Classification of Manifold Learning Based Flight Fingerprints of UAVs in Air Traffic

Ümit Çelik and Haluk Eren

***Abstract*— As the number of UAVs (Unmanned Aerial Vehicles) and the market size have been expanding rapidly in recent years, projects such as NextGen and SESAR aim to include UAVs in air traffic. Therefore, different perspectives on understanding flight patterns can contribute to more effective management of future air traffic. Analysis of flight data offers an important insight into the operations of a UAV. In this study, it is aimed to extract a flight fingerprint using different machine learning techniques by means of a public dataset and the data obtained from our experimental flights. To get the individual flight pattern, multidimensional UAV sensor data has been reduced using manifold learning methods. By comparison, the most proper manifold method that allows highest classification accuracy (CA) has been investigated. Their performances are compared using both different manifold types and different classification methods. Then, the obtained manifold is used as flight fingerprints and validated by classification techniques. Various unsupervised manifold learning techniques such as t-Distributed Stochastic Neighbor Embedding (t-SNE), Locally Linear Embedding (LLE), Isometric Feature Mapping (ISOMAP) were tried for dimension reduction. For flight fingerprint classification, supervised machine learning techniques such as k-Nearest Neighbors (k-NN), Adaboost, Neural Network, Bayes, etc., were tested. It has been observed that the highest classification accuracy is achieved with the t-SNE manifold and k-NN classification pair. The extracted fingerprint can find many application areas such as performance tests in production lines, air traffic control, risk analysis, anomaly detection, observing pilot performance, drone efficiency over time.**

***Index Terms*— UAV, manifold learning, flight fingerprints, flight classification, drone, air traffic.**

I. Introduction

THERE are many kinds of land vehicles such as bicycles, tracks, tractors, buses, passenger vehicles, trains. We know their characteristics, which could be very large, small, slow, in low maneuver capability, etc. In a similar vein, we need to know characteristics of all the air vehicles which are more complex than land vehicles. They fly over lands at different altitudes, at different speeds, and maneuver capabilities. Some of them need long runways or special land portions, and some of them need a small field within a city. There are many kinds of air vehicles but the regulations have mostly been defined for airplanes. Recently, there are some global projects involving all types of air vehicles into the air traffic information system such as NextGen [1], SESAR [2], CARATS [3]. It means, new regulations should be developed for the new sky. For international laws, agreements, and local air traffic control and services, we should know the characteristics of air vehicles under different conditions. In the near future, we may mostly see unmanned air vehicles in the sky such as air taxis, packet delivery drones, unmanned military vehicles, postal services, and unmanned helicopters. Hence, we should be ready for future unmanned life in the sky. For these reasons, we have focused on this study on UAVs. We cannot investigate all types of UAV vehicles. However, we need to restrict our study with drones to give specific examples to the small UAVs. If we can generate the fingerprints of specified drones, we can effectively arrange air, land and infrastructures, services, and privileges. Therefore, we can correctly prepare new generation regulations covering all types of unmanned air vehicles. Taking an example of our study, researchers may find fingerprints of all types of UAVs.

The aviation industry generally operates with a low profit margin, while energy costs and other uncertainties make the situation even more fragile [4]. Therefore, analysis of high dimensional flight data with machine learning techniques can increase flight operation efficiency and reliability, as well as reduce insurance costs [5]. Traditional statistical approaches rely on predefined thresholds and cannot provide sufficient information for flight analysis. On the other hand, machine learning techniques can enable comparison of high-dimensional flight data and obtaining unknown patterns. By obtaining flight fingerprints as suggested in this study, abnormal or inconsistent flight behaviour can be recognized at its first stages.

A. Problem Statement

There are different types of aircrafts such as fixed wing and rotating wing. Therefore, monitoring whether the drone is flying in accordance with its own production technology is a reassuring approach in terms of detecting anomalies in the flights of future vehicles. Thus, the quality of the flight performed after each operation and pilot performance, if any, can be measured, or it can be used to determine the most suitable three-dimensional route for the UAV.

The flight characteristics of all the airplanes in air traffic are well known by air traffic controllers and the other service providers. In addition to conventional air vehicles, we need to know all types of UAV air vehicles' characteristics depending on their intrinsic parameters under different weather conditions

Manuscript received 1 March 2022; revised 8 August 2022 and 25 October 2022; accepted 12 January 2023. Date of publication 20 January 2023; date of current version 8 May 2023. The Associate Editor for this article was L. Wang. *(Corresponding author: Ümit Çelik.)*

The authors are with the School of Civil Aviation, Firat University, 23119 Elaziğ, Turkey (e-mail: u.celik@firat.edu.tr; heren@firat.edu.tr).

Digital Object Identifier 10.1109/TITS.2023.3237159

and air traffic. Considering their altitudes and speeds, drones are suitable for visual flight rules. Therefore, in order to understand the flight pattern of drones, we need to conduct multiparameter approaches. The increase in data size makes it more complex to extract the flight characteristics of the system. For this reason, we need to find a simpler solution to obtain a clear fingerprint representing flight characteristics of each drone. Thus, we need to reduce the dimension using nonlinear dimension reduction methods which refers to unsupervised learning based techniques. Then, we will select the best method/s representing the fingerprint of the flight. After that, we should find the most suitable classification method to clearly distinguish each flight fingerprint from each other. Eventually, we will find the most efficient methods and their related stages to get flight fingerprints and its classification. In this manner, we can easily understand the flight characteristics of UAVs in the sky. It may also give us the flight quality in operations. This golden information can be used by manufacturers and service providers.

In the study, it will be decided to collect the sensor data that is effective in determining the flight characteristics during the flight. Flight data acquired on the same flight route at different altitudes, payloads and speeds are used. The size of multidimensional sensor data will be reduced by non-linear size reduction techniques. Thus, flight fingerprints representing the flight characteristics will be obtained. By comparing the obtained fingerprints, dominant sensors that can represent the flight pattern will be decided. Finally, data analysis will be performed for flight classification.

B. Proposed Approach

In this study, it is aimed to obtain a flight fingerprint representing the flight characteristic by using the flight sensor data. Multidimensional sensor data is processed with data reduction techniques to obtain a low-dimensional manifold representing a flight. In this way, it is aimed to easily compare the flight characteristics performed by different routes, altitudes or different pilots. The proposed approach can contribute to adaptive route planning studies.

