A HYBRID APPROACH FOR OUTLIER DETECTION IN PHARMACEUTICAL COLD CHAIN LOGISTICS: A CASE STUDY

Sevde Ceren Yıldız Özenç, Merve Er, Seniye Ümit Fırat

S. C. Yıldız Özenç (✉)
Department of Industrial Engineering, Institute of Pure and Applied Sciences, Marmara University, Istanbul, Turkey
Department of Industrial Engineering, Faculty of Engineering, Doğuş University, Istanbul, Turkey
e-mail: syildiz@dogus.edu.tr

M. Er
Department of Industrial Engineering, Faculty of Engineering, Marmara University, Istanbul, Turkey
e-mail: merve.er@marmara.edu.tr

S. Ü. Fırat
Department of Industrial Engineering, Faculty of Engineering, Istanbul Sabahattin Zaim University, Istanbul, Turkey
e-mail: seniye.firat@izu.edu.tr

A HYBRID APPROACH FOR OUTLIER DETECTION IN PHARMACEUTICAL COLD CHAIN LOGISTICS: A CASE STUDY

ABSTRACT

Pharmaceutical cold chain management is a complex process that involves temperature-controlled production, storage, packaging, and delivery activities. Pharmaceutical products are typically highly sensitive to environmental factors including temperature fluctuations, humidity, light and vibrations. Therefore, pharmaceutical cold chains include serious human health as well as financial and regulatory risks. The highest risk for cold chain breakage occurs during logistics activities. This study proposes a hybrid approach called Route Detection-based Support Vector Regression (RD-SVR) algorithm for detecting temperature outliers during transportation of pharmaceutical products. This classification-based framework is used for identifying cold chain breakages and safeguarding product quality. The proposed RD-SVR model is performed on a data set including one-year data logs of two vehicles gathered from one of the biggest logistics companies in Türkiye. The results are visualized and evaluated by considering the risks and routes. Unfortunately, there is very limited literature on the use of big data analytics in cold chains. Therefore, the proposed approach and example case study contributes to both theory and practice by suggesting a way for ensuring quality, safety and reliability of pharmaceutical products during transportation.

Keywords: Cold chain, Outlier detection, Pharmaceutical industry, Support vector regression

1. INTRODUCTION

A cold chain is a temperature-controlled supply chain that involves the production, storage and transportation of sensitive and perishable products such as fruits and vegetables, meat, seafood, dairy products, pharmaceuticals, and some chemicals. Cold chain storage and transportation involves many irreversible risks including product degradation and regulatory violations (Ding et al., 2023; Lin et al., 2023). Therefore, risks in cold chain storage and transportation must be thoroughly analyzed and proactive interventions should be implemented to prevent irreversible consequences. An effective cold chain management necessitates specialized packaging systems and vehicles, and continuous real time monitoring of environmental factors such as temperature, humidity, light, and vibration. The emergence of Industry 4.0 technologies and increasing digitalization of supply chains help to increase traceability of supply chain operations and develop data-driven approaches in cold chains.

Pharmaceutical cold chains, being a critical stakeholder of the cold chain, play a critical role for the healthcare industry for ensuring the safety and efficacy of medications. Pharmaceutical products are highly sensitive to environmental factors such as extreme or fluctuating temperatures and humidity. Improper environmental conditions may lead to chemical degradation, compromise the efficacy of medications, and even cause the death of humans. Therefore, all phases of pharmaceutical cold chains need to be managed effectively due to their direct and critical effects on public health.

Pharmaceutical products often require product-specific storage, packaging and transportation conditions that necessitate strict control to ensure their integrity and compliance with regulatory standards. The highest risk of temperature variations occurs during the transfer of the products. Pharmaceutical products are typically transported with refrigerated vehicles to meet the desired temperature during the route. Failure to maintain the desired temperature throughout the route, in other words, breakage of the cold chain, may cause serious health risks. Additionally, breakdown of the cold chain may lead to substantial financial risks owing to wasted inventories, regulatory penalties, and damaged reputation. The high costs of pharmaceutical products are further exacerbated by product deterioration, leading significant financial losses both for manufacturers and logistics companies. Hence, managing pharmaceutical cold chain risks is essential, as their consequences can be severe and fatal. Machine learning algorithms are suitable for risk classification and prediction in the pharmaceutical cold chain field (Konovalenko & Ludwig, 2021; Mangini et al., 2024). Outlier detection is one way to handle such risks (Zhao et al., 2013).

Transportation of temperature-sensitive pharmaceuticals is a challenging task that requires monitoring temperature deviations and detecting anomalies that could indicate a cold chain breakage. In order to contribute to this critical research area, this article employs hybrid Route Detection-based Support Vector Regression (RD-SVR) algorithm for outlier detection during transportation of temperature-sensitive pharmaceutical supply chains. Our model can be considered as a classification problem. The main objective of this research is to test if a classification-based outlier detection algorithm can be used to detect the out-of-range temperature measurements causes to breakage of the pharmaceutical cold chain, in a refrigerated transportation vehicle. The research question is “Can RD-SVR algorithm be used to detect out-of-reference temperature measurements on location-based temperature data in pharmaceutical cold chain logistics?”.

The proposed methodology is applied on data gathered from two vehicles of an international logistics company for a one-year time horizon. The developed algorithm consists of two phases. In the first phase, a route detection algorithm is developed in accordance with the sensor data. This algorithm determines all the routes of the vehicles within a one-year period, while at the same time eliminating unnecessary data. Thus, data cleaning is also performed in this phase. In the second phase, route-based outlier detection is performed using Support Vector Regression (SVR) to detect anomalous temperature of the pharmaceutical products during transportation. The advantages of using SVR can be expressed by its robustness and being convenient to multi-dimensional data, (Awad et al., 2015).

The remainder of the article is structured as follows. Section 2 presents the relevant literature. Section 3 explains the proposed approach, case study company, and data collection. Section 4 gives and discusses the results. Finally, Section 5 concludes the paper.

2. LITERATURE REVIEW

Cold chain refers to the post-production supply chain activities of temperature-sensitive products within the required temperature and humidity ranges to guarantee product safety (Joshi et al., 2009; Salin & Nayga Jr, 2003). The main definition of the cold chain can be described as controlling and monitoring the temperature of the related products along the entire supply chain (Shashi et al., 2018). Cold chain products are categorized in five groups: 1) fruits and vegetables; 2) bakery and confectionary; 3) dairy and frozen dessert; 4) meat and sea food; 5) drugs and pharmaceuticals (Allied Market Research, 2022). This classification can be categorized under two fields mainly, food cold chain and pharmaceutical cold chain (Centobelli et al., 2021).

