Yayın Geliş Tarihi (Submitted): **11/03/2024**

Yayın Kabul Tarihi (Accepted): **14/12/2024**

Makele Türü (Paper Type): **Araştırma Makalesi – Research Paper**

Please Cite As/Atıf için:
Örnek, Ö., Günal, E. Ş. and Yazıcı, A. (2024), On anomaly detection for autonomous transfer vehicles in smart factories, *Nicel Bilimler Dergisi*, 6(2), 208-227. *doi: 10.51541/nicel.1450906*

ON ANOMALY DETECTION FOR AUTONOMOUS TRANSFER VEHICLES IN SMART FACTORIES

Özlem ÖRNEK[1], Efnan ŞORA GÜNAL[2] and Ahmet YAZICI[3]

ABSTRACT

Today, autonomous transfer vehicles (ATVs) have important roles in many smart factories. Therefore, flawless and uninterrupted operation of ATVs is required for the sake of effective production in smart factories. For this reason, it is important to detect anomalies (or, abnormalities) regarding ATVs during the operation. Therefore, this study aims to detect anomalies regarding ATVs so that possible losses during production can be prevented. For this purpose, two novel methods are proposed to detect anomalies for ATVs. The first method employs exhaustive feature selection to obtain the optimal subset of features for detecting anomalies. The other method utilizes a 2-stage hybrid approach for anomaly detection. Four types of anomalies (overdue pick-up delivery activity, unexpected pedestrian density, unexpected vehicle slow-down, and unexpected vehicle behavior) are considered for this work. During the experimental work, a test environment has been established for simulating a smart factory. The experimental results indicate that the first method provides a higher accuracy whereas the second one offers a better false-negative rate in detecting anomalies regarding ATVs.

Keywords: Anomaly detection, Autonomous transfer vehicles, Smart factories.

[1] Corresponding Author, Res. Asst., Department of Computer Engineering, Faculty of Engineering and Architecture, Eskisehir Osmangazi University, Eskişehir, Türkiye. ORCID ID: https://orcid.org/0000-0002-8775-8695
[2] Assoc. Prof. Dr., Department of Computer Engineering, Faculty of Engineering and Architecture, Eskisehir Osmangazi University, Eskişehir, Türkiye. ORCID ID: https://orcid.org/0000-0001-6236-174X
[3] Prof. Dr., Department of Computer Engineering, Faculty of Engineering and Architecture, Eskisehir Osmangazi University, Eskişehir, Türkiye. ORCID ID: https://orcid.org/0000-0001-5589-2032

AKILLI FABRİKALARDA OTONOM TAŞIYICI ARAÇLARINDA ANOMALİ TESPİTİ

ÖZ

Günümüzde otonom taşıyıcı araçların (OTA) birçok akıllı fabrikada önemli rolleri var. Bu nedenle akıllı fabrikalarda etkin üretim için OTA'ların kusursuz ve kesintisiz çalışması gerekmektedir. Bunu sağlamak için OTA'lara ilişkin anomalilerin (veya anormalliklerin) operasyon sırasında tespit edilmesi önemlidir. Bu amaçla bu çalışmada OTA'lara ilişkin anormalliklerin tespit edilerek üretim sırasında olası kayıpların önlenmesi amaçlanmaktadır. OTA'lardaki anomalilerin tespiti için iki yöntem önerilmiştir. İlk yöntem, anomalilerin tespiti için en uygun özellik alt kümesini elde etmek amacıyla kapsamlı özellik seçimini kullanır. Diğer yöntemde anomali tespiti için iki aşamalı hibrit yaklaşım kullanılır. Bu çalışma için dört tür anomali (gecikmiş teslim alma faaliyeti, beklenmeyen yaya yoğunluğu, beklenmeyen araç yavaşlaması ve beklenmeyen araç davranışı) dikkate alınmıştır. Deneysel çalışma sırasında akıllı fabrikanın simüle edilebilmesi için bir test ortamı oluşturulmuştur. Deneysel sonuçlar, OTA’lara ilişkin anomalilerin tespitinde birinci yöntemin daha yüksek doğruluk sağladığını, ikincisinin ise daha iyi bir yanlış negatif oranı sunduğunu göstermektedir.

Anahtar Kelimeler: Anomali tespiti, Otonom taşıyıcı araçlar, Akıllı fabrikalar.

1. INTRODUCTION

Data-driven analysis, planning or decision-making processes are important in many sectors together with digital transformation. Together with Industry 4.0, which is referred to as digital transformation in production, autonomous robot technology is predicted to play a critical role in smart factories (Lasi et al., 2014). With this technology, data-driven processes with less human interaction have become critical in smart factories. In the factory environment, it is envisaged that all goods and materials are transported by autonomous transfer vehicles (ATV) without human intervention. While ATVs perform internal logistics transport task, which is one of the main components of production, problems can occur. And it is important to identify these problems through data analysis. For this reason, it is important to detect unusual events that cause delays in tasks and to take the necessary measures ahead, with the data collected from ATVs and traffic networks in the factory environment. Also referred to as anomaly detection in this area, an approach to identify abnormal conditions that do not comply with the expected behavior can be used (Örnek et al., 2018). By monitoring ATVs and data from the environment, comparing them with normal state characteristics, which do not include problems, errors within the factory can be detected via anomaly detection.

In general, anomalies are caused by mechanical failures, changes in system behavior, human/tool failure, or natural deviations in data from the environment (Hodge, 2004). In general, anomalies are divided into three types as point, contextual and collective anomalies (Chandola, 2009). In point anomalies, a single data instance is abnormal. In contextual anomalies, a single instance of data is abnormal in a context. In collective anomalies, a collection of related data samples is abnormal. Individual situations within a collective abnormality are not abnormal in themselves (Chandola, 2009).