As an application for the study, flight fingerprints were obtained by traveling a certain distance at different altitudes, speed, and payload. Thus, it is aimed to classify each flight data according to flight pattern. It has been decided to carry out the flights on the same route of 2 km round trip. The data obtained by the flight was taken from the UAV, and it was decided to reduce the dimension of data with different manifold learning methods. The same approach was also applied to the public drone dataset. Although there are many manifold learning methods, the reason why the t-Distributed Stochastic Neighbor Embedding (t-SNE) method is preferred is that it gives higher classification performance than other techniques. The resulting reduced size data will be used as a flight fingerprint that represents the flight performed.

Our main goal is not to get a global fingerprint which represent the flight under all types of weather conditions. In other words, the proposed fingerprint does not refer to a robust measure characterizing for individual drones. In fact, the fingerprint characterizes the actual flight performed. It means that if we want to perform various flights under different weather conditions such as rainy, cloudy or windy, we should get several fingerprints which refers to fingerprint database characterizing flight of drones for different extrinsic parameters as mentioned in Table II.

In this study, it is not aimed to create a flight fingerprint dataset. Rather, a method was proposed for researchers to create such a dataset. Flights operated under different conditions are classified as an application. As a future work, we are planning to find a global fingerprint representing the flight under any condition for an individual drone. Researchers can use the proposed method to obtain unique fingerprints of pilots and aircrafts.

C. Related Work

Most of the studies using flight data in the literature have been done for a specific improvement or anomaly detection [6], [7], [8], [9], [10]. A detailed review on the anomaly detection can be found at [11]. For example, positional information was used to obtain the trajectory of the aircraft [12]. Multisensory data was used to understand the fuel situation of the aircraft [13]. Although we conducted a large literature search, we could not find a study with a similar approach that focuses to obtain flight fingerprint from individual flights. In fact, our study does not aim to achieve an optimization or anomaly like previous studies. Our study proposes a method to obtain a flight database that can serve as a basis for studies focused on a specific target. In our study, a method on how to obtain each signature that creates such a database by using spatiotemporal flight data is proposed. Therefore, a signature database for that aim can be created by following the proposed method for the purpose of future studies. Using this signature database, a targeted global measure or fingerprint can be obtained.

In dimension reduction methods, it is aimed to preserve the general characteristics of the data and to reduce the size of the data by eliminating the dataset that does not contain valuable information [14]. We use non-linear dimensionality reduction which is conducted by manifold learning to get flight fingerprint classification. Each flight has a particular manifold. Manifold learning searches for the intrinsic low-dimensional embedding structures within high-dimensional data [15]. Manifold learning has been used in many applications in recent years such as process monitoring [16], electronic tongue classification [17], electronic nose [18]. Dimensionality reduction is a smart way to visualize high-dimensional data, but also it can be employed for raising the classification accuracy (CA).

Semi-supervised anomaly detection methods were investigated for health monitoring of a real aircraft system [8]. A real-time, data-driven aircraft health monitoring technique that detects nominal flight operation is proposed [19]. Runway deviation risk was estimated using flight data based on Bayesian network [20]. Model-based and time-based method are proposed to decide whether flight trajectories are abnormal or not [21].

In the existing literature there are many research to understand flight anomalies using flight trajectory. In these studies, different flight trajectory components were deployed. In a research study, each flight trajectory consists of flight modes such as take-off, climb, cruise, hover, descent, and landing [22]. In another one, flight trajectory was described using parameter based approach by gyro data such as roll, yaw, and pitch angle [12]. In these research studies, the data representing the trajectory consists of either directly spatial or geometric-based parameter data.

However, in our study we have divided multisensory data into two group which are intrinsic and extrinsic. Our dataset does not only focus on trajectory data like the previous studies. Rather, we used position independent multidimensional sensory data for representing a flight characteristic.

Real-time anomaly detection was conducted using FDR data by monitoring sudden changes in features and the computational time was decreased by 38.2% using maximal information compression index by Chattopadhyay et al. [23]. Aircraft fuel consumption was estimated using aircraft operational parameters extracted from FDR and ADS-B data by Cheng at al. [13]. None of the studies considered the flight data as signature.

D. Contribution and Novelty

In the existing study, nonlinear dimension reduction techniques were used to determine the UAV flight characteristics. The flight fingerprint is extracted by proposing a definition that represents the flight behaviour. Among the applied manifold learning methods, the highest performing ones were found and evaluated. In addition, many classification techniques have been tried for the performance comparison. Thus, the most suitable classification technique that can be used with the chosen manifold learning method has been determined. With this approach, a separate flight characteristic can be extracted for each UAV. Thus, after the drone prototype is manufactured, the lowest cost and most stable drone can be found for flights in different conditions for mass production. In case of remote control, pilot behaviour can also be deduced with this method. Estimation of Flight Fuel Consumption can be conducted comparing specified portions of Flight Track Data and the proposed fingerprint. In addition, mechanical problems such as engine and propeller failures that may occur during flight can be understood due to the fact that each potential mechanical problem carries a different fingerprint. It will allow diagnostics to be made during the flight.

E. Outline

The rest of the sections contains methodology including manifold learning techniques, normalization, and classification methods in Section II. In Section III, the experimental results are given for drones' flights in tables and plots including performance evaluations. After that, the discussions are provided for the experiments in Section IV. The conclusion is located in Section V.

II. Methodology

Detection of abnormal conditions on any system is one of the main goals of data science research and there are still many challenges to overcome in this field. Unlike traditional methods, the use of machine learning techniques, which include brand new theories, is becoming widespread in aviation. The use of machine learning techniques to solve the problems encountered in rapidly increasing UAV operations using flight data is also widespread. In this work, we focused on obtaining flight fingerprints using the flight sensors data recorded by drones. We aim to classify drone flights by finding their characteristics under certain flight conditions. In this way, with simultaneous flight data processing, any anomalies can be predicted before they become a serious problem. Flight fingerprint assessment system steps are given in Fig. 1.

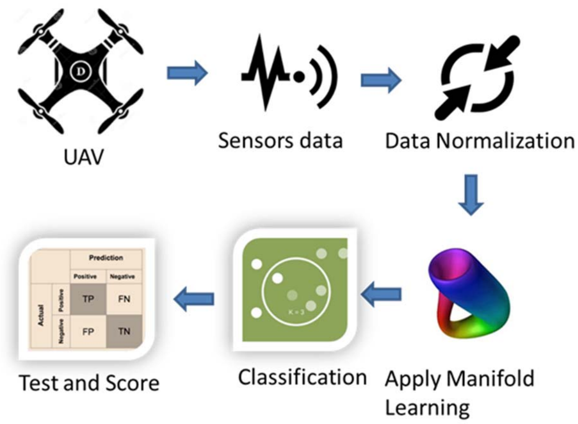


Fig. 1. Flight fingerprint assessment system.