2.1. Cold Chain Risks

A supply chain inherently involves multiple sources of risk. In a supply chain, risk can be defined as failure, damage or loss caused by unforeseen events (Zsidisin & Ritchie, 2009). There are many risk categorizations of supply chain. Yildiz Ozenc et al. (2023) provides a comprehensive literature review which integrates all risks and risk categorizations in a supply chain. One of the most important features that separate the cold chain from the supply chain is the regulatory requirements. Pharmaceutical cold chain has special regulatory requirements for transportation and storage (Ruiz-Garcia & Lunadei, 2010). Risks in the cold chain are more vital and should be kept under control. In literature, cold chain risks have been categorized in many ways. Liao et al. (2023) is stated cold chain risk as lack of infrastructure, lack of green transportation, lack of capacity building, operation problems in cold chain transport, loss of energy, temperature control and packaging. Shen & Qian (2023) identified the risks of cold supply chain as five criteria: Refrigeration risk, Supply and demand risk, Logistics risk, Information risk and Environmental risk. Lin et al. (2023) developed a hierarchy model for vaccine transportation quality and risk in cold chain with five criteria and twenty-six sub-criteria and determined the most influential sub-criteria for vaccine transportation quality and risk as local weather conditions, topography, road characteristics, cold-chain logistics standardization, and national economic development. Ding et al. (2023) identified fifteen risk factors in three risk assessment aspects, operation risk, hardware risk, and shipper risk, that threaten cold chain operations. They resulted that the most important risk aspect is shipper risk and six key factors that significantly impact cold chain logistics are excessively long container loading time, insufficient pre-cooling of goods, poor or improper product packaging, dockside or shipboard power supply system malfunction, improper cargo stacking, and human error affecting temperature settings by using the best-worst method. Arslan et al. (2023) identified five risk criteria of food cold chain. They prioritized these risks as financial risks, delivery risks, technological ability risks, environmental risks, quality risks and social risks consecutively by using Decomposed Fuzzy Set-Based FMEA approach.

As a result of the literature research conducted to identify cold chain risks, we concluded that every risk in the cold chain indicates one major risk. That risk is the breakage of the cold chain. The most important reason for the breakage of the cold chain is the situation where the temperature goes out of the limit during transportation.

2.2. Outlier Detection

Out of the limit temperatures can be interpreted as outliers in a series of data. Therefore, detecting the breakage of the cold chain can be considered as an outlier detection, also known as anomaly detection, problem. In a way, outliers can be defined as inconsistent values in a set of data (Barnett et al., 1979). There are several methodologies of outlier detection, and they are grouped in different ways in the literature (Escalante, 2005; Han et al., 2012). These methodologies can be categorized as statistical-based techniques, density-based techniques, classification-based techniques, clustering-based techniques, and also machine learning and deep learning techniques (Chandola et al., 2009). Machine learning is a useful approach in an industrial context and has a wide range of applications (Bertolini et al., 2021; Oliveira et al., 2024). Anomaly detection is one of the crucial machine learning tasks (Kang et al., 2020). Numerous studies apply machine learning techniques for anomaly detection in the existing literature (Hernandez-Jaimes et al., 2024; Mascali et al., 2023; Oliveira et al., 2024; Ravinder & Kulkarni, 2023; Olimov et al., 2022; Zhao et al., 2013). As well as machine learning algorithms, deep learning algorithms have recently become popular for anomaly detection (Abusitta et al., 2023; Seo et al., 2024; Tian et al., 2025; Zipfel et al., 2023).

Various learning algorithms such as artificial neural networks (ANNs), support vector machines (SVMs), Bayesian networks, etc., can be applied as a classification-based outlier detection method (Chandola et al., 2009). SVMs have always been a preferred method in outlier detection for various study fields (Baldomero-Naranjo et al., 2021; Thongkam et al., 2008; Y. Zhang et al., 2013). In the study, where classifiers were compared, it was concluded that these classifier models can distinguish various anomalies, and it was determined that the SVM had the highest predictive power among the classifiers (Carvalho et al., 2019). SVM can be generalized for regression (Alpaydin, 2020). This generalized version is called Support Vector Regression (SVR). SVR is a machine learning method where a model determines the significance of a variable in describing the relationship between input and output (F. Zhang & O'Donnell, 2020). SVR is used as the classification-based algorithm. Similar to classification, SVR is defined by its use of kernels, a sparse solution, and control over the margin and the number of support vectors (Awad et al., 2015).

Due to the temperature data being easy to label in a distinctive way, RD-SVR model is proposed as a classification-based outlier detection. Outlier detection helps to classify the temperature during transportation in cold chain weather if valid or invalid. The most significant aspect of the cold chain is to keep the temperature in the desired range throughout the entire supply chain activities and processes. However, in the cold chain field there are very limited number of studies which focuses on outlier detection via machine learning. There is a lack of studies using classification-based outlier detection in pharmaceutical cold chain field in best of our knowledge. This study aims to fill the gap in existing literature.

3. METHODOLOGY

This section explains the research methodology of this research study under the concept of objectives, research questions, techniques, and work plan. Based on our research question, classification-based outlier detection framework is developed as route detection-based support vector regression algorithm. Steps of the framework are planned as research planning, literature review, data collection, data mining analysis, and results. The flow chart of the developed framework is given in Figure 1.

