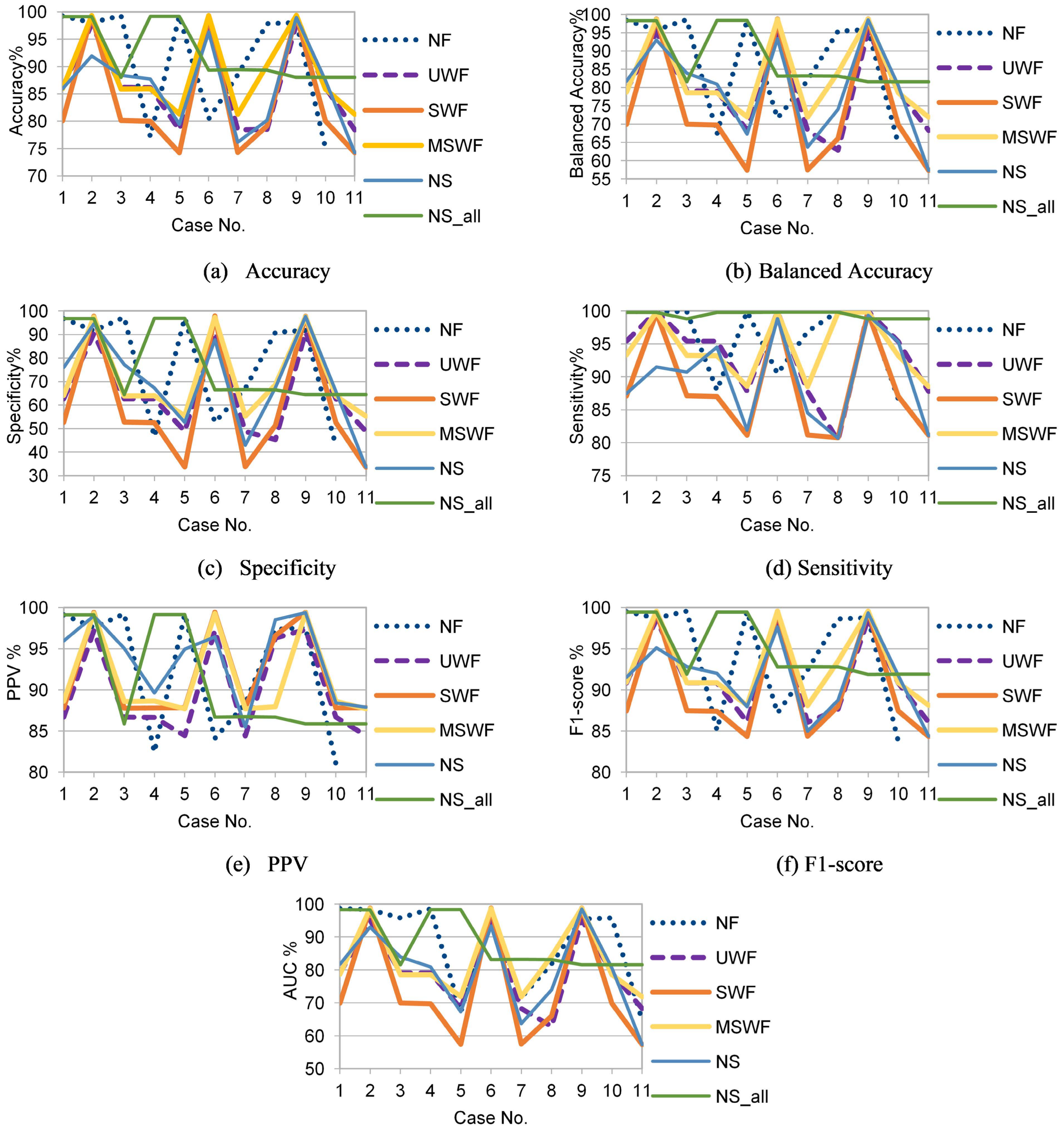


FIGURE 12. **The performance metric of applying K-NN on for opened door case.**

96.79 are achieved, respectively. Moreover, AUC up to 95.68 and 95.67 are achieved, respectively.

As shown in Fig. 10, applying the RF model on MSWF and SWF achieved accuracy up to 97.49 and 97.52 and balanced accuracy up to 96.90 and 96.95, respectively. They also achieved SPE up to 95.14 and 94.55 and SEN up to 99.39 and 99.39, respectively. PPV up to 97.16 and 96.81 and F1-score up to 97.99 and 98.02 are achieved, respectively. Also, AUC up to 96.90 and 96.95 are achieved, respectively.

According to the results of applying the SVM model in Fig. 11, using SWF and MSWF achieved accuracy up to 97.79 and 97.64 and balanced accuracy up to 97.19 and 96.97, respectively. They also achieved SPE up to 94.62 and 94.00 and SEN up to 99.76 and 99.94, respectively. Besides, PPV up to 100.00 and 96.34 and F1-score up to 98.23