In critical safety environments, anomaly detection is important to identify abnormal operating conditions. Anomalies in ATV operations can arise, potentially leading to significant consequences if left undetected. These consequences are mainly related to production delays, safety hazards, and financial impacts. Undetected anomalies can disrupt production processes, leading to costly downtime. For example, an ATV malfunction might halt material flow, creating bottlenecks in the production line. This downtime translates directly into financial losses due to lost production output. Also, anomalies in ATV behavior pose serious risks to factory personnel and equipment. Malfunctioning ATVs could collide with workers, leading to injuries. Similarly, erratic ATV movements might damage other equipment, resulting in repair costs and further production delays. Besides, the financial implications of undetected anomalies extend beyond production delays. Repairing damaged equipment, covering medical expenses for injured personnel, and potential legal liabilities arising from accidents all contribute to significant monetary losses. Thus, thanks to anomaly detection, precautions can be taken before potential system failures cause undesirable results and production interruptions.

For the anomaly detection system, classification algorithms can be utilized to accurately model the distribution of data that is taken from the environment. For accurate training during classification, the data should include all possible situations. Many algorithms define boundaries around normality during classification and automatically generate a threshold (Hodge, 2004). In classification-based anomaly detection, a classification model is created for normal or abnormal events based on labeled training data. Classification based methods used in anomaly detection include support vector machines (SVM), neural networks, Bayesian networks based, rule-based algorithms and fuzzy logic. Examples of rule-based algorithms are decision tree algorithms.

Identifying abnormal conditions of ATVs fundamentally corresponds to anomaly detection in traffic networks. There are studies in the literature using classification-based methods for the detection of anomalies in traffic networks (Barria and Thajchayapong, 2011; Chen et al., 2010; Kinoshita et al., 2014; La-inchua et al., 2013; Raiyn and Toledo, 2014). In Pan and Wu (2017), speed, acceleration and lane change data were used with the support vector machine (SVM) to effectively detect abnormal traffic conditions. Flow, occupancy and average speed data were used with SVM, Principal Component Analysis and Genetic Algorithm for the detection of abnormal traffic situations (Min et al., 2017). In another study, speed, occupancy rate and traffic flow data were used with SVM to detect incidents on urban roads (Chlyah et al., 2016). In Gakis et al., (2014), speed, occupancy and flow data were used with SVM for automatic incident detection. Classification-based anomaly detection approaches can be used to detect abnormal events in traffic networks as well (Chen and Wang et al., 2009; Jiang et al., 2010; Liu et al., 2014; Lu et al., 2014; Payne and Tignor, 1978). In Ohe et al., (1995), changes in the amount of the data, speed, and occupancy were used with neural networks to determine the traffic incidents such as accidents, stopping vehicles and obstacles on the road. Zhang et al., (2011) realized traffic incident detection using speed, position and travel time attributes with backpropagation (BP) neural networks. In Z. Zhou and L. Y. Zhou (2010), SVM, utilizing occupancy, speed, upstream detection point flow rate and downstream detection point flow rate as attributes, was used to detect traffic incidents. Traffic incident detection has been performed by using fuzzy logic and genetic algorithm together with occupancy and volume data in Srinivasan et al., (2001). In another work, anomaly detection for ATVs in smart factories was realized by using the C4.5 decision tree algorithm with time zone, mean speed, road id, road type, road content, activity status, speed limit, road status, pedestrian density, and processing time data (Örnek et al., 2018). Örnek et al., (2020) stated the anomaly status as “present” or “absent” with fuzzy logic-based anomaly detection for ATV. Değirmenci et al., (2020) determined the change in the characteristics of the intersections used by ATVs in smart factories as an abnormal situation. Zhu et al., (2021) used meteorological data and traffic measures in fuzzy logic-based algorithm for traffic incident detection. The meteorological data include different types of weather conditions, including sunshine, heavy fog, and rainstorm. And traffic measures include volume, occupancy, and speed. Xie et al., (2022) proposed a method based on ensemble learning algorithm with using volume, speed and occupancy data for traffic incident detection. ElSahly et al., (2023), used Random Forest algorithm for traffic incident detection using flow rate, speed, and occupancy at upstream and downstream stations. Also, they considered congestion levels, incident severity, incident location, and detector distance while incident detection. Zhu et al., (2024) proposed a Network Lasso (NL) based decentralized learning framework for traffic incident detection using traffic features like flow, speed, and occupancy.

Regarding the previous studies, two novel methods are proposed in this work to detect anomalies of ATVs in smart factories so that possible losses during production can be prevented. While the first method employs exhaustive feature selection to obtain the optimal subset of features for the detection of anomalies, the other method utilizes a 2-stage hybrid approach for anomaly detection. Four types of anomalies (overdue pick-up delivery activity, unexpected pedestrian density, unexpected vehicle slow-down, and unexpected vehicle behavior) are considered for this work. For the experimental work, a test environment has been established for simulating a smart factory. The experimental results indicate that the first method provides a higher accuracy whereas the second one offers a better false-negative rate in detecting anomalies regarding ATVs.

The organization of the remaining of the study is as follows: The proposed methods are introduced in Section 2. The experiments and experimental results are given in Section 3. Finally, the conclusions of the paper are provided in Section 4.

2. PROPOSED METHODS

In this paper, two novel methods are proposed to detect anomalies of Autonomous Transfer Vehicles in Smart Factories. The first method employs exhaustive feature selection to obtain the optimal subset of features for the detection of anomalies. The other method utilizes a 2-stage hybrid approach for anomaly detection. In the following subsections, the test environment and dataset are first explained. Then, the proposed methods are introduced.