All sensor data is recorded during the UAV flight. Sensor data to extract flight characteristics can be processed at the ground station or on the drone. In the study, both sensor data from the existing database and from our flights were used. By applying manifold learning to high-dimensional sensor data, a fingerprint representing flight is obtained. It would be much faster and simpler to classify and evaluate this fingerprint compared to high-dimensional data. In our study, even if we focused on the classification of fingerprints, in the continuation of this study, after determining the flight quality, the route risk factor for a particular route can be estimated on a map.

In this study, flight sensors can be divided into two classes as internal and external. Internal sensors are sensors that display the instantaneous state of the aircraft such as accelerometer, gyroscope sensors. External sensors, on the other hand, are sensors carried by the aircraft and observe the situation around the aircraft. Flight sensors are listed in the Table I.

Flight parameters can be divided into 3 classes which are intrinsic, extrinsic factors, and user programmed. Intrinsic parameters are data generated by installed sensors required for aircraft flight. Extrinsic parameters are parameters related to meteorology in the outside world that affect flight. User programmed parameters are flight conditions determined by the pilot. In this study, firstly, the flight parameters that can be obtained from the characteristic flight fingerprint were tried to be determined. The parameters which are collected by the sensors given in Table I are summarized in Table II.

TABLE I
FLIGHT SENSORS AND ACQUIRED PARAMETERS

Flight Sensors	
Internal	**External**
• Accelerometer • Gyroscope • IMU • Internal temperature • Battery voltage, current	• Magnetometer • Barometric pressure • GPS • Airspeed sensor • Wind direction • Outside temperature • Pressure

TABLE II
FLIGHT PARAMETERS

Parameters		
Intrinsic	**Extrinsic**	**User Programmed**
• Battery voltage • Battery current • Position (XYZ) • Orientation (XYZW) • Velocity (XYZ) • Angular velocity (XYZ) • Linear acceleration (XYZ) • Internal temperature	• Wind speed • Wind angle • Pressure • Outside temperature • GPS (XYZ)	• Altitude • Speed • Payload

In different datasets used in this study, data on some internal and external parameters were collected under user-programmed flight conditions. Therefore, in this study, we aim to determine user determined flight conditions by using intrinsic and extrinsic parameters as a sample application. In order to find the optimal parameter sets, classification accuracy showing certain altitude, payload and speed conditions are calculated. In this way, the contribution of each parameter to the classification will be obtained. In order to determine the parameter sets from which we can obtain characteristic flight fingerprints, firstly, the data in each parameter group, e.g. accelerometer XYZ, were normalized. Then, the data were reduced to 2D by applying the dimension reduction technique. After then, the classification accuracy of this parameter was calculated with the k-NN classification method. The result obtained is evaluated as the classification strength of this parameter. After this process is obtained for all parameters, the parameter classification strength table is obtained.

In the second step, classification accuracy is calculated with the aforementioned steps by using all intrinsic and extrinsic parameters in the dataset as seen in Table II. Some sensor data may not reflect flight characteristics and may have a negative impact on flight classification accuracy. Therefore, in order to determine these parameters, the classification effect of each sensor data on the flight characteristics is tested and the sensor data with the lowest effect on the classification accuracy (CA) is not used.

In this way, it is aimed to achieve the most characteristic flight fingerprint. Procedure for calculating parameter classification performance and procedure for parameter set optimization are given in Fig. 2a and 2b, respectively. In the study, the accelerometer XYZ and Gyro XYZ data recorded during the flight and the public dataset were reduced to a 2-dimensional dataset with different manifold learning methods. This resulting dataset is considered as a flight fingerprint.

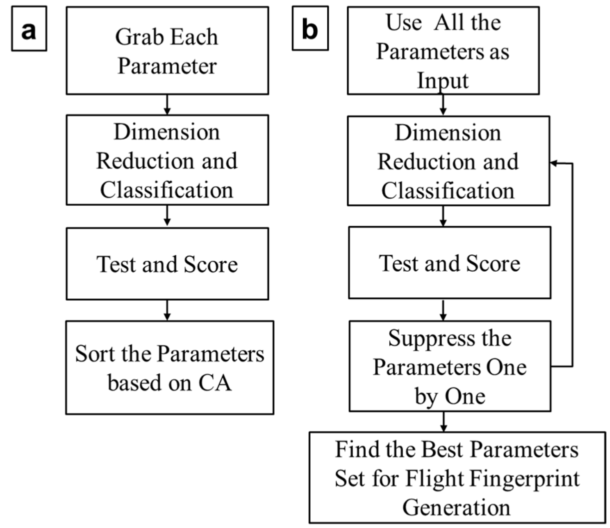


Fig. 2. a) Procedure for determining each parameter classification performance b) procedure for parameters set optimization.

TABLE III
DESCRIPTION OF VARIABLES FOR THE DATA RECORD

Variable	Dimensions	Unit
Wind speed	1	m/s
Wind angle	1	deg
Battery voltage	1	V
Battery current	1	A
Orientation x; y; z; w	4	quaternion
Velocity x; y; z	3	m/s
Linear acceleration x; y; z	3	m/s^2

A. Datasets

The publicly available dataset [24], [25] was used in the experiments. The public dataset was obtained by performing 209 different flights. The sensor data of the dataset in Table III were used in the experiments. Public dataset includes 257896 data points for each sensor for 209 different flights at different conditions. Flight data were collected during the autonomous flights; whose dimension is totally 14D as seen in Table III.

In addition, sensor data was collected using the DJI Phantom 4 drone to validate the results tested on the public dataset. Our own dataset includes 14131 data points for each sensor for 6 different flights at different altitudes. We used the same pilot for the flights at different altitudes. The flights were performed within the same hour and no significant weather changes were observed during the flights. Descriptions of variables for our data are given in Table IV.