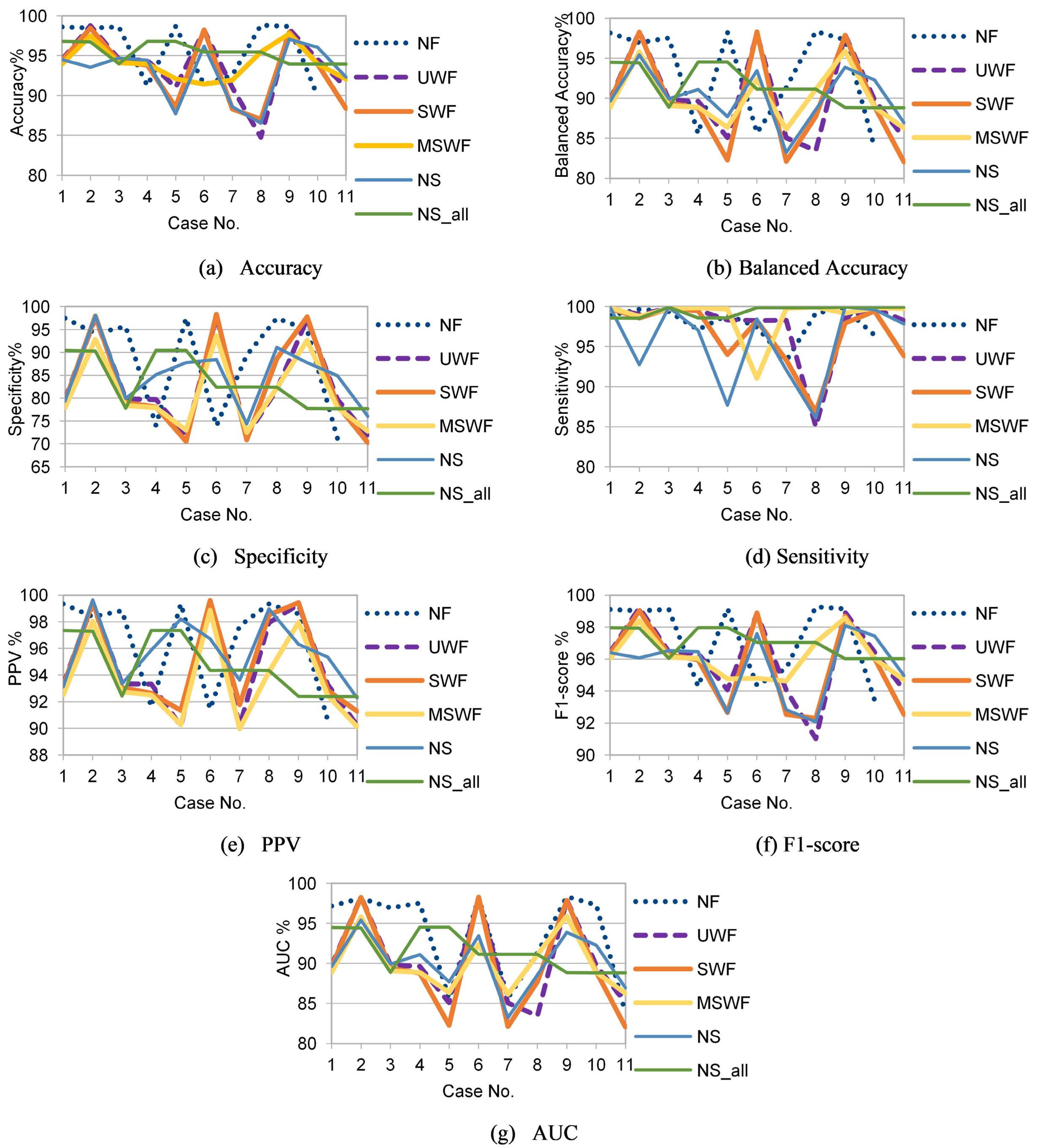
(g) AUC

FIGURE 13. The performance metric of applying LDA on for opened door case.

and 98.11 are achieved, respectively. Moreover, AUC up to 97.19 and 96.97 are achieved, respectively.

2) TESTING RESULTS FOR THE OPENED DOOR

The performance metrics of applying K-NN, LDA, NB, RF, and SVM classification models on data generated from the Testing2 dataset are presented in Figs. 12-16. Testing2 measurements were taken mostly when the door closed. These measurements are quite different from the training dataset, which were taken mostly when the door closed. In Fig. 12, applying the K-NN model on SWF and MSWF achieved accuracy up to 98.53 and 98.87 and balanced accuracy up to 98.35 and 98.37, respectively. They also achieved SPE up to 98.12 and 97.50 and SEN up to 98.66 and 99.26, respectively. Besides, PPV up to 99.53 and 99.34 and



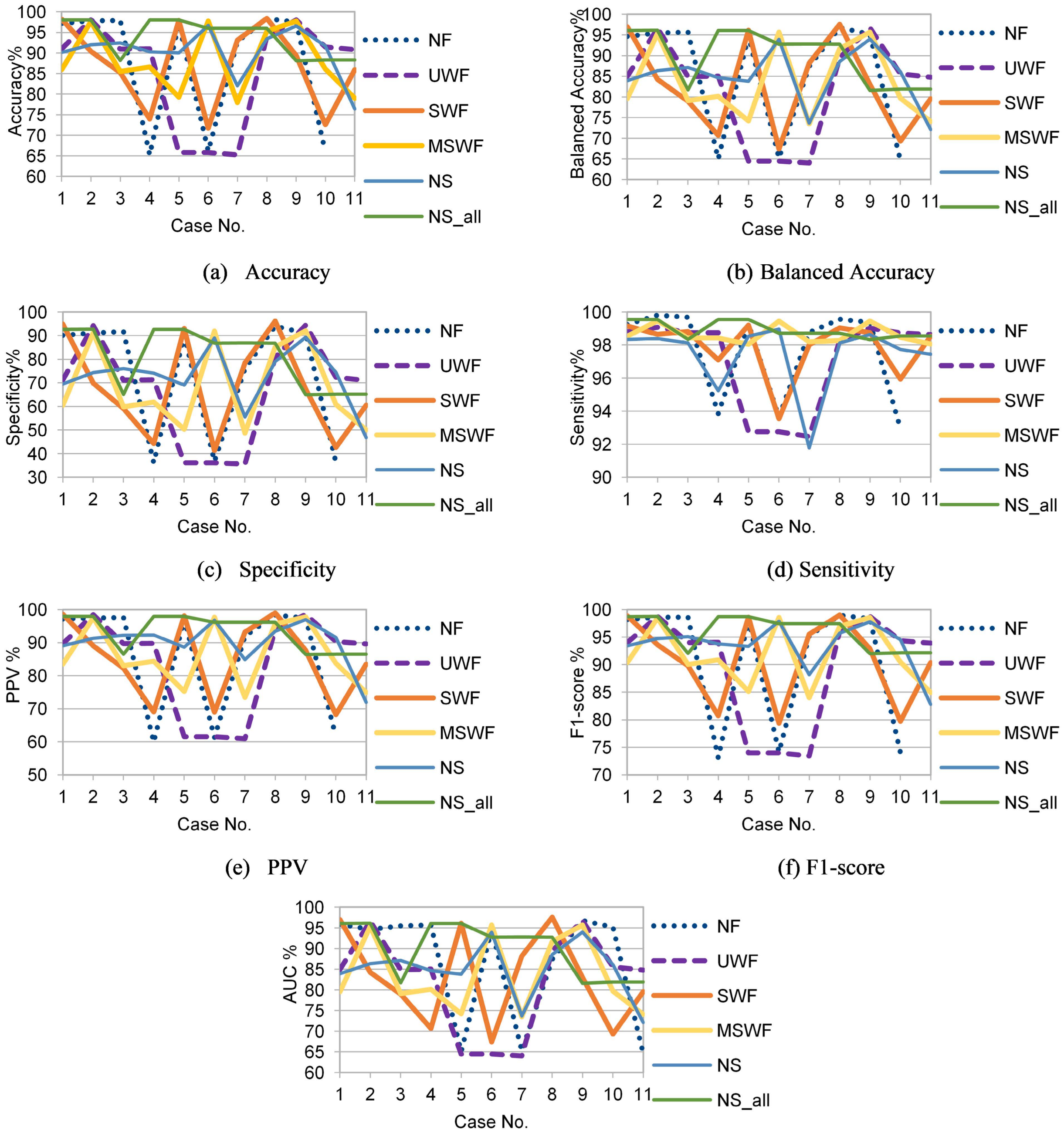
(a) Accuracy (b) Balanced Accuracy (c) Specificity (d) Sensitivity (e) PPV (f) F1-score (g) AUC