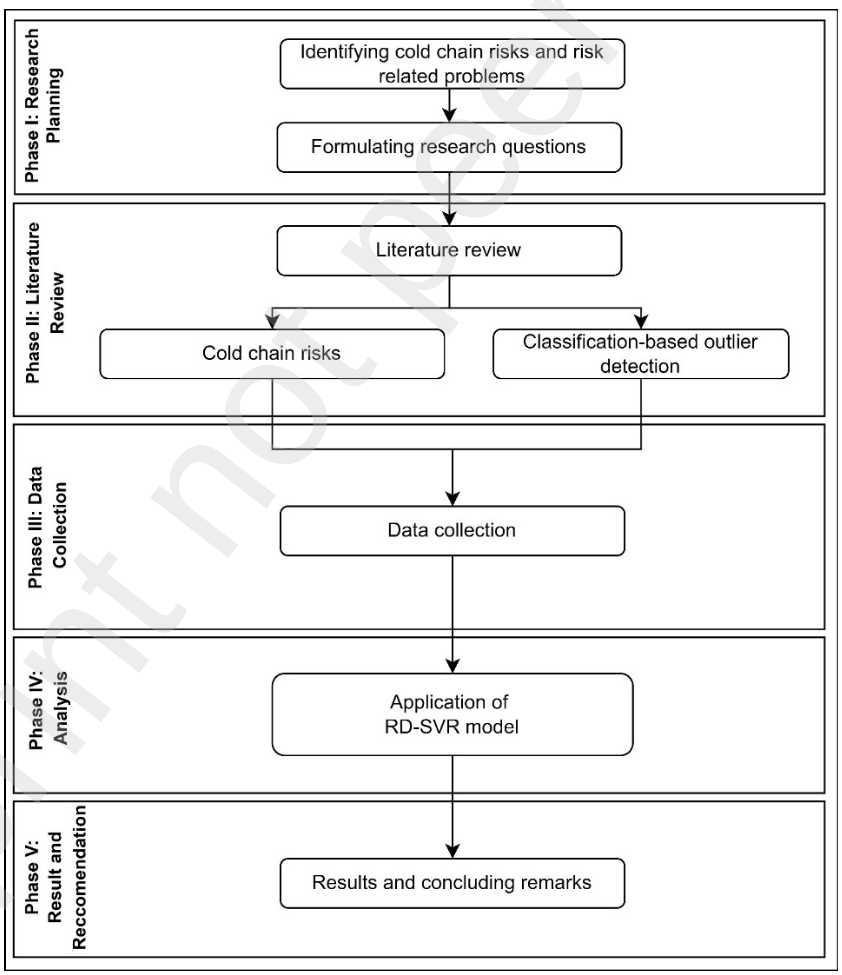

Figure 1. Flowchart of Research Methodology

In Phase I, the research objectives are determined, and the research question is formulated. There are mainly three objectives of the study. The first objective is to identify the anomalous temperature in pharmaceutical cold chain logistics which may cause the risks and damage the product safety. Second objective is to develop an algorithm for analyzing, interpreting and managing the risks in pharmaceutical cold chain. Finally, the third objective is to investigate and analyze the performance of developed techniques in this field. While reaching these objectives, route detection-based support vector regression algorithm is proposed. In frame of these purposes, a research question is conducted. The research question of the study is "Can RD-SVR algorithm be used to detect out-of-reference temperature measurements on location-based temperature data in pharmaceutical cold chain logistics?"

In Phase II, a comprehensive literature review was conducted. The literature review was divided into two topics basically: cold chain risks and classification-based outlier detection. The first topic aims to identify cold chain risks and determine the most addressed risk in literature. The second topic is directly related to choosing proper algorithm for analyzing and handling the risks.

Phase III is the data collection phase. In frame of the research question and the research objectives, cooperation was made with one of the most significant logistics companies in the sector. The company provides cold chain transportation data. The company is engaged in the storage and transportation of pharmaceutical products. The cold chain structure is illustrated in Figure 2.

Figure 2. Cold Chain structure

The company's operation center is located in Kocaeli, Türkiye. Cold products are stored in cold rooms of the warehouse. They are then transported to pharmaceutical warehouses by refrigerated vehicles according to the demand. From pharmaceutical warehouses, they are delivered to customers. The temperature is constantly monitored both in the storage area and in transportation. The company provided us with the temperature data during transportation over a one-year period. Further information about the data is given in Sub-section 3.1.

Phase IV is the analysis phase. Detecting anomalous temperature in cold chain logistics can be considered as a binary classification problem. For this purpose, a route detection-based support vector regression algorithm is developed to detect anomalous temperature along the route as a classification-based outlier detection.

Phase V is the conclusion phase. In this phase all the results of the analyses are presented and discussed in the context of the research question. Also, the performance of the developed model is evaluated. Detailed results and discussions will be given in the relevant sections later in the paper.

3.1. Cold Chain Data

The quality and safety of cold chain products must be ensured from the producer to the end consumer during storage and transportation. Pharmaceutical products are transported in the cold chain with refrigerated vehicles in the cooperated company. When transporting products, the temperature value in the vehicle should be between 2°C and 8°C. The company's refrigerated vehicles are cooled with air conditioners. There are two air conditioners in each vehicle in case of any breakdowns. Monitoring the temperature and humidity inside vehicles is carried out using data loggers. These devices are used to monitor the internal environmental conditions of vehicles and collect data.

The case study company provides storage and transportation services for pharmaceutical cold chains. They use top-notch technology in both storage and transportation to keep the products in the desired temperature range. The desired temperature range is between 2°C and 8°C. The air conditioners keep the temperature in the desired range. The situation in the storage area is more stable and under control. In case of any malfunction, backup air conditioners are activated, and the technical team intervenes immediately. For this reason, out-of-limit situations are rarely experienced in the storage area. However, situation in transportation is a bit challenging since the Türkiye's road characteristics has some difficulties. Air conditioners fail more often. The temperature and the humidity inside the vehicles are monitored via an online tracking system. The system warns in out-of-limit situations. Consequently, we chose to investigate the transportation in cold chain rather than storage as a research area.

The company provided us with one year of data from the data loggers in the two vehicles on route base. Data loggers are measuring devices used in temperature and humidity measurement, which record the temperature or temperature-humidity values of the environment in which it is located until its capacity is full, can store the recorded information in its memory and allow the information in its memory to be transferred to the computer when necessary. There are three data loggers in each refrigerated vehicle of the company. Two of them measure the temperature inside the vehicle and one of them measure the humidity inside the vehicle. Each data logger measures approximately every 2 minutes. Thus, in the transportation of the data, we have two temperature values, one humidity value, vehicle identity, date, time, latitude and longitude information between April 2023 and March 2024. An example of the given data includes heads and tails of one file is given in Table 1.

Table 1. An example of the data given

TYPE	VALUE	unitid	dtupdategmt	lat	lon
2	27	54808	2023-08-01 00:01:10	40,984723	29,251308
5	59	54808	2023-08-01 00:01:10	40,984723	29,251308
3	26,799999	54808	2023-08-01 00:01:10	40,984723	29,251308
2	26,9	54808	2023-08-01 00:03:10	40,984723	29,251308
5	59	54808	2023-08-01 00:03:10	40,984723	29,251308
3	26,799999	54808	2023-08-01 00:03:10	40,984723	29,251308
2	26,9	54808	2023-08-01 00:05:11	40,984723	29,251308
5	59	54808	2023-08-01 00:05:11	40,984723	29,251308
3	26,799999	54808	2023-08-01 00:05:11	40,984723	29,251308
⋮	⋮	⋮	⋮	⋮	⋮
2	29,099999	54808	2023-08-30 09:38:23	40,881196	29,260263
5	70	54808	2023-08-30 09:38:23	40,881196	29,260263
3	29	54808	2023-08-30 09:38:23	40,881196	29,260263
2	29,099999	54808	2023-08-30 10:10:26	40,881196	29,260263
5	70	54808	2023-08-30 10:10:26	40,881196	29,260263
3	29	54808	2023-08-30 10:10:26	40,881196	29,260263
2	29,2	54808	2023-08-30 10:12:27	40,881196	29,260263
5	70	54808	2023-08-30 10:12:27	40,881196	29,260263
3	29,099999	54808	2023-08-30 10:12:27	40,881196	29,260263

In the data, there are 3 “type” values in the first column which are TYPE 2, TYPE 2 and TYPE 5. These values are data logger records. TYPE 2 and TYPE 3 represent the temperature values and TYPE 5 represent the humidity values inside the vehicle. In the second column, “value” represents the temperature and humidity values. In third column, “unitid” represents the vehicle identity. The fourth column shows the date and time of the records. The fifth and sixth columns show the coordinate of the vehicle. In this study, TYPE 2 values are used for analysis as temperature data. In-vehicle temperature data from TYPE 2 for August and December (in order to see the differences between summer and winter months) are given in Figure 3 and Figure 4. The red lines represent the temperature range which is desired to be within all temperature values. The blue dots demonstrate the temperature values.