TABLE IV
DESCRIPTION OF VARIABLES FOR OUR DATA

Variable	Dimensions	Unit
Battery voltage	1	V
Battery current	1	A
Velocity XYZ	3	m/s
Magnetometer XYZ	3	µT
Accelerometer XYZ	3	m/s2
Gyroscope XYZ	3	°/s

TABLE V
MACHINE LEARNING TECHNIQUES USED IN THIS WORK

Machine Learning Techniques	
Supervised Learning	**Unsupervised Learning**
• k-NN [26] • SVM (Support Vector Machines) [27] • SGD (Stochastic Gradient Descent) [29] • Neural Network [32] • Naïve Bayes • Logistic Regression • AdaBoost [35]	• Manifold Learning o ISOMAP [28] o Local Linear Embedding [30, 31] o t-SNE [33] • PCA [34]

B. Machine Learning Techniques

Clustering is an unsupervised machine learning technique that provides subsets of each data, taking into account the similarities between the parameters. Supervised techniques are trained using the training dataset. These techniques produce a function that is inferred from the input and the correct output. Machine learning techniques used in this work are listed in Table V.

In this study, it is aimed to obtain a fingerprint that can represent the flight by processing the multidimensional sensor data of a particular flight using unsupervised dimension reduction techniques that can enable identifying previously unknown patterns in the multisensory data. Firstly, we applied data normalization to multidimensional data. The purpose of data normalization is to transform data belonging to different units or unitless parameters so that they have a similar distribution. We tested different data normalization approaches and obtained higher classification accuracy by normalizing the data in between -1 and 1. After normalization is applied to the dataset as a pre-process, the data is processed by ML techniques.

Considering sensors for a simple UAV, it is seen that there are dozens of dimensional sensor datasets which do not always have a linear relationship between each other. Manifold learning techniques are nonparametric dimension reduction implementations and they can be used to reduce the dimension of data without losing the non-linear interactions in the dataset. The reduced data keeps the envelope of the original big data. The high-dimensional dataset obtained in many applications can be modelled as a manifold in a lower dimension using manifold learning techniques.

PCA found numerous applications for data driven schemes in which dimension reduction is needed. PCA is appropriate for obtaining linear parameters related to the underlying problem dataset. However, it does not work well for intrinsic nonlinear relations in the dataset. Manifold learning can be expressed as an extended version of PCA. With its enhanced structure, we can obtain different nonlinear schemes of the dataset [36].

Fig. 3. Experimental flight route that the sensor data is captured by DJI Phantom 4 drone (Route 2).

III. EXPERIMENTAL RESULTS

In this work, we aim to obtain fingerprints specific to each flight that specifically includes flight characteristics. We tested the proposed approach on two different datasets. The publicly available dataset [24] was used in the experiments as the first dataset. Further information about the datasets can be found in the Datasets section. The public dataset was captured during autopilot controlled flight and it consists of sensors data as described in Table III. As a second dataset, the sensors data was collected using the DJI Phantom 4 drone. We used the same pilot for the flights at different altitudes. The flights were carried out within the same hour and no significant weather changes were observed during the flights.

The flights were carried out between point 1 and point 2 as seen in Fig. 3. In the experiments, a total of 2 km round trip was travelled on a certain route. Flights were performed at different altitudes. Accelerometer, gyroscope, and GPS sensor data were recorded during the flight. In the experiments, flights were carried out at 15m, 30m, 45m, 60m, 75m, and 90m altitudes. Approximately 2500 data points were collected for each flight. In Fig. 4, the raw data plots of gyroscope XYZ and accelerometer XYZ for 15m and 90m are given over time. It is difficult to estimate the direct flight characteristic by looking at the raw data plots.

The public dataset was obtained by performing 209 different flights [24]. Flights at different altitudes (25m, 50m, 75m, 100m), speeds (4m/s, 6m/s, 8m/s, 10m/s, 12m/s) and payload mass (no payload, 250g, 500g) have been carried out. In this work, by keeping two of the three variable parameters constant, the variation of the relevant pattern for different cases was analysed.

Flights were divided into 4 groups in order to obtain the fingerprint of each flight parameter determined by the user, while keeping the other parameters constant. The flight parameter details are given in Table VI. For example, for classification experiments, the altitude of the flights in the public dataset in the 1st group is variable, and the speed and payload are selected as constant.

TABLE VI
FLIGHT GROUPS FOR DIFFERENT ALTITUDES, SPEEDS, AND PAYLOADS

Group ID	Speed (m/s)	Payload (gr)	Altitude (m)	Route	Flight ID
1	**4**	500	50	R1	1
	6				2
	8				3
	10				4
	12				5
2	4	**0**	25	R1	6
		250			7
		500			8
3	8	250	**25**	R1	9
			50		10
			75		11
			100		12
4	6	0	**15**	R2	13
			30		14
			45		15
			60		16
			75		17
			90		18

A. Comparison of Manifold Learning Techniques for Flight Classification

First, the flight dataset was reduced to 2D using linear and non-linear dimension reduction methods for certain flight conditions as listed in Group 1 in Table VI, and data classification was conducted by k-NN classification. In this experiment, it was aimed to determine the most proper manifold learning method that can achieve the highest classification accuracy. As a result of our experiments, it was concluded that the highest classification accuracy was obtained with t-SNE manifold learning. The classification accuracy and running time for each manifold learning technique is given in Table VII.

In the experiments, the software was run on a computer with Intel I7 10th generation and 16GB RAM. Running time was measured for decreasing the data from 14 dimensions (14D) to 2 dimensions (2D) for 5881 items and classifying the 2D output data with k-NN. The data collected during a flight was used to generate a flight fingerprint representing the current flight. It was not intended to obtain a real-time fingerprint in this work. However, one of the future studies may be obtaining flight anomalies by comparing instant fingerprints obtained during flight with previously extracted fingerprints. Thus, it is checked whether the actual fingerprint and the previous observed fingerprint match each other as an envelope.

In addition, for comparison with nonlinear dimension reduction techniques, the data was reduced to 2 dimensions with PCA, which is a linear dimension reduction technique, and classification accuracy was obtained. Considering the computational complexity of using nonlinear size reduction techniques such as t-SNE, it can be assumed that it is more advantageous to prefer the PCA method if the PCA technique provides higher classification accuracy. However, in the experiments, it was observed that the classification accuracy performance after size reduction with PCA was lower than that of t-SNE.

B. Comparison of Flight Classification Methods Based on 2D Manifolds

Initially, flight classification was performed using sensor data collected at different flight speeds while keeping payload and altitude constant. In Experimental Group 1, flights data with 500gr payload, 50m altitude and flight speeds of 4m/s, 6m/s, 8m/s, 10m/s and 12 m/s were used as listed in Table VI.