FIGURE 14. The performance metric of applying NB on for opened door case.

F1-score up to 99.08 and 99.29 are achieved, respectively. Moreover, AUC up to 98.35 and 98.37 are achieved, respectively.

In the case of applying the LDA model, Fig. 13, the SWF and MSWF achieved accuracy up to 99.31 and 99.35 and balanced accuracy up to 98.80 and 98.70, respectively. They also achieved SPE up to 97.92 and 97.56 and SEN up to 99.71 and 99.84, respectively. PPV up to 99.44 and 99.34 and F1-score up to 99.56 and 99.59 are achieved, respectively. AUC up to 98.80 and 98.70 are achieved, respectively.

According to applying the NB model in Fig. 14, using SWF and MSWF achieved accuracy up to 98.45 and 97.74 and balanced accuracy up to 98.29 and 95.91, respectively. They also achieved SPE up to 98.35 and 93.66 and SEN up to 99.75 and 99.97, respectively. Besides, PPV up to 99.58 and 98.86 and



(a) Accuracy (b) Balanced Accuracy

(c) Specificity (d) Sensitivity

(e) PPV (f) F1-score

(g) AUC

FIGURE 15. The performance metric of applying RF on for opened door case.

F1-score up to 99.02 and 98.56 are achieved, respectively. Moreover, AUC up to 98.29 and 95.91 are achieved, respectively.

In Fig. 15, applying the RF model on MSWF and SWF achieved accuracy up to 98.42 and 97.79 and balanced accuracy up to 97.57 and 95.72, respectively. They also achieved SPE up to 96.11 and 91.98 and SEN up to 99.21 and 99.47, respectively. PPV up to 98.96 and 97.73 and F1-score up to 99.00 and 98.59 are achieved, respectively. Also, AUC up to 97.57 and 95.72 are achieved, respectively.

According to the results of applying the SVM model in Fig. 16, using SWF and MSWF achieved accuracy up to 99.39 and 98.20 and balanced accuracy up to 98.84 and 96.12, respectively. They also achieved SPE up to 97.88 and 92.32 and SEN up to 99.80 and 99.92, respectively. Besides, PPV up to 100.00 and 97.79 and F1-score up to 99.62 and

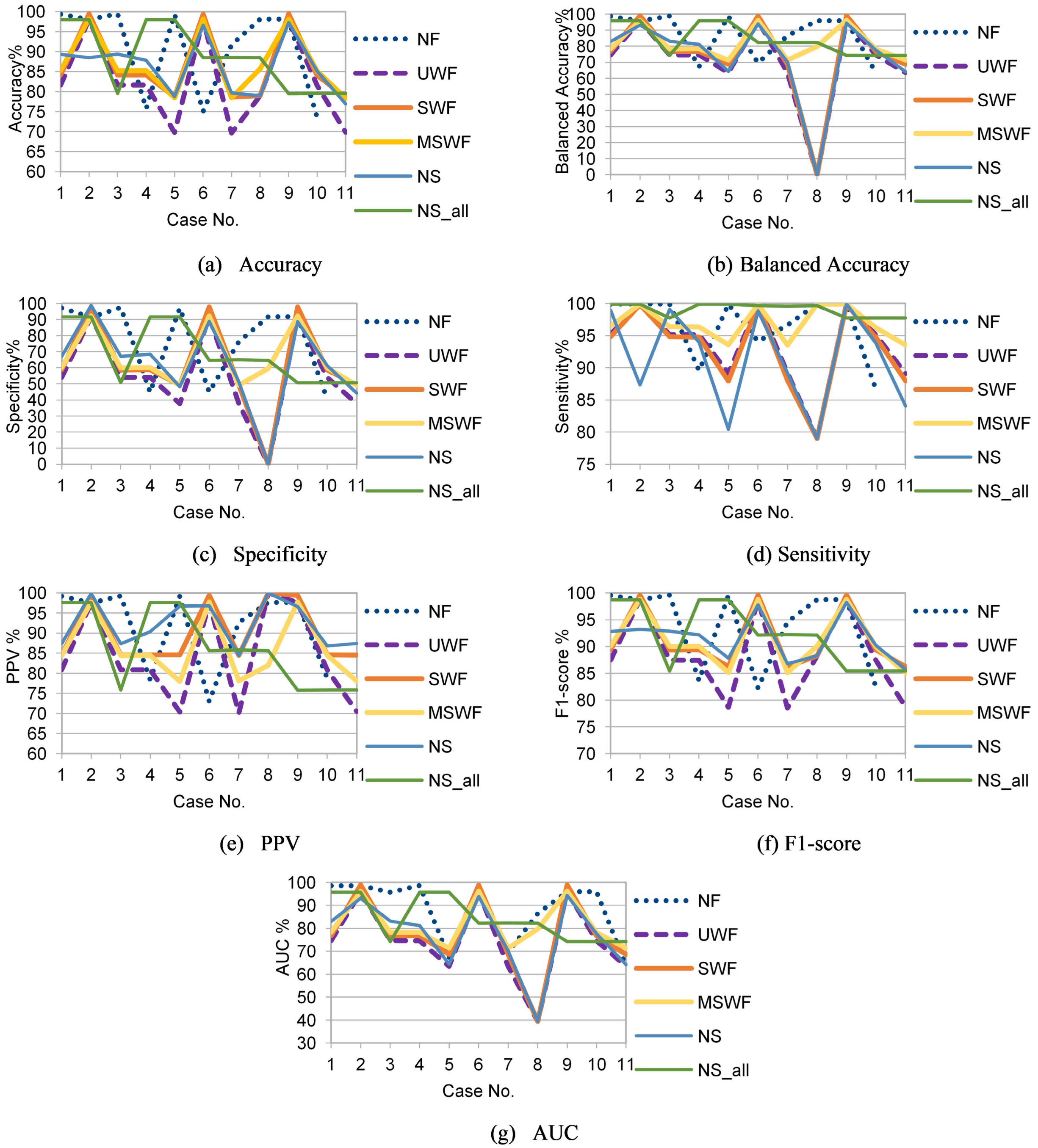


FIGURE 16. The performance metric of applying SVM on for opened door case.