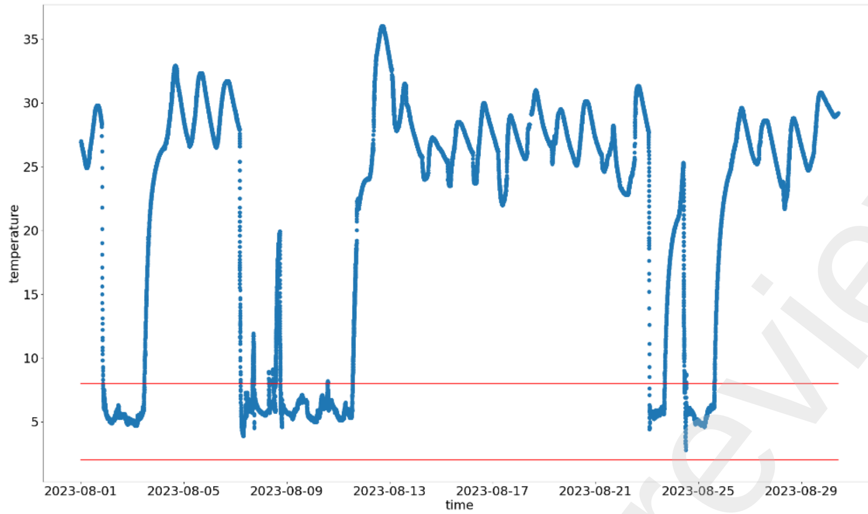

Figure 3. Temperature vs Time Graph for August

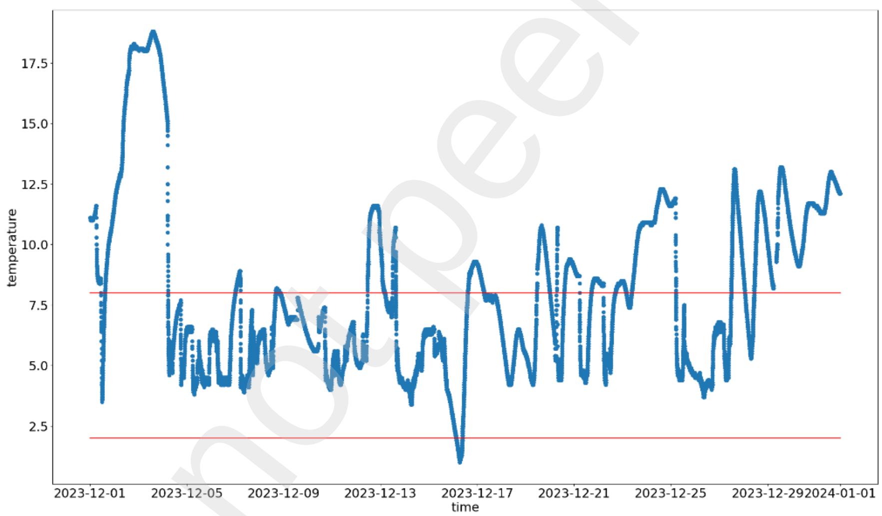

Figure 4. Temperature vs Time Graph for December.

Transportation data is gathered as monthly basis for two refrigerated vehicles over one year period. Each refrigerated vehicle has 12 data files monthly. Each data file has approximately 150000 columns.

3.2. RD-SVR Model

A hybrid approach, Route Detection-based Support Vector Regression (RD-SVR) is developed for classification-based outlier detection for pharmaceutical cold chain to detect anomalous temperature during transportation of pharmaceutical products. The proposed model consists of two consecutive algorithms: Route Detection (RD) algorithm and Support Vector Regression (SVR) algorithm. The route detection algorithm is developed entirely based on the unique cold chain data on hand. In addition to determining the routes, the algorithm also performs big data cleaning, which is one of the most important processes for sensor data. Then, the support vector regression algorithm determines the outliers based on each route during transportation as a classification-based outlier detection method. The flowchart of the proposed RD-SVR model is demonstrated in Figure 5.