Flight classification was conducted by using the sensor data collected for different payloads by keeping the speed and altitude constant. In Experimental Group 2, flights data with a speed of 4m/s, an altitude of 25m, and payloads of 0gr, 250gr, and 500gr were used. Flight classification was carried out using sensor data collected at different altitudes by keeping speed and payload constant. In Experimental Group 3, 250 gr payload, 8m/s speed, 25m, 50m, 75m and 100m altitude flight data were used. In Experimental Group 4, we collected flight data at different altitudes, corresponding to 6 different classes, 15m, 30m, 45m, 60m, 75m and 90m, using a DJI Phantom 4 drone.

Sensor data dimensions for certain flights are reduced to 2D data using t-SNE manifold learning. The most effective parameters that can be used in flight classification have been tried to be revealed. The highest classification accuracy was achieved using wind speed angle, battery voltage current, XYZW orientation, XYZ velocity and XYZ acceleration data for flight groups 1, 2 and 3. It was observed that the angular velocity sensor data decreased the classification accuracy from 0.984 to 0.942, therefore the angular velocity sensor data was not used in the experiments.

In Group 4 experiments, XYZ accelerometer, XYZ gyroscope, XYZ magnetometer, XYZ velocity, battery current and voltage data corresponding to 14D data were used for flight altitude classification. The data was normalized from -1 to 1 to process the sensor data with the same weight. We applied t-SNE manifold learning to reduce the data size from 14D to 2D for classification. In fact, we measured CA by reducing the dimensionality to 2D, 3D, and 4D, resulting in 0.973, 0.980, 0.980, respectively. Since there was no significant performance difference, we continued by reducing the data to 2 dimensions. 90% data was used for k-NN training and 10% data was used to test the model. We also tested the cross-validation technique with 5 and 10 folds and the CA was measured as 0.961 and 0.966, respectively. As seen in Table VII, CA was measured 0.969 by the training/test separation method. Due to the low impact of the cross-validation technique on the results, we continued the experiment with the training/test split method.

In this study, the original 14D and 2D dimension reduced data were classified using k-NN, AdaBoost, Neural Network, Naïve Bayes, SVM, Logistic Regression, and SGD classifiers. Classification accuracy (CA) values were calculated for each case. Neighbour number 5 and Euclidean Distance are used for k-NN classifier. Manhattan and Chebyshev distance with neighbourhood from 2 to 8 was also tested, but we could not observe any notable improvement on CA. A v-SVM type with linear kernel function and regression cost equal to 1 was used for SVM. For SGD, hinge loss function,

TABLE VII

CLASSIFICATION RESULTS FOLLOWING DIMENSION REDUCTION WITH DIFFERENT MANIFOLDS

Dimension Reduction Method	Running Time (Seconds)	k-NN Classification Accuracy
t-SNE Manifold	15	0.97
ISOMAP Manifold	8	0.78
LLE Manifold	2.5	0.76
MDS Manifold	48	0.73
Spectral Embedding Manifold	6	0.62
PCA	1.5	0.82

TABLE VIII

COMPARISON OF FLIGHT CLASSIFICATIONS AT DIFFERENT FLIGHT SPEEDS

Model	CA	F1	Precision	Recall
k-NN	0.969	0.969	0.970	0.969
AdaBoost	0.961	0.961	0.961	0.961
Neural Network	0.823	0.822	0.824	0.823
Naive Bayes	0.383	0.352	0.372	0.383
SVM	0.338	0.326	0.376	0.338
Logistic Regression	0.330	0.248	0.236	0.330
SGD	0.238	0.205	0.220	0.238

ridge regularization and constant learning optimization with 1000 iterations were used. For the neural network, 100 hidden layers, ReLu activation and 400 iterations were used. Ridge regularization was used for logistic regression. For Adaboost, the number of estimators equal to 50, learning rate of 1 and SAMME.R classification was used. It was observed that the method with the highest classification accuracy was the k-NN classifier. The main reason k-NN performs better might be that it can handle non-linear data. The fact that the t-SNE nonlinear dimension reduction technique yields better results than PCA linear dimension reduction technique means that the data exhibit nonlinear behavior. A comparison of the classifiers used to classify 2-dimensional manifolds is given in Table VIII.

Fig. 5 represents the confusion matrix for the k-NN classification, which is obtained by reducing 14D data with different flight speeds to 2D using t-SNE manifold learning. As seen in the confusion matrix, flights with Flight IDs of 1, 2, 3, 4, 5 were classified with a classification accuracy of 0.969.

C. Flight Fingerprint Classification for Different Flight Conditions

After normalizing the 4 groups of 14-dimensional flight data, it was reduced to 2 dimensions with t-SNE manifold learning.

Then the data were classified using k-NN, AdaBoost, Neural Network, Naïve Bayes, SVM, Logistic Regression and SGD classifiers. The obtained CA values are given in Table IX and Fig. 6 for comparison. We observed that k-NN had the highest CA for classification of any experimental group. AdaBoost outperformed than Neural Networks for classifying flight conditions, but still had acceptable CA. There was a

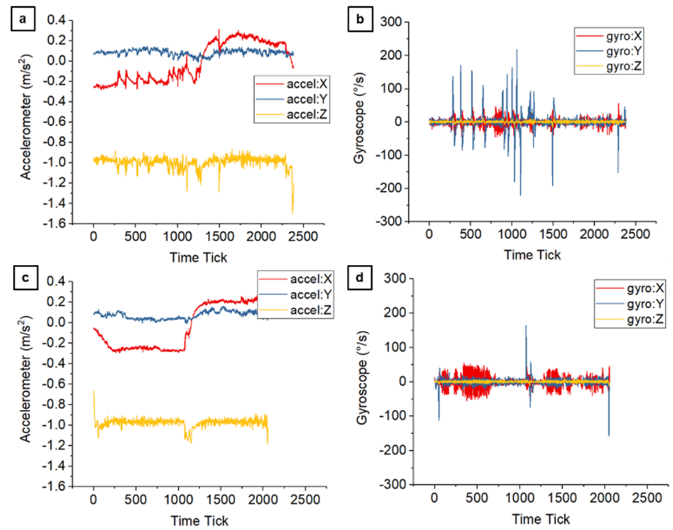


Fig. 4. a) Accelerometer XYZ data for 15m altitude, b) gyroscope XYZ data for 15m altitude, c) Accelerometer XYZ data for 90m altitude, d) gyroscope XYZ data for 90m altitude.