2.1. Test Environment and Dataset

A test environment has been established for simulating a smart factory at Smart Factory and Robotics Laboratory of Eskişehir Osmangazi University. The test environment contains, workstations, industrial robot arm and shelves for receiving and delivering operations, ATV roads, traffic signs and pedestrian crossings for transportation. A sample photograph of the laboratory is shown in Figure 1.

Figure 1. Smart factory and robotics laboratory

GAZEBO was used to create the virtual test environment in the laboratory and OpenStreetMap (OSM) standard was used to create HD Map. Figure 2 gives a visual of the Smart Factory virtual environment created in the GAZEBO simulation environment. Figure 3 shows the environment HD Map visual.

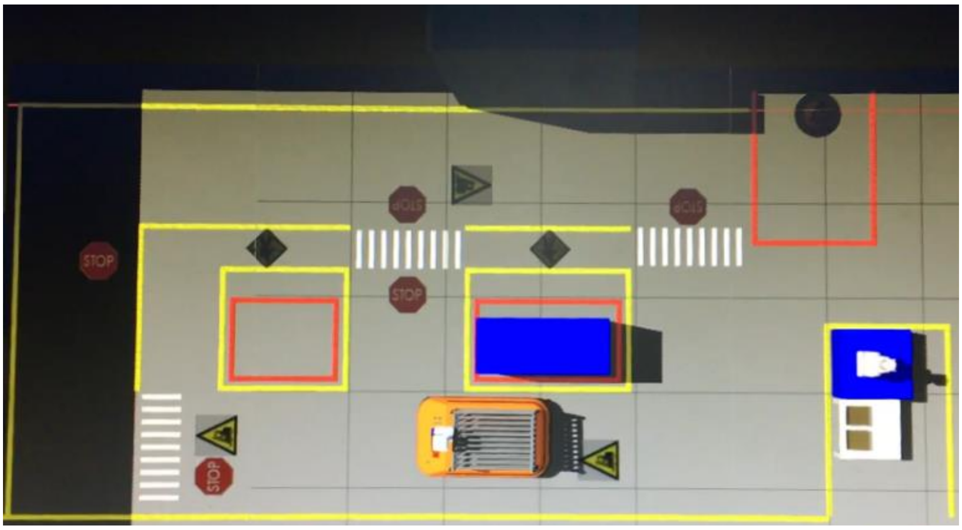

Figure 2. GAZEBO smart factory simulation

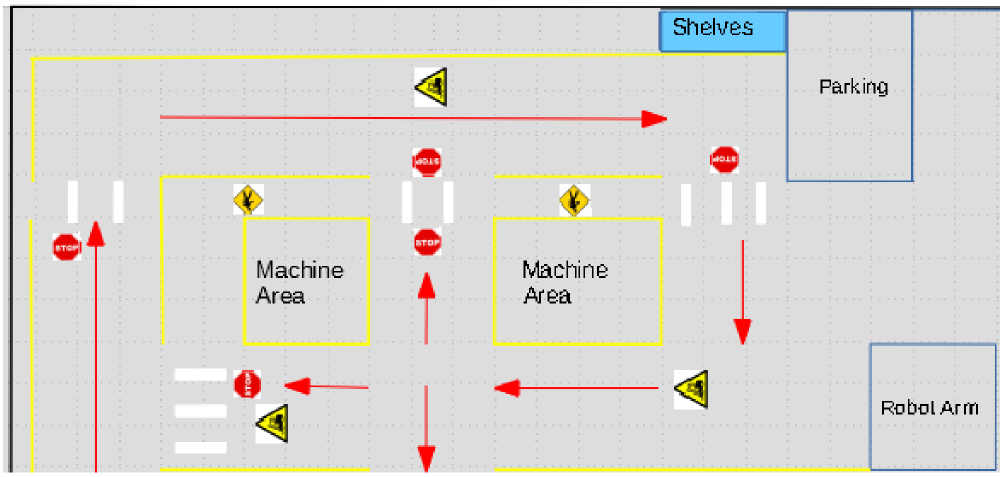

Figure 3. Smart factory hd map

Abnormalities in the internal traffic network can occur for many different reasons. In this study, four types of anomalies were considered. Abnormal conditions, anomaly descriptions and anomaly types are listed in Table 1.

Table 1. List of anomalies

Anomaly	Anomaly Definition	Anomaly Type
Stopping	Overdue Pick-up Delivery Activity	Pick-up Delivery
Slow-Down	Unexpected Pedestrian Density	Pedestrian Density
	Unexpected Vehicle Slow-down	Vehicle Slow-down
Unknown	Unexpected Vehicle Behavior	Vehicle

The dataset was obtained by simulation. The data were taken from two road sections with “pedestrian path” or “filling and unloading point” as they are representative of critical areas in the simulated factory layout. The data were collected in 15-minute intervals as this frequency is mostly aligned with typical ATV operation cycles and potential anomaly occurrences considering the size of the factory environment.