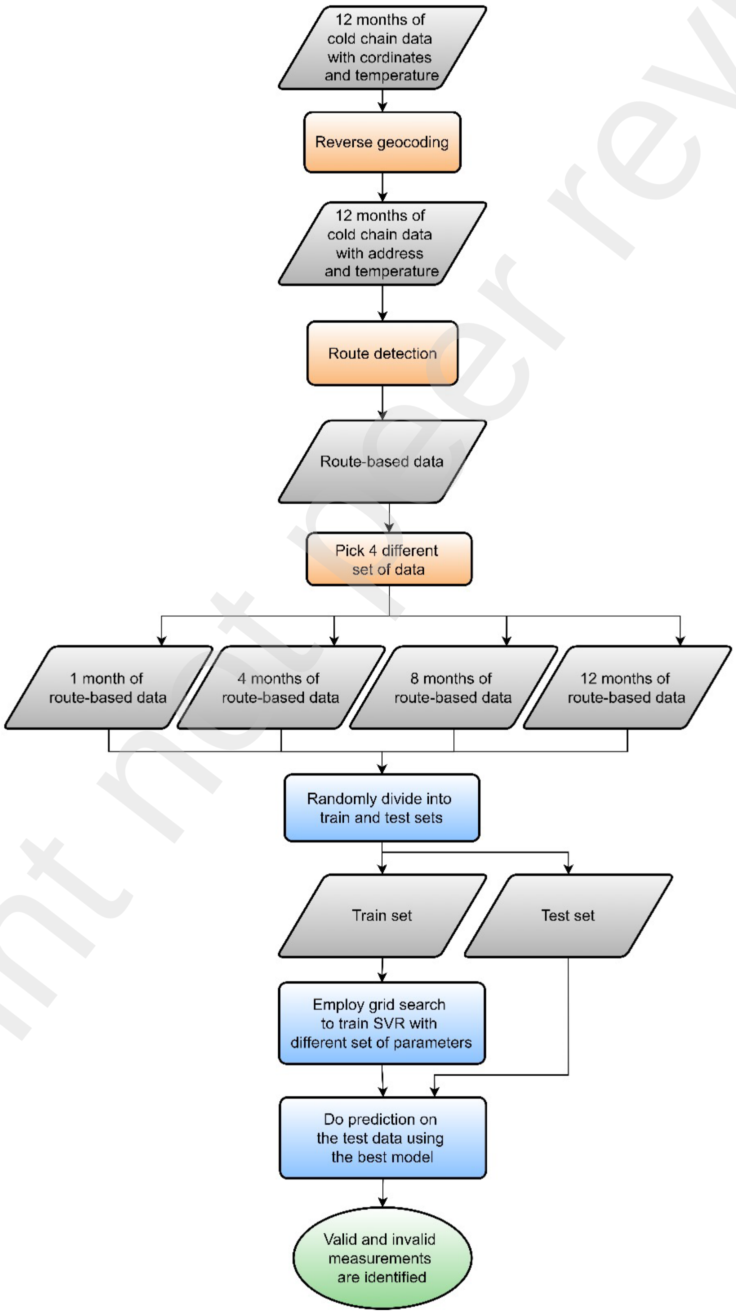

Figure 5. Flowchart of RD-SVR Model

3.2.1. Route Detection Algorithm

Sensor data is considered big data. It is very critical to handle big data cleaning in order to accurately identify anomalies in the field of pharmaceutical cold chain. For this purpose, a rule-based algorithm named route detection is developed to identify the routes while pruning the data. RD algorithm analyses the overall 12 months of data row by row. It aims to identify each distinct route that the refrigerated vehicle takes, by tracking the change in the location and temperature measurements. This process helps to prune the data by identifying the routes and neglecting the measurement taken when vehicle is inactive.

Reverse geocoding is an essential process to determine the routes of refrigerated vehicles. Location information of the refrigerated vehicles is given as latitude and longitude information in the data. However, to determine routes, the address of the vehicle is needed to know instead of the coordinates. Therefore, reverse geocoding is employed to transform latitude and longitude to the address.

In order to apply reverse geocoding, all latitude and longitude pairs of each vehicle from different data files are combined into a set of coordinates. To accomplish this, only a unique set of coordinates can be worked on. Eventually, approximately 300000 unique coordinates data remained. A Python script is prepared, and reverse geocoding service is used to convert the data. The script communicates with the service by sending a coordinate and writing the result back into a file.

As we know, the route of the vehicles always starts with the operation management center of the company, which is located in Kocaeli, Türkiye. The end point depends on customer demand. Since the final destination may vary, the routes of the vehicles should be extracted from the data. To find the routes, we checked the changing addresses starting from Kocaeli in the data on a city-by-city basis by considering temperature inside the vehicle. Each vehicle must maintain an inside temperature between 2°C and 8°C along the route. The cooling inside the vehicles begins before loading. Thus, the interior temperature values during loading meet the required temperature values. We set the threshold as the interior temperature of the vehicle meets the 8°C or under very first time while the vehicle is at Kocaeli location. Then, we eliminated the data before this threshold.

The key aspect of the route detection algorithm is to determine the city where the vehicle finishes unloading. Because the final destination of the routes can be changed according to customer demand. The vehicles should be active on the routes, and they do not have to be active after delivering the pharmaceutical products to the final destination. In other words, the air conditioners are switched off on the way back and the temperature inside the vehicle starts to increase. Even if the vehicles are inactive data loggers inside the vehicles record the temperature and humidity in almost two minutes. Therefore, the data on when the vehicles are inactive should be eliminated. While eliminating the data, we examined the cumulative increase in temperature and set a threshold for the temperature. When the temperature reached 10°C, we assumed that the vehicle was on the way for returning to the operation center. Temperature data between 8°C and 10°C on the return should not be considered as an outlier. Besides, these data constitute a very small part of the whole data. The data higher than 10°C on the way returning is eliminated by the developed RD algorithm. Like this way the RD algorithm prunes the data. As illustrated in Figure 6, most of the unnecessary data was eliminated by using the RD algorithm.

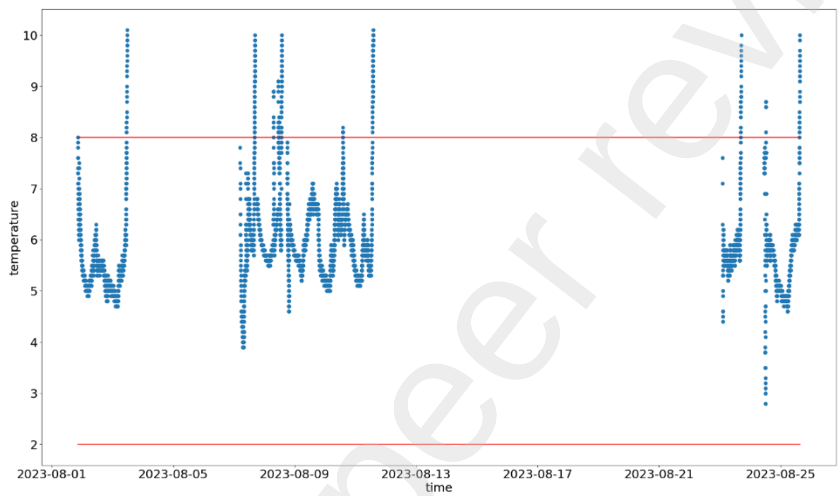

Figure 6. Temperature vs Time for August after data pruning with the RD algorithm.

All routes over one year period are determined monthly for each vehicle by using the developed RD algorithm. For example, one of the vehicles had six routes in August. These routes are given in Figure 7.