98.85 are achieved, respectively. Moreover, AUC up to 98.84 and 96.12 are achieved, respectively.

3) TIME CONSUMPTION COMPARISON

Time consumption is an essential metric for evaluating and comparing the proposed technique performance. As shown in Fig. 17, the proposed technique methods (SWF and MSWF) provided an acceptable time consumption.

For all models except LDA, using MSWF consumed the least time in the worst case. For LDA, it consumed less than one additional second compared to NF, UWF, and NS, but it is less than NS_all by about 163.8 seconds. Using SWF with the K-NN model, it consumed less time than NF, NS, and NS_all, but it consumed less than one additional second than UWF and MSWF. For LDA, it consumed less time than MSWF, NS, and NS_all, but it consumed less than one additional

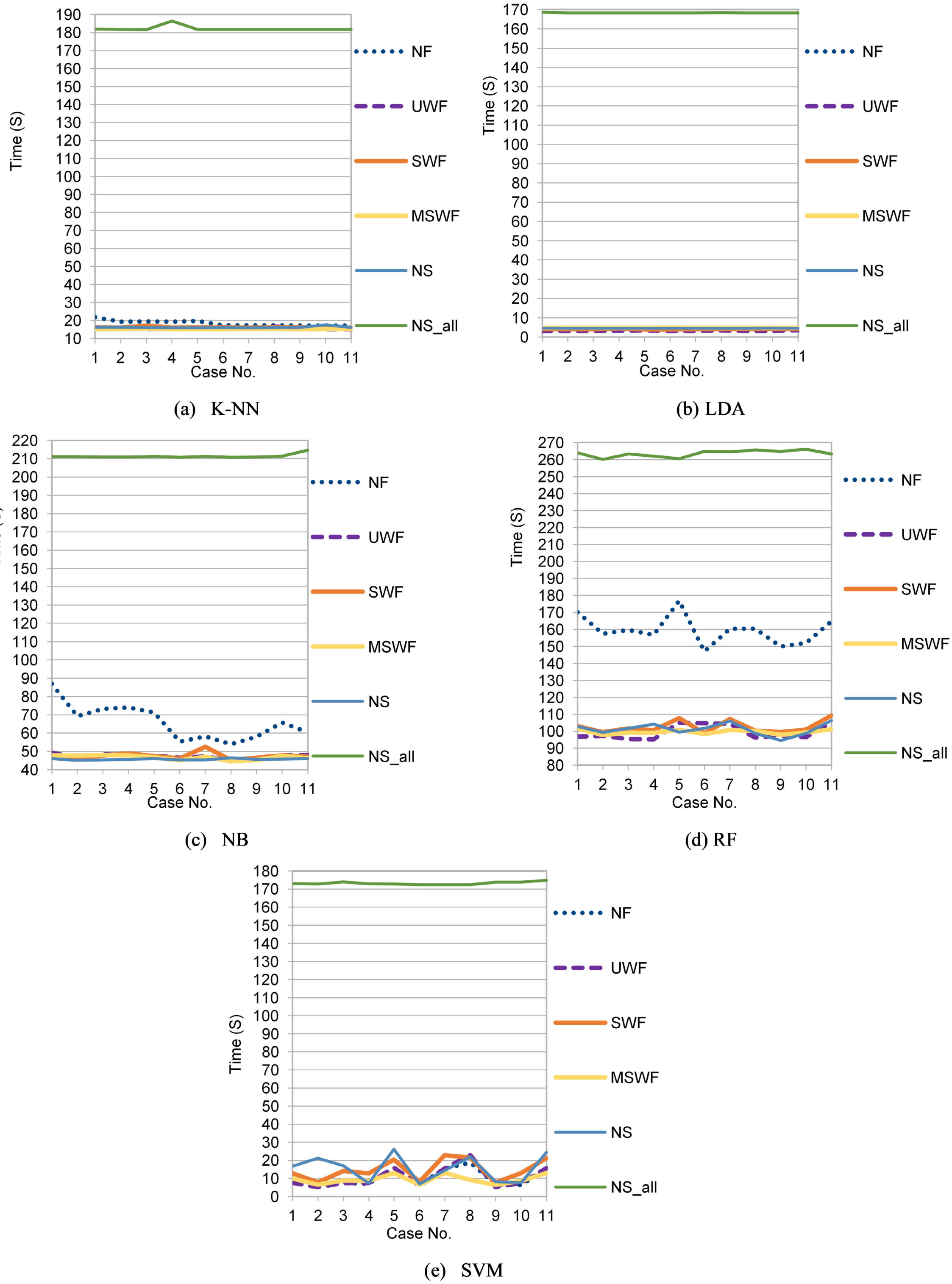


FIGURE 17. The time consumption applying various classification models on the six data cases.

second compared to NF and UWF. For NB, it consumed less time than NF, UWF, and NS_all, but it consumed less than one and two additional seconds compared to MSWF and NS, respectively. According to the RF results, it consumed less time than NF and NS_all, but it consumed less than 2, 3, and 7 additional seconds compared to NS, UWF, and MSWF,

TABLE 5. Balanced accuracy ranges summary.