In the proposed method, eight attributes were used to detect types of anomalies. These attributes and their descriptions are listed in Table 2. These attributes provide a multi-faceted view of ATV performance, enabling the detection of various anomaly types. The time zone attribute captures the temporal context of ATV activity. By dividing the day into 15-minute intervals, the system can learn typical patterns of behavior within each time slot and identify deviations as potential anomalies. For example, unusual ATV slowdowns during peak operating hours could indicate a problem. The mean speed attribute directly reflects ATV movement efficiency. Unexpectedly low mean speeds could suggest congestion, obstacles, or mechanical issues. Conversely, unusually high speeds could signal reckless driving or control system malfunctions. The road ID attribute identifies the specific road segment where the ATV is operating. By associating data with road IDs, the system can account for variations in expected behavior based on road characteristics, such as the presence of pedestrian crossings or loading zones. The road content attribute indicates whether the road segment includes a pedestrian path or a pick-up/delivery point. This distinction is crucial because it influences the expected interactions between ATVs and other entities in the environment. For instance, slower speeds and more frequent stops would be anticipated near pedestrian paths and loading zones. The activity state attribute denotes whether the ATV is currently engaged in a pick-up/delivery activity or not. This information helps distinguish between normal stops for loading/unloading and unexpected stops due to anomalies. The road state attribute signals whether the road segment is open or closed. This is critical for identifying situations where ATVs are attempting to access restricted areas or encountering unexpected road closures. The pedestrian density attribute quantifies the number of pedestrians present on the road segment containing a pedestrian path. High pedestrian density can lead to ATV slowdowns and increased stopping frequency, so this attribute helps differentiate between normal fluctuations in pedestrian traffic and unexpected congestion. Finally, the process time attribute measures the time taken for pick-up/delivery operations at designated points. Unusually long processing times could indicate delays in loading or unloading procedures, suggesting potential issues with equipment or personnel. By analyzing these attributes in combination, the anomaly detection system can identify patterns that deviate from expected behavior and pinpoint potential anomalies in ATV operation within the simulated smart factory environment.

In the training dataset, there are 937, 477, 362, 176 and 92 instances respectively belonging to “normal”, “anomaly-vehicle”, “anomaly-vehicle-slow-down”, “anomaly-pedestrian-density” and “anomaly-pick-up-delivery” classes. In the test dataset, there are 159, 124, 66, 50 and 63 instances, respectively.

Table 2. Attributes for anomaly detection

Attribute Name	**Value**	**Description**
Time Zone	(08:00 – 16:00)	15-minute intervals
Mean Speed	(0-20)	Mean speed value for the road segment
Road ID	(I/II)	The ID of the path
Road Content	(0/1)	Information that the road segment contains. If the value is 0, it includes the pedestrian path, if 1, it contains the pick-up/delivery.
Activity State	(0/1)	Pick-up/delivery activity status information within the road segment. If the value is 0, it is passive, if it is 1, it is active.
Road State	(0/1)	Information whether the road is open or closed. If the value is 0, it is closed, if it is 1, it is open.
Pedestrian Density	(0-50)	Pedestrian density value of the road segment containing the pedestrian path.
Process Time	(0-15)	The processing time value for the path segment with pick-up/delivery.

ATVs are internal logistics components that perform transport tasks in factories, and it can form a large traffic network within the factory. To manage this traffic network efficiently, it is important to detect unexpected situations (anomalies) that may occur in the network. For this purpose, Exhaustive Search based anomaly detection tests are performed to detect anomalies that may occur in the factory internal logistics traffic network. In another experiment, a hybrid anomaly detection method is proposed to increase the accuracy of anomaly detection.

2.2. Exhaustive Search-Based Anomaly Detection

Exhaustive Search is a general problem-solving technique that solves all possible situations for a solution and checks whether each situation provides the problem. It is also known as brute-force search or generate and test. It is easy to apply. If there is a solution, the exhaustive search will always find it. However, the cost is proportional to the number of possible situations, and in some problems, the size of the situations can grow very quickly. It is therefore used for problems where the number of states is limited.

The four types of anomalies given in Table 1 and the absence of anomaly (labeled as ‘normal’) are considered as a 5-class classification problem. Using eight attributes in Table 2 for Exhaustive Search, 255 different attribute subsets were obtained. During the attribute selection process, classification is performed using Naive Bayes, Support Vector Machine by Sequential Minimal Optimization (SMO (SVM)), J48, Random Tree, Random Forest, Partial Decision (PART) and Logistic Model Tree (LMT) classification algorithms. For each classification algorithm used with attribute selection, it is aimed to achieve anomaly detection with the highest possible accuracy. Figure 4 shows the process diagram of the proposed method.