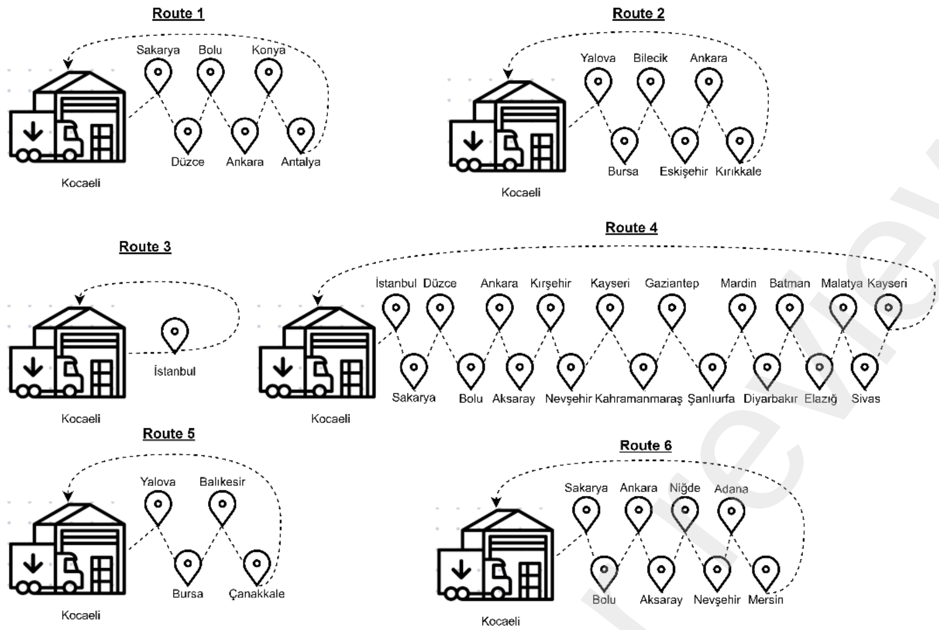

Figure 7. All routes of a refrigerated vehicle in August.

All the routes of each vehicle are demonstrated on the maps of Türkiye. As an example, Route 2 of the vehicle in August is given in Figure 8. The route is stated as a green line on the map. Coordinates of the vehicle with valid temperature are stated as a blue pin and coordinates of the vehicle with invalid temperature stated as a red pin on the map in Figure 9. Figure 10 demonstrates the route after data elimination as a part of the RD algorithm with valid and invalid temperature. These figures were created as HTML files. When the curser hovers on the pins, it shows the temperature, date and time of the measurements. A map-based representation of all routes can be shared upon request.

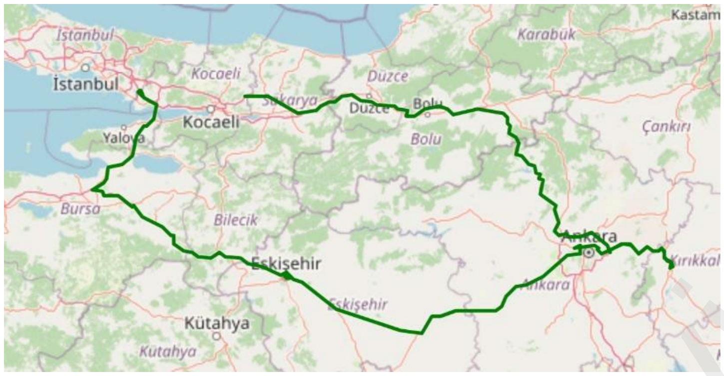

Figure 8. Map-based representation of Route 2.

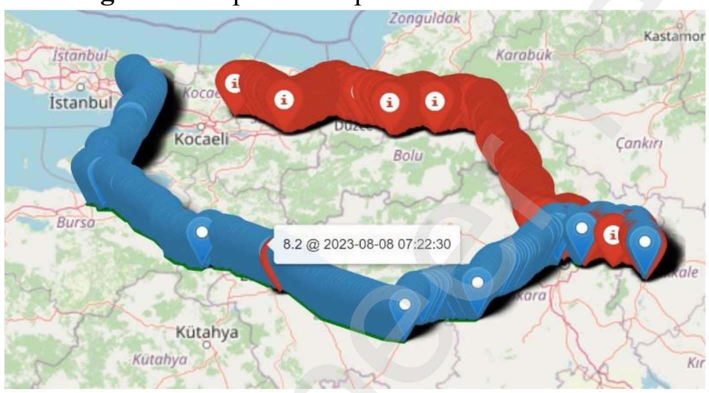

Figure 9. Map-based representation of Route 2 with location pins for each temperature measurement.

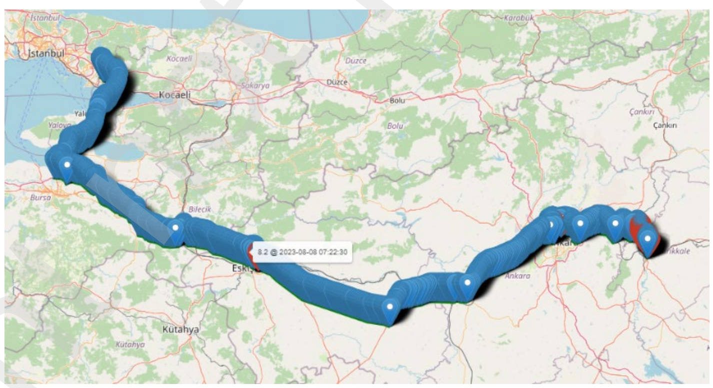

Figure 10. Map-based representation of Route 2 with location pins for each temperature measurement after data elimination with the RD algorithm.

3.2.2. Support Vector Regression

If the training dataset has class labels, outlier detection can be considered as a classification problem and any classification methods can be used (Han et al., 2012). In our problem, the class label is very clear: Valid and invalid temperature. Support Vector Machine (SVM) is selected as a classification method for outlier detection in this study. The SVM can be used in several ways in binary classification. As a variation of SVM, we used Support Vector Regression (SVR) which is proposed by (Drucker et al., 1996) and used only for classification problems, (Fadil et al., 2021). SVR is a data mining algorithm used to predict a real number by mapping an input data to a function, using training data, (H. Yu & Kim, 2012). To match the requirement for SVR, we labeled our data as 1 for valid and 0 for invalid temperature values.

SVR can be used with custom and predefined kernels which have several parameters. In context of this study, linear, polynomial, Radial Basis Function (RBF) and sigmoid kernels and c and gamma parameters are used. The parameter c determines the cost of the mislabeled predictions. Picking a higher c value helps the model to accurately classify all training examples. The gamma parameter determines the significance of a single training example. Picking a higher gamma means that other examples need to be closer to have an impact.

In this study the SVR implementation which is provided by Sckit-learn is used (Buitinck et al., 2013; Pedregosa et al., 2011). The default values of the parameters in this library are rbf for kernel, scale for gamma and 1 for c.

4. RESULTS

Hyper parameter optimization is used to select the optimal parameters (Claesen & De Moor, 2015; T. Yu & Zhu, 2020). Here, the terms parameters include the kernel function, c and gamma values. Possible sets of parameters are processed through grid search. The sets of all parameters are given in Table 2.

Table 2. Parameter values for SVR.

Parameter Name	**Parameter Value**			
Kernel	linear	polynomial	rbf	sigmoid
C	0.1	1	10	100
gamma	auto		scale	

The cross-validation value for the grid search is taken as 3 and the candidate sets of parameters and scored based on Means Squared Error (MSE).