		Predicted Class				
		1	2	3	4	5
Actual Class	1	100.0%	0.0%	0.0%	0.0%	0.0%
	2	0.0%	99.2%	0.0%	0.8%	0.0%
	3	0.0%	0.0%	92.9%	0.0%	3.0%
	4	0.0%	0.8%	2.0%	96.6%	1.5%
	5	0.0%	0.0%	5.1%	2.5%	95.5%

Fig. 5. Confusion matrix for flight classifications at different flight speeds.

TABLE IX

COMPARISON OF CLASSIFICATION ACCURACY (CA) FOR 4 GROUPS OF EXPERIMENTS

Model	Different Speeds	Different Payloads	Different Altitudes	Different Altitudes (Our Data)
k-NN	0.969	0.998	0.986	0.998
AdaBoost	0.961	0.993	0.970	0.996
Neural Network	0.823	0.978	0.873	0.988
Naive Bayes	0.383	0.611	0.491	0.516
SVM	0.338	0.516	0.407	0.384
Logistic Regression	0.330	0.494	0.368	0.359
SGD	0.238	0.519	0.329	0.336

sudden drop in performance when we used Naive Bayes, SVM, Logistic Regressions and SGD.

After the 14D sensor data were normalized, they were classified with k-NN. In addition, after the 14D data was normalized, it was converted into 2D data with t-SNE and classified with k-NN again. Obtained results are given in Table X. After the data was reduced to 2D with t-SNE manifold learning, classification accuracy was calculated as 0.986. Although the data size was reduced from 14D to 2D, the classification accuracy was calculated to be 0.1 higher. Manifold learning reduces size-related complexity. It has been

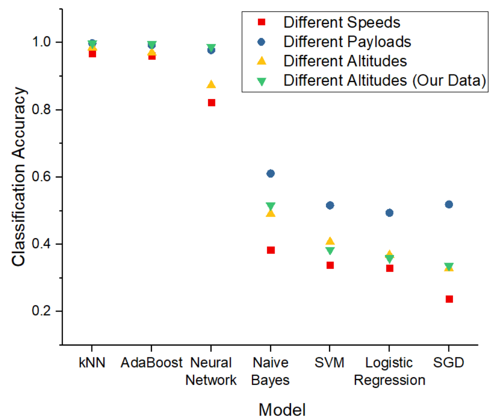


Fig. 6. Comparison of Classification Accuracy (CA) for 4 groups of experiments.

TABLE X
COMPARISON OF CLASSIFICATION ACCURACY (CA) CLASSIFIED BASED ON ORIGINAL DATA AND REDUCED 2D DATA USING K-NN CLASSIFIER

Data Dimensions	**Different Speeds**	**Different Payloads**	**Different Altitudes**	**Different Altitudes (Our Data)**
2D (Manifold Learning)	0.984	0.996	0.986	0.998
14D (Original Dataset)	0.861	0.932	0.897	0.697

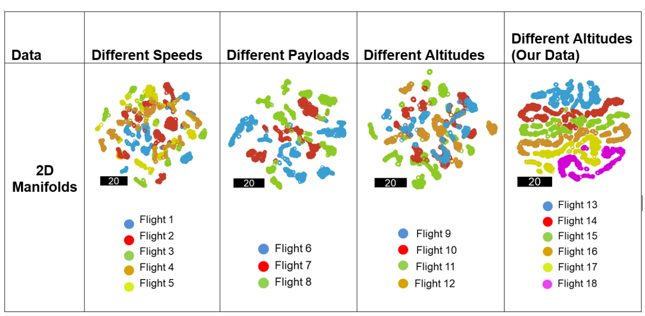


Fig. 7. Flight fingerprints based on 2D manifolds for each flight groups.

observed that flight fingerprints can be classified with a higher CA after data dimension reduction.

Flight fingerprints based on 2D manifolds for each flight groups are given in Fig. 7. The fingerprints are derived from 14D flight data by reducing the dimension to 2D using t-SNE manifold learning method.

IV. DISCUSSIONS

In this study, experimental studies have been carried out to test the most suitable algorithms that will enable us to obtain characteristic flight patterns representing flights using sensor data. The recorded flight data obtained by aircrafts can be converted in a proper manner to get flight traces. When a plane crash occurs, experts need flight traces to understand the cause.

In a similar vein, if the flight of UAV is recorded, it means that we have the flight traces of the drone. In this situation, we can use flight fingerprints to understand possible anomalies. We can classify flights with high classification accuracy, but it is clear that different weather conditions can create a different flight fingerprint. This is something we already want, because it is a "new" flight with different characteristics.

First of all, it is aimed to compare algorithms that reduce the size of high-dimensional datasets without losing the information containing the characteristics of the flights. Afterwards, it is aimed to classify the reduced-size data representing a flight by dataset classifiers to determine the most appropriate algorithm set that can produce the highest classification performance. Extensive use of supervised and unsupervised machine learning algorithms makes interpretation of flight data more efficient and relevant predictions more precise.

The proposed manifold learning approach is most suitable for classifying flight anomalies and further analysing the extracted fingerprints. Unsupervised learning algorithms are a more proper way to analyse multi model datasets for different cases captured by flight sensors. However, supervised methods also need to be considered for classification of anomalies and parameters. After flight data normalization, we use unsupervised machine learning for dimension reduction, and then supervised machine learning for flight data classification.

Determining proper machine learning algorithms is significant for flight data processing and monitoring applications. Considering the many flight sensors, the volume of data collected is huge. Therefore, it is necessary to reduce data sizes and optimize algorithms to speed up the data analysis and outlier identification process.

Sensors may sometimes output abnormally under extreme conditions such as excessive temperature and excessive humidity. Such data do not indicate flight abnormality. In other words, outlier data may come from sensors that do not belong to the flight characteristics of the UAV. In traditional methods, outliers need to be detected and filtered out. Since the manifold represents the overall data, the effect of the outliers is reduced.

Flight data monitoring (FDM) parameters are captured at different data rates and may require preprocessing to consider these data time-critical. Preprocessing of data requires choosing the right supervised or unsupervised machine learning methods for operations such as data size reduction without losing characteristic information, classification of flight patterns. The resulting characteristic flight patterns allow for the prediction of abnormal or inconsistent behaviour. Such situations are more difficult to understand by observing the data stream from the sensor. Unexpected outliers in patterns are important and extensive research is needed to further define this situation.