	NF		UWF		SWF		NS		MWF		NS_all	
	Worst	Best	Worst	Best	Worst	Best	Worst	Best	Worst	Best	Worst	Best
K-NN	76.73	98.30	75.29	98.28	77.24	98.35	89.86	93.86	85.13	98.37	89.75	94.70
LDA	64.64	98.64	NA-62.86	97.24	57.47	98.80	57.34	98.51	71.88	98.70	81.60	98.32
NB	75.23	98.35	83.45	98.32	82.06	98.29	83.24	96.08	86.15	95.91	88.82	95.48
RF	64.63	96.78	64.45	96.72	67.41	97.57	72.14	96.97	73.47	96.95	81.61	96.68
SVM	63.54	98.67	NA-63.45	97.24	NA-68.65	98.84	NA-64.20	96.04	71.32	96.97	74.22	96.25

TABLE 6. A comparison between NF and MSWF worst cases based on their performance metrics.

	KNN		LDA		NB		RF		SVM	
	NF	MSWF	NF	MSWF	NF	MSWF	NF	MSWF	NF	MSWF
ACC	84.36	90.66	75.13	81.28	72.01	84.17	65.22	77.96	73.58	78.38
Balanced ACC	76.73	85.13	64.64	71.88	75.23	86.15	64.63	73.47	63.54	71.32
SPE	58.53	73.17	42.63	55.26	71.17	72.48	36.12	48.75	40.44	49.11
SEN	89.86	90.99	86.64	88.51	70.72	80.69	92.96	97.49	86.64	93.53
PPV	83.46	91.25	81.01	87.68	89.43	89.96	59.89	73.45	73.00	78.02
F1-score	89.40	94.03	83.73	88.09	81.25	88.79	73.12	84.04	82.23	85.08
AUC	76.73	84.08	64.64	71.88	75.23	86.15	64.63	73.47	63.54	71.32
Time	21.85	15.51	3.90	4.85	86.89	47.99	170.13	101.61	18.68	13.05

respectively. Also related to SVM results, it consumed less time than UWF, NS, and NS_all, but it consumed around 4 and 10 additional seconds more than NF and MSWF, respectively.

E. DISCUSSION

In this section, the results presented in the previous section and summarized in Table 5 are discussed. Using MSWF greatly enhanced the accuracy range compared to using NF, UWF, SWF, or NS. Although using NS_all achieved better accuracy than MSWF for the worst cases (the cases from the eleven cases that achieved lowest accuracy), using MSWF was more efficient than using NS_all regarding time consumption. Using NS_all achieved better accuracy than MSWF for the worst cases because in Testing 2 most of the measurements were taken with the door opened while most of the Training measurements were taken with the door closed. Hence, using a dynamic fusion equation (NS and NS_all) is preferable when the place environment is dynamic, but using a static fusion equation (SWF and MWF) is preferable when the place environment is not very dynamic.

The good results of MSWF are logical since only MSW has similar correlation plots for the training and testing sets, which means a stable correlation among the four variables. Hence, using MSWF improved the ranges of the eleven cases accuracy, notably the lower bound. Although the five algorithms (K-NN, LDA, NB, RF, and SVM) are from different MLAs categories, MSWF had the same positive effect on the accuracy results of them with different detection accuracy ranges. So, the proposed technique is not biased to a specific MLAs category. On applying K-NN, using MSWF raised the minimum bound of balanced accuracy from 76.73 to 85.13, while applying LDA raised it from 64.64 to 71.88. Applying NB raised the minimum bound from 75.23 to 86.15 while applying RF raised it from 64.63 to 73.47. As for applying SVM, it was raised from 63.54 to 71.32.

MWF provided an acceptable accuracy percentage using a static fusion equation, which requires zero time for production. On the other hand, NS_all provided higher accuracy; however, the dynamic equation's production has O(n) time complexity. Besides, the fusion equation of NWF, using either SWF or MSWF, does not require to be altered according to changes made in the number or the type of the sensors. Consequently, it does not limit the system's scalability. Contrariwise, the dynamic equation (NS and NS_all), which is based on the sensors' correlation, requires reproduction.

Using SWF and NS, they provided lower improvement in detection accuracy for K-NN, NB, and RF. However, for LDA, they showed degraded accuracy for cases 5, 7, and 11. As mentioned in Table 2, these cases include using CO_2 without Light, and the correlation between CO_2 and Light differed significantly between Training and Testing2 datasets (Figures 3 and 5). The worst case is 11, where CO_2 was used with humidity as the correlation between CO_2 and Humidity differed extremely between Training and Testing2 datasets. This problem did not appear for MSWF and NS_all because they have similar correlation plots of the training and testing sets. Moreover, LDA is a parametric machine learning algorithm. The data characteristics' changes affected the LDA estimated parameters (mean and covariance) calculated in the training phase. Hence, these issues affected the accuracy of detection.

Using UWF did not enhance detection accuracy, and the worst case is case 8 (Temperature and Humidity) for the same reasons. In this case, the SPE is NA (Not Available) and is replaced by zero for plotting, Figure 8-c. SPE is NA when a classifier cannot recognize negative class patterns and, in turn, Balanced ACC, which is the average of SPE and SEN, is NA [65]. The Balanced accuracy of 62.86 is the lowest accuracy, excluding case 8. In the case of SVM, UWF, SWF, and NS have the same problem in case 8. SVM could not recognize the negative class patterns.

According to the previous discussion and Table 6, which compares performance metrics between NF and MSWF, it could be concluded that using MSWF is a good compromise between high accuracy and low time consumption. Hence, using MSWF can provide an efficient occupancy detection system.

V. CONCLUSION

Environmental sensor-based occupancy detection systems are beneficial in many essential applications because sensor data processing requires fewer processing capabilities and a smaller storage size. Besides, it maintains the privacy of individuals. However, uncertainty and unreliability are the main problems of using environmental sensor data because sensor data tends to be incomplete and noisy. NWF, the proposed technique, is a linear weighted fusion technique based on neutrosophy. By using neutrosophy, the proposed method handled the sensor's data uncertainty. As a result, occupancy detection accuracy is enhanced. Also, using multiple types of sensors increased the reliability by using a variety of sensor types. Moreover, it minimizes the detection time by using only one feature for training and testing, which saves some energy. Additionally, using a predefined fusion equation instead of a dynamic one consumes no time to produce the equation. The equation is also not limited to a specific number or type of features, which does not limit the system scalability. Accordingly, the experimental results proved enhancement in accuracy ranges and time consumption using NWF. Thus, using NWF makes the occupancy detection system more efficient. For future works, more investigation is required for the proposed technique effect on multiclass classification problems such as occupancy estimation. In occupancy estimation, classes represent the number of occupants or occupancy level. Also, the applicability of using the proposed technique to fuse more heterogeneous resources such as images, audios, and videos is suggested for occupancy detection and other different applications.