A total of 12 months of data was given to the algorithm as 4 different data sets. The data sets contain 1-month, 4-months, 8-months and 12-months of data. The reason for giving data to the algorithm as four different data sets is to investigate the performance of the developed RD-SVR model on different data sizes. Each data set is randomly split into two parts, 20% test and 80% train. For each train set, grid search returns the parameter set which gives the best result. The best set for each run is given in Table 3.

Table 3. Optimal set of parameters for each data set.

Data sets	C	Kernel	Gamma
1-month	100	rbf	auto
4-months	100	rbf	scale
8-months	100	rbf	scale
12-months	100	rbf	scale

For each set of parameters determined by the grid search, R^2 score and MSE are calculated using the test sets. Results are given in Table 4.

Table 4. Results for each optimal set of parameters.

Data set	Parameters	R^2	MSE
1-month	c:100, kernel:rbf, gamma:auto	0.7958	0.0039
4-months	c:100, kernel:rbf, gamma:scale	0.9183	0.0084
8-months	c:100, kernel:rbf, gamma:scale	0.9136	0.0077
12-months	c:100, kernel:rbf, gamma:scale	0.9174	0.0073

According to the results, SVR provides high accuracy based on a high R^2 value and low error terms, indicating that the SVR algorithm performs well for this study. The fitted graph of the SVR predictions for all data sets is given in Figure 11. According to the graphs, for both valid and invalid samples, the predictions from SVR align with the training data.

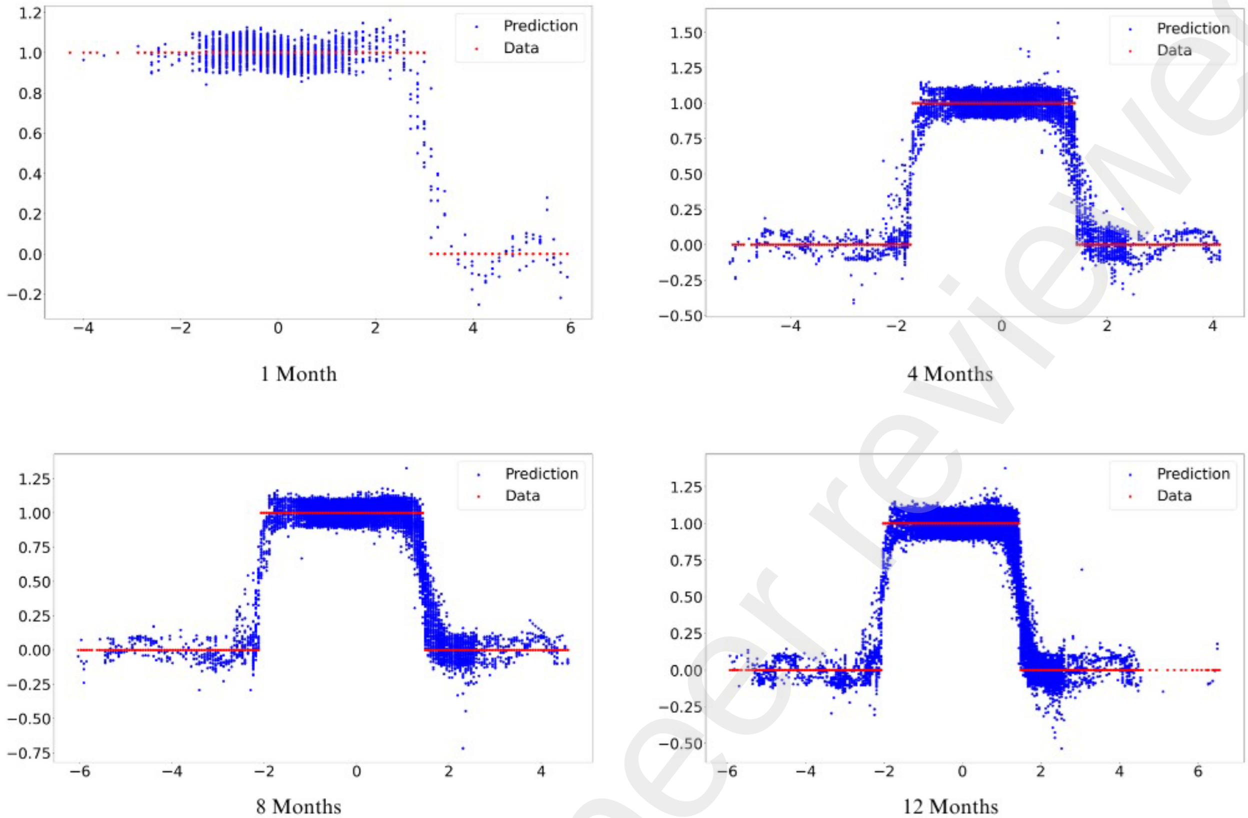

Figure 11. Visual representation of SVR's train set and predictions on for all data sets

Classification results of the RD-SVR model for all data sets are given as confusion matrix in Figure 12.