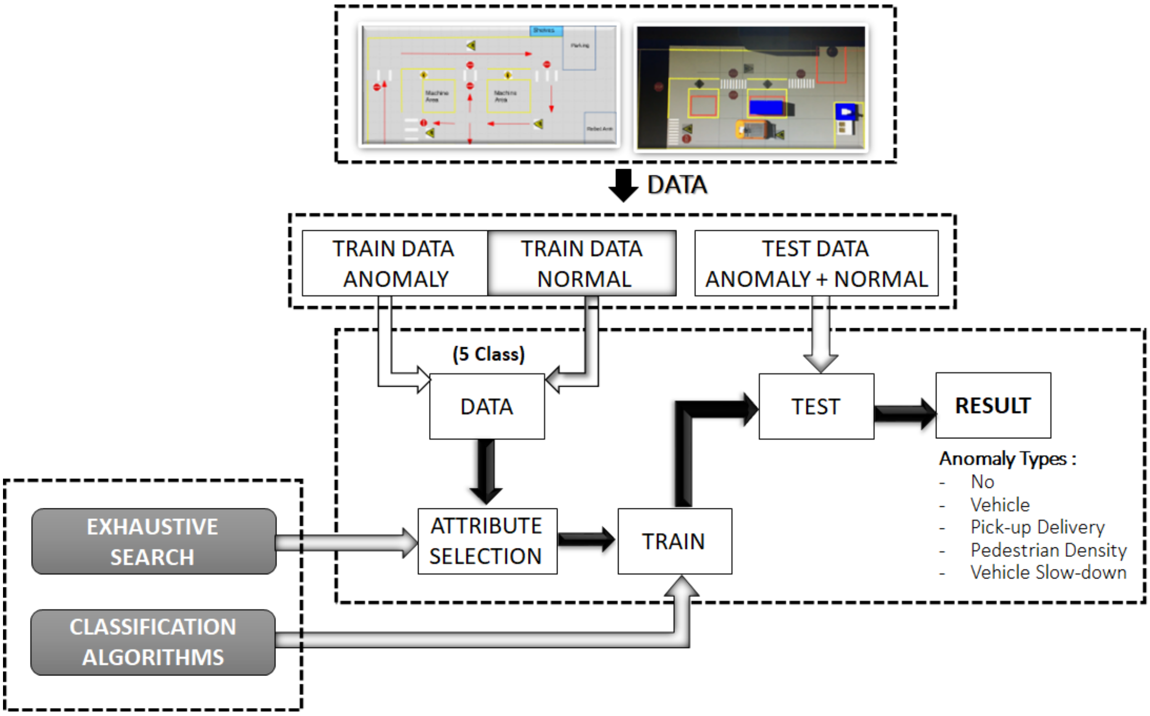

Figure 4. Process diagram of exhaustive search-based anomaly detection

2.3. A Hybrid Approach to Anomaly Detection

Hybrid anomaly detection takes place in two stages. In the first stage, the presence of anomaly is determined, while the type of anomaly is determined in the second stage. The first stage is considered as a 2-class classification problem, that is, the conditions to be detected are corresponding to “anomaly” and “normal” classes. Class “anomaly” indicates the presence of an anomaly whereas class “normal” indicates that there is no anomaly. Thus, at the first stage, it is determined whether there is an anomaly in the ATV or whether it is a normal situation. After this determination, in the second stage, types of anomalies of the instances, which are classified as “anomaly” at the first stage, are determined. The classes for the second stage (4-class) are “anomaly-vehicle”, “anomaly-vehicle-slow-down”, “anomaly-pedestrian-density” and “anomaly-pick-up-delivery”. In this method, previously mentioned classification algorithms are used too.

To prevent misclassification that may occur due to the differences of anomaly types, hybrid anomaly detection aims to detect more anomalies in the first stage. Then, in the second stage, it is aimed to make more accurate detection of an anomaly by just focusing on anomaly types. Figure 5 illustrates the process diagram of the proposed method.

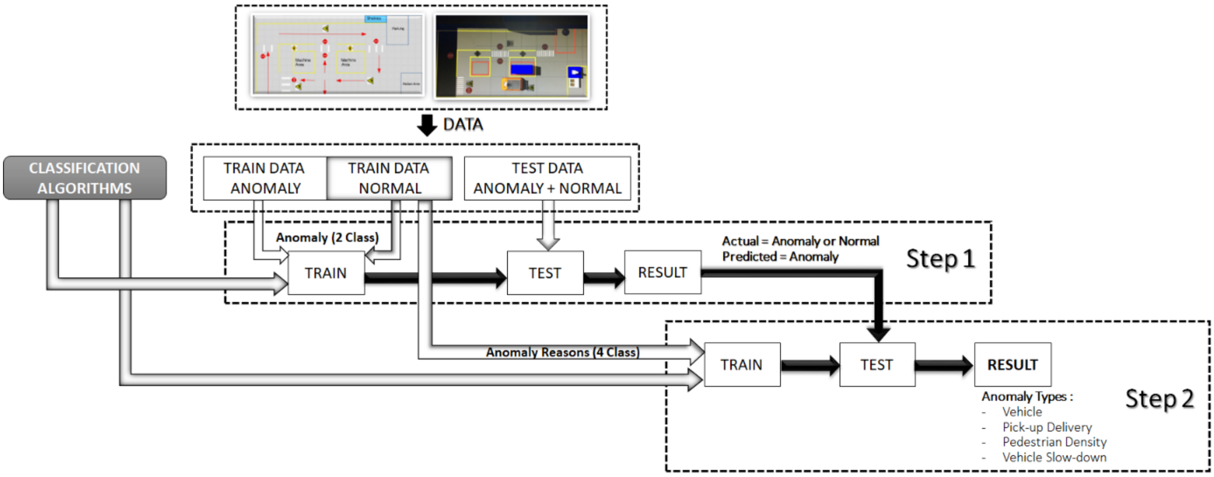

Figure 5. Process diagram of Hybrid anomaly detection

3. EXPERIMENTAL WORK

Using the data obtained by simulating the movements of ATVs, using 8 different attributes from the experimental environment, machine learning algorithms were used to infer whether the movements are normal or anomaly. For this purpose, trainings were carried out with different classification algorithms when all 8 different attributes obtained from the experimental environment were used. After this experimental study, to determine which of the 8 attributes has the highest discrimination in terms of classification, the highest performance with the least attribute was tried to be handled with the exhaustive search method. After that, a hybrid method was proposed by an independent experimental study, which classified the normal-anomaly type in the first stage and the anomaly type in the second stage. The results are presented comparatively.

3.1. Initial Results without Feature Selection

In the first phase of the experiments, initial classification results were obtained using the initial feature set without feature selection for 5-class classification. The corresponding results are provided in Table III. According to Table III, in the case of using eight attributes to determine the behavior of the ATV, Random Tree Classifier gives the best Accuracy (Acc) and False Negative Rate (FNR) values. Since a low FNR value means less missed values, a small numerical value will show that the classifier has a high performance in this respect.