REFERENCES

[1] R. F. R. Suleiman and M. I. Nebil, "Implementation of statistical learning model for room occupancy detection," *Eur. J. Mol. & Clin. Med.*, vol. 7, no. 8, pp. 3737–3746, 2021.

[2] Y. Li, J.-L. Wang, Z.-H. Tian, T.-B. Lu, and C. Young, "Building lightweight intrusion detection system using wrapper-based feature selection mechanisms," *Comput. Secur.*, vol. 28, no. 6, pp. 466–475, 2009.

[3] P. E. Lopez-de-Teruel, F. J. Garcia, and O. Canovas, "Practical passive localization system based on wireless signals for fast deployment of occupancy services," *Future Gener. Comput. Syst.*, vol. 107, pp. 692–704, Jun. 2020.

[4] M. Monzón-Chavarrías, S. Guillén-Lambea, S. García-Pérez, A. L. Montealegre-Gracia, and J. Sierra-Pérez, "Heating energy consumption and environmental implications due to the change in daily habits in residential buildings derived from COVID-19 crisis: The case of barcelona, Spain," *Sustainability*, vol. 13, no. 2, p. 918, Jan. 2021.

[5] N. A. Azmi, M. Arıcı, and A. Baharun, "A review on the factors influencing energy efficiency of mosque buildings," *J. Cleaner Prod.*, vol. 292, Apr. 2021, Art. no. 126010.

[6] X. Yang, L. Gao, J. Zheng, and W. Wei, "Location privacy preservation mechanism for location-based service with incomplete location data," *IEEE Access*, vol. 8, pp. 95843–95854, 2020.

[7] P. Kraipeerapun and S. Amornsamankul, "Room occupancy detection using modified stacking," in *Proc. 9th Int. Conf. Mach. Learn. Comput.*, Feb. 2017, pp. 162–166.

[8] M. Frontczak, S. Schiavon, J. Goins, E. Arens, H. Zhang, and P. Wargocki, "Quantitative relationships between occupant satisfaction and satisfaction aspects of indoor environmental quality and building design," *Indoor Air*, vol. 22, pp. 119–131, Sep. 2011.

[9] L. M. Candanedo, V. Feldheim, and D. Deramaix, "A methodology based on hidden Markov models for occupancy detection and a case study in a low energy residential building," *Energy Buildings*, vol. 148, pp. 327–341, Aug. 2017.

[10] S. Fong, C. Bhatt, D. Korzun, S. H. Yang, and L. Yang, "Internet of breath (IoB): Integrative indoor gas sensor applications for emergency control and occupancy detection," in *Lecture Notes in Real-Time Intelligent Systems* (Advances in Intelligent Systems and Computing), vol. 756, J. Mizera-Pietraszko, P. Pichappan, and L. Mohamed, Eds. Cham, Switzerland: Springer, 2019, doi: 10.1007/978-3-319-91337-7_32.

[11] P. W. Tien, S. Wei, and J. Calautit, "A computer vision-based occupancy and equipment usage detection approach for reducing building energy demand," *Energies*, vol. 14, no. 1, p. 156, Dec. 2020.

[12] D. Liu, X. Guan, Y. Du, and Q. Zhao, "Measuring indoor occupancy in intelligent buildings using the fusion of vision sensors," *Meas. Sci. Technol.*, vol. 24, no. 7, Jul. 2013, Art. no. 074023.

[13] J. Zou, Q. Zhao, W. Yang, and F. Wang, "Occupancy detection in the office by analyzing surveillance videos and its application to building energy conservation," *Energy Buildings*, vol. 152, pp. 385–398, Oct. 2017.

[14] A. B. Khalifa, I. Alouani, M. A. Mahjoub, and A. Rivenq, "A novel multi-view pedestrian detection database for collaborative intelligent transportation systems," *Future Gener. Comput. Syst.*, vol. 113, pp. 506–527, Dec. 2020.

[15] Z. Chen, M. K. Masood, and Y. C. Soh, "A fusion framework for occupancy estimation in office buildings based on environmental sensor data," *Energy Buildings*, vol. 133, pp. 790–798, Dec. 2016.

[16] M. K. Masood, Y. C. Soh, and C. Jiang, "Occupancy estimation from environmental parameters using wrapper and hybrid feature selection," *Appl. Soft Comput.*, vol. 60, pp. 482–494, Nov. 2017.

[17] J. Ahmad, H. Larijani, R. Emmanuel, M. Mannion, and A. Javed, "Occupancy detection in non-residential buildings—A survey and novel privacy preserved occupancy monitoring solution," *Appl. Comput. Informat.*, vol. 17, no. 2, pp. 279–295, Apr. 2021.

[18] R. Frodl and T. Tille, "A high-precision NDIR CO_2 gas sensor for automotive applications," *IEEE Sensors J.*, vol. 6, no. 6, pp. 1697–1705, Dec. 2006.

[19] Y. Jeon, C. Cho, J. Seo, K. Kwon, H. Park, S. Oh, and I.-J. Chung, "IoT-based occupancy detection system in indoor residential environments," *Buildings Environ.*, vol. 132, pp. 181–204, Mar. 2018.

[20] E. Soltanaghaei and K. Whitehouse, "Practical occupancy detection for programmable and smart thermostats," *Appl. Energy*, vol. 220, pp. 842–855, Jun. 2018.

[21] H. Elkhoukhi, Y. NaitMalek, A. Berouine, M. Bakhouya, D. Elouadghiri, and M. Essaaidi, "Towards a real-time occupancy detection approach for smart buildings," *Proc. Comput. Sci.*, vol. 134, pp. 114–120, Jan. 2018.

[22] A. Pratama, W. Widyawan, A. Lazovik, and M. Aiello, "Multi-user low intrusive occupancy detection," *Sensors*, vol. 18, no. 3, p. 796, Mar. 2018.

[23] B. W. Hobson, D. Lowcay, H. B. Gunay, A. Ashouri, and G. R. Newsham, "Opportunistic occupancy-count estimation using sensor fusion: A case study," *Building Environ.*, vol. 159, Jul. 2019, Art. no. 106154.