One of the most important preliminary steps is parameter selection. One of the first goals should be to reduce the number of parameters without losing the characteristic data representing certain situations. A poor parameter selection leads to a poor FDM analysis. It should be determined which parameter set provides more efficient analysis for certain situations and appropriate algorithms should be preferred.

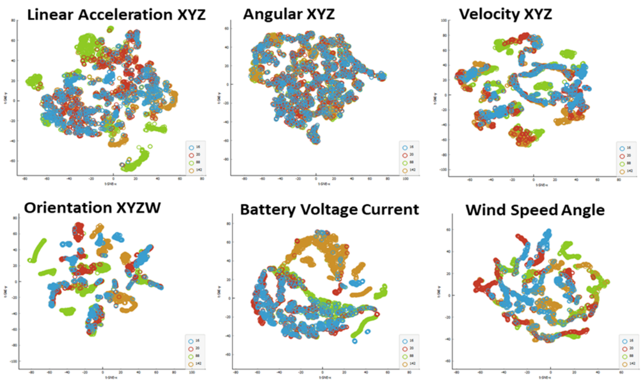


Fig. 8. 2D patterns visually represent the characteristic behaviour of the sensor data.

TABLE XI
CLASSIFICATION PERFORMANCE OF EACH PARAMETER

Parameters	k-NN	AdaBoost	Neural Network
Orientation XYZW	0.880	0.836	0.808
Wind speed angle	0.810	0.762	0.759
Velocity XYZ	0.771	0.704	0.623
Battery voltage current	0.731	0.708	0.694
Linear acceleration XYZ	0.646	0.620	0.611
Angular XYZ	0.343	0.326	0.313

Each flight sensor data reduced to 2D by t-SNE manifold learning is given in Fig. 8. These 2D patterns visually represent the characteristic behaviour of sensor data in distinguishing particular flights. Each colour in the patterns represents a specific flight condition. For example, it is seen that the angular XYZ sensor does not show a characteristic behaviour in separating the flights. We summarized classification performance of each parameter in Table XI. Indeed, an approach that can take into account different types of data and automatically select parameters that provide characteristic information may provide more efficient results than a manually determined parameter set.

Characteristic flight fingerprint representing a particular flight was obtained using t-SNE manifold learning as previously described. t-SNE allows mapping high-dimensional data to low-dimensionality while preserving the significant structure of the original data [33]; and it attempts to 'extract' clustered local groups instead of trying to 'unroll' it. ISOMAP preserves geodesic pairwise distances; LLE preserves local properties; PCA maximizes variance; and MDS preserves Euclidean pairwise distances [33]. t-SNE is very powerful because of this 'clustering' vs. 'unrolling' approach. Significant structure of the original data is important for a particular flight classification, therefore t-SNE performs better than other dimension reduction techniques.

The main reason k-NN performs better might be that it can handle non-linear data [37]. The fact that the t-SNE nonlinear dimension reduction technique yields better results than PCA linear dimension reduction technique means that the data exhibits nonlinear behavior.

V. CONCLUSION

In this study, a flight fingerprint representing the flight pattern was obtained using multisensory flight data. Experiments were conducted at different altitudes, speeds and payloads for a given distance. Each flight data was classified after 2D manifold extraction. Therefore, flight fingerprints representing flight behaviour were obtained. Flight characteristics were classified according to different speeds, payloads and altitudes. In our experiments, we tried out the public dataset and flight sensor data captured by our field experiments using real drones. According to the performance analysis, it was seen that t-SNE manifold learning and k-NN classifier pair provided the highest classification performance for both datasets. Flight classification performances were obtained for the extracted fingerprints. It was observed that the 2D manifold could be classified with higher CA, although the data size was reduced from 14D to 2D. This approach can be used for many state of the art technologies such as production line tests, risk analysis, air traffic control, pilot performance, drone efficiency, crash analysis, and diagnostic applications. As a future work, this approach could be extended for all types of aircraft including manned and unmanned aerial vehicles.

ACKNOWLEDGEMENT

The author(s) received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

[1] C. H. Fleming, M. Spencer, J. Thomas, N. Leveson, and C. Wilkinson, "Safety assurance in NextGen and complex transportation systems," *Saf. Sci.*, vol. 55, pp. 173–187, Jun. 2013.
[2] A. Sipe and J. Moore, "Air traffic functions in the NextGen and SESAR airspace," in *Proc. IEEE/AIAA 28th Digit. Avionics Syst. Conf.*, Oct. 2009, pp. 2.A.6-1–2.A.6-7.
[3] Study Group for the Future Air Traffic Systems. (2010). *Long Term Vision for the Future Air Traffic Systems, CARATS Collaborative Actions for renovation of Air Traffic Systems*. [Online]. Available: http://www.mlit.go.jp/common/000128185.pdf
[4] J. Zuidberg, "Identifying airline cost economies: An econometric analysis of the factors affecting aircraft operating costs," *J. Air Transp. Manage.*, vol. 40, pp. 86–95, Aug. 2014.
[5] E. Smart, "Detecting abnormalities in aircraft flight data and ranking their impact on the flight," Ph.D. thesis, Inst. Ind. Res., Univ. Portsmouth, Portsmouth, U.K., 2011.
[6] M. Memarzadeh, B. Matthews, and I. Avrekh, "Unsupervised anomaly detection in flight data using convolutional variational auto-encoder," *Aerospace*, vol. 7, no. 8, p. 115, Aug. 2020.
[7] K. Sheridan, T. G. Puranik, E. Mangortey, O. J. Pinon-Fischer, M. Kirby, and D. N. Mavris, "An application of DBSCAN clustering for flight anomaly detection during the approach phase," in *Proc. AIAA Scitech Forum*, Jan. 2020, p. 1851.
[8] L. Basora, P. Bry, X. Olive, and F. Freeman, "Aircraft fleet health monitoring with anomaly detection techniques," *Aerospace*, vol. 8, no. 4, p. 103, Apr. 2021.
[9] J. Zhong, Y. Zhang, J. Wang, C. Luo, and Q. Miao, "Unmanned aerial vehicle flight data anomaly detection and recovery prediction based on spatio-temporal correlation," *IEEE Trans. Rel.*, vol. 71, no. 1, pp. 457–468, Mar. 2022.
[10] J. Leoni, F. Zinnari, E. Villa, M. Tanelli, and A. Baldi, "Flight regimes recognition in actual operating conditions: A functional data analysis approach," *Eng. Appl. Artif. Intell.*, vol. 114, Sep. 2022, Art. no. 105016.
[11] L. Basora, X. Olive, and T. Dubot, "Recent advances in anomaly detection methods applied to aviation," *Aerospace*, vol. 6, no. 11, p. 117, Oct. 2019.
[12] B. Wang, D. Liu, Y. Peng, and X. Peng, "Multivariate regression-based fault detection and recovery of UAV flight data," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3527–3537, Jun. 2020.