Table 3. Initial classification results (5 Class problem) (without feature selection)

Classification Algorithm	***Acc (%)***	***Average FNR***
Naive Bayes	60,4	0,28
SMO (SVM)	57,4	0,50
J48	83,8	0,20
Random Tree	**84,6**	**0,19**
Random Forest	82,3	0,23
PART	83,5	0,24
LMT	74,7	0,37

3.2. Experimental Results for Exhaustive Search-Based Anomaly Detection (5-class)

Next, with the help of an exhaustive search, 255 different combinations of 8 attributes were obtained. Then, these combinations were tested with the classification algorithms. The highest general Acc rates and anomaly classes average FNR obtained with Naive Bayes, SMO, J48, Random Tree, Random Forest, PART and LMT algorithms and the respective attributes are given in Table 4. The highest Acc among the algorithms was obtained by PART with 7 attributes. According to the results of the PART classifier with exhaustive search, all attributes except PT attribute were found to contribute.

Table 4. Exhaustive search based anomaly detection results

Classification Algorithm	**Attributes**								***Acc (%)***	***FNR (%)***
	Time Zone	Mean Speed	Road ID	Road Content	Activity State	Road State	Pedestrian Density	Process Time		
Naive Bayes	✓	✓			✓	✓	✓	✓	64,10	0,26
SMO (SVM)	✓	✓			✓	✓	✓	✓	60,90	0,53
J48	✓	✓				✓	✓	✓	83,80	0,20
Random Tree	✓	✓	✓	✓	✓	✓	✓	✓	84,60	**0,19**
Random Forest	✓	✓	✓	✓	✓	✓	✓	✓	82,30	0,23
PART	✓	✓	✓	✓	✓	✓	✓		**85,90**	0,22
LMT	✓	✓	✓			✓		✓	77,70	0,30

When we consider our problem as a 5-class problem, the Random Tree algorithm provides the highest performance. Comparing the experimental results in sections A and B, the highest success rate is obtained when all 8 of the attributes are used in the Random Tree and Random Forest classifiers. On the other hand, when the exhaustive search method is applied with the decision tree based J48 classifier, we still get the same success rate (83.2) when 5 attributes are used instead of 8. Moreover, Naive Bayes, SMO, PART, LMT algorithms provide higher performance rates by using fewer attributes. Of these, Naive Bayes provides 6% improvement with 6 attributes, 6% improvement with SMO (SVM) and 6 attributes, 3% improvement with PART and 7 attributes, and 4% improvement with LMT and 5 attributes. When the table is considered as a whole, the highest score among all success rates was the PART classifier with 7 attributes and a success rate of 85.9%. Except for J48, Random Forest and Random Tree classifier algorithms, in all other classifier algorithms, the optimal search operation increased the success rates by identifying the attributes with the highest distinctiveness for the classifiers. When using 8 attributes, the highest performance was obtained as 84.6%, while with the application of exhaustive search algorithm, it increased to 85.9% with the use of 7 attributes.

3.3. Experimental Results for Hybrid Anomaly Detection

In the method, which consists of two stages, in the first stage, all the data in the training and test datasets have been used. In the training dataset to be used for the first stage of Hybrid Anomaly Detection for 2 Class, there are 937 and 1107 instances respectively belonging to “normal” and “anomaly” classes. In the test dataset to be used for the first stage of Hybrid Anomaly Detection for 2 Class, there are 159 and 303 instances respectively belonging to “normal” and “anomaly” classes. As a result of the first classification, the data was classified as “anomaly” or “normal”.

In the second stage, only the “anomaly” class data from the training data set used in the first stage was taken according to the anomaly types for training. For the test, the instances which are the "anomaly" of the actual and predicted value of the class because of the first stage test constitute the second stage test data set. At this stage, there are 477, 362, 176 and 92 data belonging to “anomaly-vehicle”, “anomaly-vehicle-slow-down”, "anomaly-pedestrian density” and “anomaly-pick-up-delivery” classes, respectively. The test dataset is not constant since it depends on the first stage.

For the first stage, experimental studies were carried out with Naïve Bayes, SMO, J48, Random Tree, Random Forest, PART and LMT classifiers used, and the results are given in Table 5. In this table, average Acc values are given and, while False Negative Error Rate

(FNR) is given for the value of the samples found to be absent while the anomaly was present.

In our study, it is important to minimize “false negative" (missed) values in the “anomaly” class. Therefore, in addition to Acc values, FNR values by class were also taken into consideration. Accordingly, LMT, which has a high Acc and the second lowest FNR, can be the best choice.

Table 5. Hybrid anomaly detection step 1 classification results

Classification Algorithm	**Evaluation Metrics**	
	Step 1	
	Acc (%)	***FNR (%)***
Naive Bayes	65,8	0,271
SMO (SVM)	69,5	0,096
J48	87,2	0,119
Random Tree	87,0	**0,063**
Random Forest	87,0	0,066
PART	88,5	0,089
LMT	**89,4**	0,116

In the second stage, Naïve Bayes, SMO, Random Tree and Random Forest classification algorithms were used for “vehicle”, “vehicle slow-down”, "pedestrian density” and “pick-up delivery” classes. With the classification algorithms, experimental studies were carried out using the training data set and the test data sets obtained from the first stage. Accurate estimates of “anomaly" class from the results of all classification algorithms used in the first stage constitute the second stage test datasets. Each test set obtained from the algorithms in the first stage was used as the test data set for each algorithm used in the second stage. The test dataset which was formed with the result of each algorithm used in the first stage was used as the test dataset for each algorithm in the second stage. And because of the using these test sets, the algorithms providing the highest average Acc in the second stage are given in Table 6. In this table, average Acc values are given and, while average False Negative Error Rate (FNR) is given for the value of the samples found to be different from real anomaly class. This average FNR value does not include the samples whose actual value is not an anomaly, but which is predicted as anomaly. Table 6 presents the first stage and second stage Acc together.