[24] L. Rueda, K. Agbossou, A. Cardenas, N. Henao, and S. Kelouwani, "A comprehensive review of approaches to building occupancy detection," *Building Environ.*, vol. 180, Aug. 2020, Art. no. 106966.

[25] J. L. G. Ortega, L. Han, N. Whittacker, and N. Bowring, "A machine-learning based approach to model user occupancy and activity patterns for energy saving in buildings," in *Proc. Sci. Inf. Conf. (SAI)*, Jul. 2015, pp. 474–482.
[26] T. Ekwevugbe, N. Brown, V. Pakka, and D. Fan, "Real-time building occupancy sensing using neural-network based sensor network," in *Proc. 7th IEEE Int. Conf. Digit. Ecosyst. Technol. (DEST)*, Jul. 2013, pp. 114–119.
[27] K. P. Lam, M. Höynck, B. Dong, B. Andrews, Y.-S. Chiou, R. Zhang, and J. Choi, "Occupancy detection through an extensive environmental sensor network in an open-plan office building," *IBPSA Building Simul.*, vol. 145, pp. 1452–1459, Jul. 2009.
[28] B. Dong, B. Andrews, K. P. Lam, M. Höynck, R. Zhang, Y.-S. Chiou, and D. Benitez, "An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network," *Energy Buildings*, vol. 42, no. 7, pp. 1038–1046, Jul. 2010.
[29] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models," *Energy Buildings*, vol. 112, pp. 28–39, Jan. 2016.
[30] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: A survey," *Multimedia Syst.*, vol. 16, no. 6, pp. 345–379, 2010.
[31] J. Chaney, E. H. Owens, and A. D. Peacock, "An evidence based approach to determining residential occupancy and its role in demand response management," *Energy Buildings*, vol. 125, pp. 254–266, Aug. 2016.
[32] P. Christodoulou, A. Christoforou, and A. S. Andreou, "A hybrid prediction model integrating fuzzy cognitive maps with support vector machines," in *Proc. 19th Int. Conf. Enterprise Inf. Syst.*, 2017, pp. 554–564.
[33] N. S. Fayed, M. Abu-Elkheir, E. M. El-Daydamony, and A. Atwan, "Sensor-based occupancy detection using neutrosophic features fusion," *Heliyon*, vol. 5, no. 9, Sep. 2019, Art. no. e02450.
[34] A. Abdelgawad and M. Bayoumi, "Data fusion in WSN," in *Resource-Aware Data Fusion Algorithms for Wireless Sensor Networks* (Lecture Notes in Electrical Engineering), vol. 118. Boston, MA, USA: Springer, 2012, doi: 10.1007/978-1-4614-1350-9_2.
[35] E. Hailemariam, R. Goldstein, R. Attar, and A. Khan, "Real-time occupancy detection using decision trees with multiple sensor types," in *Proc. Symp. Simulation Archit. Urban Design*, 2011, pp. 141–148.
[36] Z. Yang, N. Li, B. Becerik-Gerber, and M. Orosz, "A multi-sensor based occupancy estimation model for supporting demand driven HVAC operations," in *Proc. Symp. Simulation Archit. Urban Design*, 2012, pp. 1–8.
[37] T. Ekwevugbe, N. Brown, and D. Fan, "A design model for building occupancy detection using sensor fusion," in *Proc. 6th IEEE Int. Conf. Digit. Ecosystems Technol. (DEST)*, Jun. 2012, pp. 1–6.
[38] Z. Yang, N. Li, B. Becerik-Gerber, and M. Orosz, "A systematic approach to occupancy modeling in ambient sensor-rich buildings," *Simulation*, vol. 90, no. 8, pp. 960–977, Aug. 2014.
[39] B. Ai, Z. Fan, and R. X. Gao, "Occupancy estimation for smart buildings by an auto-regressive hidden Markov model," in *Proc. Amer. Control Conf.*, Jun. 2014, pp. 2234–2239.
[40] Q. Hua, H.-B. Chen, Y.-Y. Ye, and S. X.-D. Tan, "Occupancy detection in smart buildings using support vector regression method," in *Proc. 8th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (IHMSC)*, Aug. 2016, pp. 77–80.
[41] K. Tutuncu, O. Cataltas, and M. Koklu, "Occupancy detection through light, temperature, humidity and CO_2 sensors using ANN," *Int. J. Ind. Electron. Elect. Eng.*, vol. 5, no. 2, pp. 63–67, Feb. 2017.
[42] D. Alghamdi, "Occupancy detection: A data mining approach," *Int. J. Sci. Eng. Res.*, vol. 7, no. 6, pp. 168–172, Jun. 2016.
[43] T. H. Pedersen, K. U. Nielsen, and S. Petersen, "Method for room occupancy detection based on trajectory of indoor climate sensor data," *Building Environ.*, vol. 115, pp. 147–156, Apr. 2017.
[44] S. H. Kim and H. J. Moon, "Case study of an advanced integrated comfort control algorithm with cooling, ventilation, and humidification systems based on occupancy status," *Building Environ.*, vol. 133, pp. 246–264, Apr. 2018.
[45] M. Koklu and K. Tutuncu, "Tree based classification methods for occupancy detection," in *Proc. IOP Conf. Ser., Mater. Sci. Eng.*, vol. 675, 2019, Art. no. 012032.
[46] H. Elkhoukhi, Y. NaitMalek, M. Bakhouya, A. Berouine, A. Kharbouch, F. Lachhab, M. Hanifi, D. El Ouadghiri, and M. Essaaidi, "A platform architecture for occupancy detection using stream processing and machine learning approaches," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 17, Sep. 2020, Art. no. e5651.
[47] D. Giri, S. Shreya, P. Kumari, and R. Yadav, "Indoor human occupancy detection using machine learning classification algorithms & their comparison," in *Proc. IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1110, 2021, Art. no. 012020.
[48] S. I. Kampezidou, A. T. Ray, S. Duncan, M. G. Balchanos, and D. N. Mavris, "Real-time occupancy detection with physics-informed pattern-recognition machines based on limited CO_2 and temperature sensors," *Energy Buildings*, vol. 242, Jul. 2021, Art. no. 110863.
[49] C. Wang, J. Jiang, T. Roth, C. Nguyen, Y. Liu, and H. Lee, "Integrated sensor data processing for occupancy detection in residential buildings," *Energy Buildings*, vol. 237, Apr. 2021, Art. no. 110810.
[50] S. Yang, Z. Huang, C. Wang, X. Ran, C. Feng, and B. Chen, "A real-time occupancy detection system for unoccupied, normally and abnormally occupied situation discrimination via sensor array and cloud platform in indoor environment," *Sens. Actuators A, Phys.*, vol. 332, Dec. 2021, Art. no. 113116.
[51] Y. Zhong, S. Fong, S. Hu, R. Wong, and W. Lin, "A novel sensor data pre-processing methodology for the Internet of Things using anomaly detection and transfer-by-subspace-similarity transformation," *Sensors*, vol. 19, no. 20, p. 4536, Oct. 2019.
[52] S. Mittal, K. Gopal, and S. L. Maskara, "Preprocessing methods for context extraction from multivariate wireless sensors data—An evaluation," in *Proc. Annu. IEEE India Conf. (INDICON)*, Dec. 2013, pp. 1–6.
[53] G. Dong and H. Liu, *Feature Engineering for Machine Learning and Data Analytics*. Boca Raton, FL, USA: CRC Press, 2018.
[54] F. Smarandache, *Neutrosophic Perspectives: Triplets, Duplets, Multisets, Hybrid Operators, Modal Logic, Hedge Algebras. And Applications*, 2nd ed. Brussels, Belgium: Pons, 2017, p. 335. [Online]. Available: https://digitalrepository.unm.edu/math_fsp/27
[55] Y. Guo and H. D. Cheng, "New neutrosophic approach to image segmentation," *Pattern Recognit.*, vol. 42, no. 5, pp. 587–595, May 2009.
[56] W. Li, I. Santos, F. C. Delicato, P. F. Pires, L. Pirmez, W. Wei, H. Song, A. Zomaya, and S. Khan, "System modelling and performance evaluation of a three-tier cloud of things," *Future Gener. Comput. Syst.*, vol. 70, pp. 104–125, May 2017.
[57] Z. Zhang, X. Chen, J. Ma, and J. Shen, "SLDS: Secure and location-sensitive data sharing scheme for cloud-assisted cyber-physical systems," *Future Gener. Comput. Syst.*, vol. 108, pp. 1338–1349, Jul. 2020.
[58] L. Candanedo, "UCI machine learning repository: Occupancy detection data set," School Inf. Comput. Sci., UCI Mach. Learn. Repository, Univ. California, Irvine, CA, USA, Tech. Rep., 2016. Accessed: May 12, 2021.
[59] L. Candanedo, "Occupancy-detection-data," GitHub, San Francisco, CA, USA, Tech. Rep., 2015. Accessed: Nov. 9, 2021.
[60] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informat.*, vol. 17, no. 1, pp. 168–192, Jan. 2021.
[61] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced data sets," *J. Inf. Eng. Appl.*, vol. 3, no. 10, pp. 1–13, 2013.
[62] D. L. Olson and D. Delen, "Performance evaluation for predictive modeling," in *Advanced Data Mining Techniques*. Berlin, Germany: Springer, 2008, doi: 10.1007/978-3-540-76917-0_9.
[63] Q. Gu, L. Zhu, and Z. Cai, "Evaluation measures of the classification performance of imbalanced data sets," in *Int. Symp. Intell. Comput. Appl.*, pp. 461–471. Springer, 2009.
[64] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, *arXiv:2010.16061*.
[65] R. Trevethan, "Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice," *Frontiers Public Health*, vol. 5, p. 307, 2017, doi: 10.3389/fpubh.2017.00307.