[13] C. Huang and X. Cheng, "Estimation of aircraft fuel consumption by modeling flight data from avionics systems," *J. Air Transp. Manage.*, vol. 99, Mar. 2022, Art. no. 102181.

[14] I. Levner, "Feature selection and nearest centroid classification for protein mass spectrometry," *BMC Bioinf.*, vol. 6, no. 1, pp. 1–14, 2005.

[15] Y. Huang, G. Kou, and Y. Peng, "Nonlinear manifold learning for early warnings in financial markets," *Eur. J. Oper. Res.*, vol. 258, no. 2, pp. 692–702, Apr. 2017.

[16] C. Liu, Z. Kong, S. Babu, C. Joslin, and J. Ferguson, "An integrated manifold learning approach for high-dimensional data feature extractions and its applications to online process monitoring of additive manufacturing," *IISE Trans.*, vol. 53, no. 11, pp. 1215–1230, 2021.

[17] J. X. Leon-Medina, M. Anaya, F. Pozo, and D. Tibaduiza, "Nonlinear feature extraction through manifold learning in an electronic tongue classification task," *Sensors*, vol. 20, no. 17, p. 4834, Aug. 2020.

[18] J. X. Leon-Medina, M. Anaya, F. Pozo, and D. A. Tibaduiza, "Application of manifold learning algorithms to improve the classification performance of an electronic nose," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2020, pp. 1–6.

[19] H. Lee, H. J. Lim, and A. Chattopadhyay, "Data-driven system health monitoring technique using autoencoder for the safety management of commercial aircraft," *Neural Comput. Appl.*, vol. 33, no. 8, pp. 3235–3250, Apr. 2021.

[20] D. J. Barry, "Estimating runway veer-off risk using a Bayesian network with flight data," *Transp. Res. C, Emerg. Technol.*, vol. 128, Jul. 2021, Art. no. 103180.

[21] M. Aksoy, O. Ozdemir, G. Guner, B. Baspinar, and E. Koyuncu, "Flight trajectory pattern generalization and abnormal flight detection with generative adversarial network," in *Proc. AIAA Scitech Forum*, Jan. 2021, p. 0775.

[22] B. Wang, D. Liu, X. Peng, and Z. Wang, "Data-driven anomaly detection of UAV based on multimodal regression model," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2019, pp. 1–6.

[23] H. Lee, G. Li, A. Rai, and A. Chattopadhyay, "Real-time anomaly detection framework using a support vector regression for the safety monitoring of commercial aircraft," *Adv. Eng. Informat.*, vol. 44, Apr. 2020, Art. no. 101071.

[24] A. T. Rodrigues et. al. "Data collected with package delivery quadcopter drone," Carnegie Mellon Univ., Jul. 2020, pp. 1–15, doi: 10.1184/R1/12683453.v3.

[25] T. A. Rodrigues et al., "In-flight positional and energy use data set of a DJI Matrice 100 quadcopter for small package delivery," *Sci. Data*, vol. 8, no. 1, pp. 1–8, 2021.

[26] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statist.*, vol. 46, no. 3, pp. 175–185, Feb. 2012.

[27] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1997.

[28] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[29] M. Qi, W. Chen, Y. Wang, Z.-M. Ma, and T.-Y. Liu, "Convergence analysis of distributed stochastic gradient descent with shuffling," *Neurocomputing*, vol. 337, pp. 46–57, Apr. 2019.

[30] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[31] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *J. Mach. Learn. Res.*, vol. 4, pp. 119–155, 2003.

[32] D. Baughman and Y. Liu, "Classification: Fault diagnosis and feature categorization," in *Neural Networks in Bioprocessing and Chemical Engineering*. New York, NY, USA: Academic, 1995, pp. 110–171.

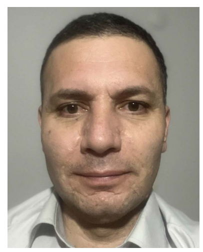
[33] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.

[34] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

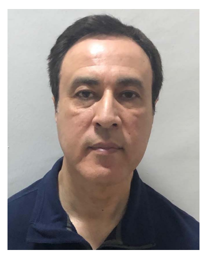
[35] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, pp. 119–139, Aug. 1995.

[36] J. Chen, S. J. Deng, and X. Huo, "Electricity price curve modeling and forecasting by manifold learning," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 877–888, Aug. 2008.

[37] N. Malini and M. Pushpa, "Analysis on credit card fraud identification techniques based on KNN and outlier detection," in *Proc. 3rd Int. Conf. Adv. Electr., Electron., Inf., Commun. Bio-Inform. (AEEICB)*, Feb. 2017, pp. 255–258.



Ümit Çelik received the bachelor's degree in computer engineering from Çankaya University and the master's and Ph.D. degrees in materials science and engineering from Istanbul Technical University. Before joining Firat University, Elaziğ, Turkey, he was a Post-Doctoral Fellow at the University of California at Davis (UC Davis). He is currently an Assistant Professor with the Department of Aircraft Maintenance and Repair, Firat University.



Haluk Eren received the B.Sc., M.Sc., and Ph.D. degrees in electrical and electronic engineering from Firat University (FU), Turkey, in 1989, 2002, and 2011, respectively. In 2008, he went to Ohio University with a Ph.D. Research Scholarship with the Department of Electrical Engineering and Computer Science (EECS). From 2009 to 2010, he was a Researcher at the Photogrammetric Computer Vision Laboratory (PCVLab), OSU. Later, he was appointed as an Assistant Professor with the Computer Science of FU in 2011, and was associated with the Digital Forensics Engineering Department with courtesy. He is currently an Assistant Professor at the School of Civil Aviation, Department of Air Traffic Control, FU. His research interests include computer vision, intelligent transportation, biomedical/social/ecosystem informatics, air traffic control, and smart city. He was an Associate Editor of IEEE ITSC in 2018, ITS Newsletter in 2016, and IEEE ACCESS to date. He has been on the editorial board of IEEE ITS Podcast since 2015.