Table 6. Hybrid anomaly detection step 2 classification results

Step 2 Classification Algorithm	Evaluation Metrics Step 2	
	Acc (%)	*FNR (%)*
Naive Bayes	74,8	0,191
SMO (SVM)	64,5	0,356
J48	82,3	0,148
Random Tree	78,0	0,198
LMT	80,9	0,214
PART	78,0	0,197
Random Forest	**83,0**	**0,142**

The success of the second stage depends on the algorithm used in the first stage. As a result of the experiments, for each algorithm in the first stage, the highest performance rate in the second stage was obtained with the Random Forest algorithm. Different Acc and FNR values are obtained for each first and second stage algorithm. The highest performance rate and lowest FNR was obtained by LMT + Random Forest. According to Table 6, LMT and Random Forest algorithms provide the best performance rates for hybrid anomaly detection. Because it gives the best results in terms of Acc value and FNR value.

The confusion matrices obtained as a result of hybrid anomaly detection by LMT and Random Forest algorithms are given in Table 7 and Table 8 for each stage of the proposed method.

Table 7. The confusion matrix for the first stage classification (lmt) process

Actual Class	Predicted Class	
	“normal”	“anomaly”
“normal”	145	14
“anomaly”	35	268

Table 8. The confusion matrix for the second stage classification (random forest) process

Actual Class	Predicted Class			
	“anomaly-vehicle”	“anomaly-vehicle-slow-down”	“anomaly-pedestrian-density”	“anomaly-pick-up-delivery”
“normal” (incorrectly classified as “anomaly” at the first stage)	3	5	6	0
“anomaly-vehicle”	120	0	0	0
“anomaly-vehicle-slow-down”	6	56	0	0
“anomaly-pedestrian-density”	1	0	22	0
“anomaly-pick-up-delivery”	0	27	0	36

4. CONCLUSIONS

ATVs are an important part of smart factories. The detection of anomalies for ATVs within the factory with the concept of smart factories is of great importance in terms of sustainability, efficiency and safety in production within the factory. For this reason, in our work, Exhaustive Search based anomaly detection and hybrid anomaly detection studies have been performed for ATVs in smart factories. Exhaustive search-based anomaly detection experiments showed that the best results for 5-class classification were obtained with 85.9% Acc and 0,22 FNR in the PART algorithm. Then, hybrid anomaly detection was used to increase the accuracy and decrease the FNR obtained. As a result of the experiments, the presence of anomaly was determined by the LMT algorithm with 89,39% Acc and 0,116 FNR in the first stage. And then the anomaly types of the instances determined as “anomaly" class in the first stage were determined with 83,0% Acc and 0,142 FNR in the second stage with the Random Forest algorithm.

As can be seen from the results obtained by using the same data set, the hybrid anomaly detection method is more successful than the exhaustive search-based anomaly detection method. In the first stage of hybrid anomaly detection, anomalies were detected with higher Acc compared to exhaustive search. The cost of the missed (undetected) anomaly is high. And in the first and second stages of hybrid anomaly detection, the fact that FNR values were reduced compared to exhaustive search has improved the study.

By accurately identifying anomalies, the proposed methods can contribute to reduced production downtime, enhanced workplace safety, and cost savings. By promptly detecting and addressing anomalies, factories can minimize disruptions and maintain smooth production flows. Also, accurate anomaly detection helps prevent accidents involving ATVs, safeguarding factory personnel from potential injuries. Moreover, preventing equipment damage, reducing repair costs, and mitigating potential liability through effective anomaly detection leads to significant financial savings.

In future studies, the data obtained from the environment in which ATVs are used can be expanded and diversified, and the impact of additional environmental features on anomaly detection can be explored to further improve accuracy and provide a more comprehensive understanding of ATV behavior.

ETHICAL DECLARATION

This work is supported by the Scientific and Technical Research Council of Turkey (TUBITAK), Contract No 116E731, project title: "Development of Autonomous Transport Vehicles and Human-Machine / Machine-Machine Interfaces for Smart Factories" and the Scientific and Technical Research Council of Turkey (TUBITAK), Program Name 2209-B - Undergraduate Thesis Support Program for Industrial Oriented, project title: "Anomaly Detection for Autonomous Transporter Vehicles in Smart Factories".

REFERENCES

Barria, J. A. and Thajchayapong, S. (2011), Detection and classification of traffic anomalies using microscopic traffic variables, *IEEE transactions on intelligent transportation systems*, 12(3), 695-704.

Barros, R. C., De Carvalho, A. C. and Freitas, A. A. (2015), *Automatic design of decision-tree induction algorithms*, Springer International Publishing, Germany.

Breiman, L. (2001), *Random forests*, Machine learning, Kluwer Academic Publishers, Netherlands, 45(1), 5-32, 2001.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., (1983), *Classification and regression trees*. Routledge, New York.

Chandola, V., Banerjee, A. and Kumar, V. (2009), Anomaly detection: A survey, *ACM computing surveys (CSUR)*, 41(3), 1-58.

Chen, L., Cao, Y. and Ji, R. (2010), Automatic incident detection algorithm based on support vector machine, *In 2010 Sixth International Conference on Natural Computation* , 2, 864-866, IEEE.

Chen, S. and Wang, W. (2009), Decision tree learning for freeway automatic incident detection, *Expert systems with applications*, 36(2), 4101-4105.