NOHA S. FAYED received the B.Sc. degree in computer science from the Faculty of Computers and Information, Mansoura University, Egypt, in 2007, the Software Development Diploma degree from the Information Technology Institute, Egypt, in 2008, and the master's degree in information technology from Mansoura University, in 2016, where she is currently pursuing the Ph.D. degree in information technology with the Faculty of Computers and Information. She worked as a Demonstrator at the Faculty of Computers and Information, Mansoura University, from 2009 to 2016, and has been a Teaching Assistant, since 2016. Her research interests include wireless sensors networks, data analysis, the IoT, and machine learning.

MOHAMMED M. ELMOGY (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Faculty of Engineering, Mansoura University, Mansoura, Egypt, and the Ph.D. degree from the Informatics Department, MIN Faculty, Hamburg University, Hamburg, Germany, in 2010. From July 2016 to August 2019, he worked as a Visiting Researcher with the Department of Bioengineering, University of Louisville, Louisville, KY, USA. He is currently a Professor with the Information Technology Department, Faculty of Computers and Information, Mansoura University. He advised more than 35 master's and doctoral graduates. He has authored/coauthored over 220 research publications in peer-reviewed reputed journals, book chapters, and conference proceedings. His current research interests include computer vision, medical image analysis, machine learning, pattern recognition, and biomedical engineering. He is also a Professional Member of the ACM Society. He has served as a Technical Program Committee Member for many workshops and conferences. He has also served as a reviewer for various international journals.

AHMED ATWAN received the B.Sc., M.Sc., and Ph.D. degrees in electronics and communications engineering from the Faculty of Engineering, Mansoura University, Egypt, in 1988, 1998, and 2004, respectively. From 2000 to 2004, he worked as a Lecturer. He worked with the Faculty of Computers and Information Sciences (FCIS), Mansoura University, as an Assistant Professor, from 2005 to 2009, and later as an Associate Professor, from 2010 to 2015, where he has been a Professor, since 2016. Since 2019, he has been a Professor with the Faculty of Computer and Information Technology, Northern Border University, Saudi Arabia. His research interests include pattern recognition, networking, signal processing, and image processing.

EMAN EL-DAYDAMONY received the B.S., M.S., and Ph.D. degrees in electrical communications from the Faculty of Engineering, Mansoura University, Egypt, in 1998, 2003, and 2008, respectively. She worked as a Teaching Assistant, from 1998 to 2002, a Teacher, from 2003 to 2007, and a Lecturer, from 2008 to 2010. From 2010 to 2017, she worked as a Lecturer with the Faculty of Computers and Information Sciences (FCIS), Mansoura University. Since 2018, she has been an Associate Professor with FCIS. Her current research interests include computer vision, pattern recognition, medical image processing, biomarker discovery, and bioinformatics.