Chlyah, M., Dardor, M. and Boumhidi, J. (2016), Multi-agent system based on support vector machine for incident detection in urban roads, *In 2016 11th International Conference on Intelligent Systems: Theories and Applications (SITA)*,1-6, IEEE.

Değırmencı, E., Örnek, Ö. and Yazici, A. (2020), Learning Intelligent Factory Traffic Characteristics and Anomali Detection with Contextual Multi-Arm Slot Machine, *In*

2020 28th Signal Processing and Communications Applications Conference (SIU), 1-4, IEEE.

ElSahly, O. and Abdelfatah, A. (2023), An incident detection model using random forest classifier, *Smart Cities*, 6(4), 1786-1813.

Gakis, E., Kehagias, D. and Tzovaras, D. (2014), Mining traffic data for road incidents detection, *In 17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 930-935, IEEE.

Hodge, V. and Austin, J. (2004), A survey of outlier detection methodologies, *Artificial intelligence review*, 22, 85-126.

Jiang, G., Niu, S., Li, Q., Chang, A. and Jiang, H. (2010), Automated incident detection algorithms for urban expressway, *In 2010 2nd International Conference on Advanced Computer Control*, 3, 70-74, IEEE.

Kinoshita, A., Takasu, A. and Adachi, J. (2014), Real-time traffic incident detection using probe-car data on the Tokyo Metropolitan Expressway, *In 2014 IEEE International Conference on Big Data (Big Data)*, 43-45, IEEE.

La-inchua, J., Chivapreecha, S. and Thajchayapong, S. (2013), A new system for traffic incident detection using fuzzy logic and majority voting, *In 2013 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 1-5, IEEE.

Landwehr, N., Hall, M. and Frank, E. (2005), Logistic model trees, *Machine learning*, 59, 161-205.

Lasi, H., Fettke, P., Kemper, H. G., Feld, T. and Hoffmann, M. (2014), Industry 4.0, *Business and information systems engineering*, 6, 239-242.

Liu, Q., Lu, J., Chen, S. and Zhao, K. (2014), Multiple Naïve bayes classifiers ensemble for traffic incident detection, *Mathematical Problems in Engineering*.

Lu, J., Liu, Q., Yuan, L., and Chen, S. (2014), Grafted Decision Tree for Freeway Incident Detection, *In CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems*, 467-477.

Maimon, O. Z., and Rokach, L. (2014), *Data mining with decision trees: theory and applications*, 81, Second edition, 328, World Scientific Publishing Co., Inc., USA.

Min, Z., Yanlei, L., Dihua, S. and Senlin, C. (2017), Highway traffic abnormal state detection based on PCA-GA-SVM algorithm, *In 2017 29th Chinese Control And Decision Conference (CCDC)*, 2824-2829, IEEE.

Ohe, I., Kawashima, H., Kojima, M., and Kaneko, Y. (1995), A method for automatic detection of traffic incidents using neural networks, *In Pacific Rim TransTech Conference, 1995 Vehicle Navigation and Information Systems Conference Proceedings, 6th International VNIS, A Ride into the Future*, 231-235, IEEE.

Örnek, Ö., Gülbandılar, E. and Yazıcı, A. (2020). Akıllı fabrikalardaki otonom taşiyicilar için bulanik mantik tabanli anomali tespiti, *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi,* 28(1), 53-61.

Örnek, Ö., Vatan, S., Sarıoğlu, S., and Yazıcı, A. (2018), Trafik Ağlarında Anomali Tespiti, *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi*, 26(3), 132-138.

Örnek, Ö., Vatan, S., Sarıoğlu, S. andYazıcı, A. (2018), Anomaly detection for autonomous transfer vehicles in smart factories, *In 2018 6th International Conference on Control Engineering and Information Technology (CEIT)*, 1-5, IEEE.

Pan, B. and Wu, H. (2017), Urban traffic incident detection with mobile sensors based on SVM, *In 2017 XXXIInd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)*, 1-4, IEEE.

Payne, H. J. and Tignor, S. C. (1978), *Freeway incident-detection algorithms based on decision trees with states,* Transportation Research Record, (682).

Raiyn, J. and Toledo, T. (2014), Real-time road traffic anomaly detection, *Journal of Transportation Technologies*, 4(03), 256.

Srinivasan, D., Cheu, R. L., and Poh, Y. P. (2001), Hybrid fuzzy logic-genetic algorithm technique for automated detection of traffic incidents on freeways, *In ITSC 2001, 2001 IEEE Intelligent Transportation Systems, Proceedings (Cat. No. 01TH8585),* 352-357, IEEE.

Xie, T., Shang, Q. and Yu, Y. (2022). Automated Traffic Incident Detection: Coping With Imbalanced and Small Datasets, *IEEE Access*, 10, 35521-35540.

Zhang, Z., Lin, X. and Hu, B. (2011), Algorithm design of traffic incident automatic detection based on mobile detection, *In Proceedings of 2011 IEEE International Conference on Service Operations, Logistics and Informatics*, 331-335, IEEE.

Zhou, Z. and Zhou, L. Y. (2010), An automatic incident of freeway detection algorithm based on support vector machine, *In 2010 International Symposium on Intelligence Information Processing and Trusted Computing*, 543-546, IEEE.

Zhu, C., Guo, Z. and Ke, J. (2021), Advanced fuzzy-logic-based traffic incident detection algorithm, *Advances in Fuzzy Systems*, 1-11.

Zhu, Q., Qin, A. K., Abeysekara, P., Dia, H. and Grzybowska, H. (2024), Decentralised Traffic Incident Detection via Network Lasso, *arXiv preprint arXiv:2402.18167.*