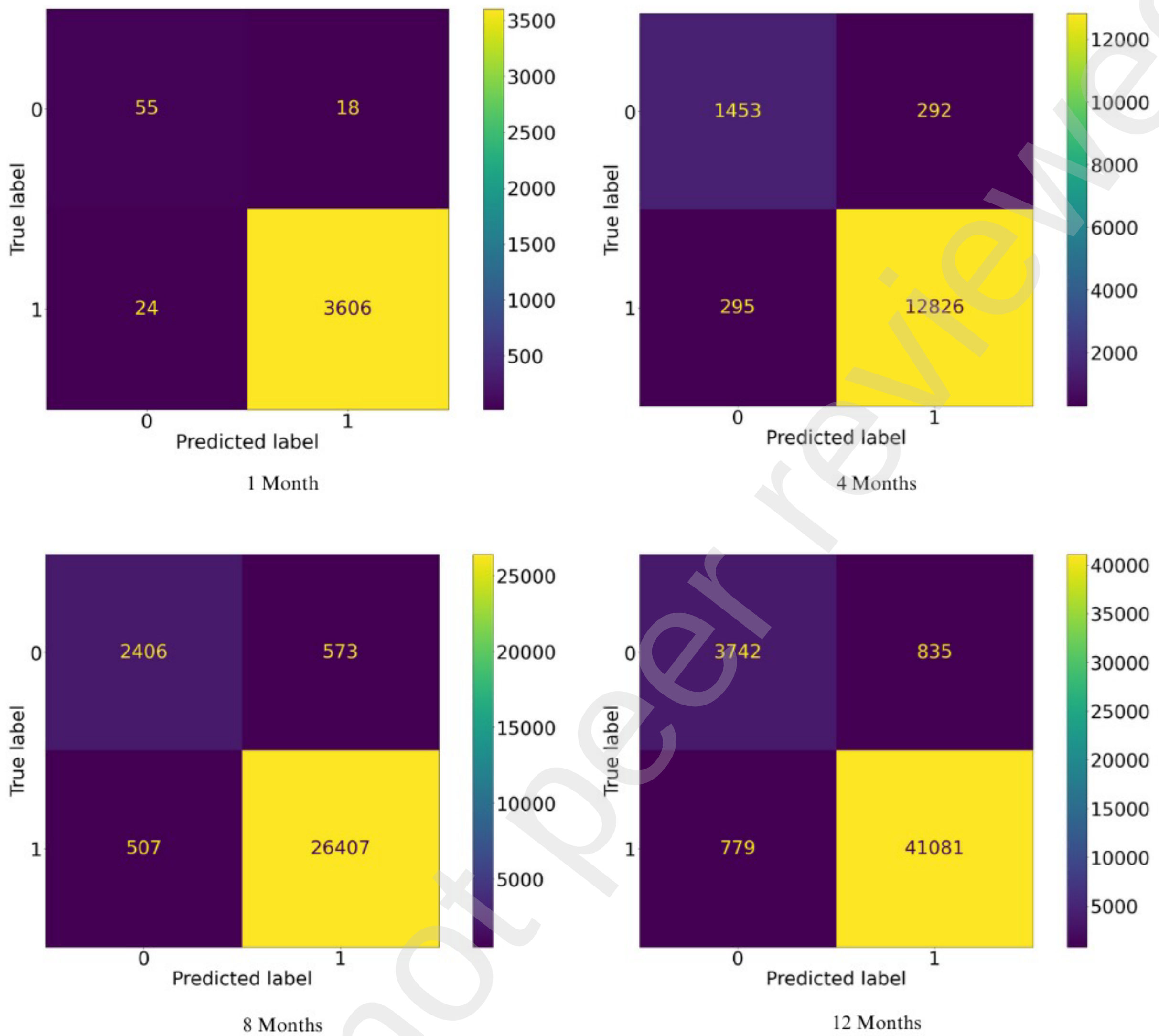

Figure 12. Confusion matrices for all data sets

Anomalous temperature detection using proposed RD-SVR model results for each data set are given in Table 5. The table shows Precision, Recall, F1-Score and Accuracy value for each label in the data sets. In the table, label represented as 0 and 1 which indicates invalid temperature and valid temperature, and support represents the number of data point of each label in each data set. The results of the data sets are compared according to Precision, Recall, F1-Score and Accuracy. As the data size increases, precision, recall and F1-score of anomalous temperature is getting higher which indicates that the RD-SVR model is suitable for big data. Additionally, the accuracy of the model is high for each data set.

Table 5. RD-SVR Model Results

Data Set	Label	Precision	Recall	F1-Score	Support	Accuracy
1 Month	0	0.70	0.75	0.72	73	0.99
	1	1.00	0.99	0.99	3630	
4 Months	0	0.83	0.83	0.83	1745	0.96
	1	0.98	0.98	0.98	13121	
8 Months	0	0.83	0.81	0.82	2979	0.96
	1	0.98	0.98	0.98	26914	
12 Months	0	0.83	0.82	0.82	4577	0.97
	1	0.98	0.98	0.98	41860	

This paper also evaluates the performance of the developed RD-SVR model for different data sizes. As indicated in Figure 12 and Table 5, our proposed hybrid approach performs well for each data set.

5. DISCUSSION AND CONCLUSION

The cold chain is an important part of global trade. Any failure to transport cold products in a healthy and safe way affects the links in the supply chain, whether companies or customers. Therefore, cold chain storage and transportation involve significant risks, and the effects of these risks are substantial as well. Pharmaceutical transportation, one of the most crucial aspects of the cold chain, uniquely involves both regulatory risks and human health risks. Due to these risks, a more careful approach is necessary, along with specific studies in this area. Recently, with advanced technologies, monitoring real-time data has made it easier to enhance the visibility and traceability of the supply chain. Data mining approaches can be utilized in cold chain management, particularly for detecting outliers. Advanced algorithms are required for analyzing large volumes of data collected from sensors and monitoring devices.

Real-time temperature monitoring and the ability to quickly identify points of disruption in the cold chain are essential to prevent product degradation, ensure regulatory compliance, and improve operational efficiency. Previous literature highlighted the usage potential of different data mining techniques in cold chain logistics. However, the increasing volume and complexity of data from sensors and GPS systems in modern cold chains present challenges for traditional outlier detection methods and highlight the need for advanced researches in this area. In order to contribute to this area, a classification-based outlier detection framework was developed to effectively clean big data and detect outliers using suitable data mining algorithms along the route. To achieve this, the RD-SVR model was proposed entirely for the requirements of the problem in this study. The developed model can be considered as one of the most appropriate algorithms for detecting anomalous in big data over a one-year period with high accuracy.

We believe that this study contributes significantly to both theory and practice. Theoretically, no similar studies using classification-based outlier detection have been encountered in the literature on the cold chain; thus, this study contributes to filling that gap by developing highly suitable model for the nature of the problem in this field. From a practical standpoint, this study is a case study, demonstrating a real-world application within the company, which makes its implementation both relevant and feasible. The pharmaceutical products transported in the cold chain are highly expensive and include some for rare diseases. Preventing disruptions in the cold chain is crucial for the company's costs. Therefore, this study helps mitigate potential cold chain disruptions, reducing costs for the company and minimizing health risks for the patients.

6. REFFERENCES

Abusitta, A., de Carvalho, G. H. S., Wahab, O. A., Halabi, T., Fung, B. C. M., & Al Mamoori, S. (2023). Deep learning-enabled anomaly detection for IoT systems. *Internet of Things*, *21*, 100656.

Allied Market Research. (2022). *Cold Chain Logistics Market Size, Share, Competitive Landscape and Trend Analysis Report, by End-use Industry, Business Type: Global Opportunity Analysis and Industry Forecast, 2021-2030*. Available at https://www.alliedmarketresearch.com/cold-chain-logistics-market. Accessed on 26th July 2024.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Arslan, Ö., Karakurt, N., Cem, E., & Cebi, S. (2023). Risk Analysis in the Food Cold Chain Using Decomposed Fuzzy Set-Based FMEA Approach. *Sustainability (Switzerland)*, *15*(17). https://doi.org/10.3390/su151713169

Awad, M., Khanna, R., Awad, M., & Khanna, R. (2015). Support vector regression. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, 67–80.

Baldomero-Naranjo, M., Martinez-Merino, L. I., & Rodriguez-Chia, A. M. (2021). A robust SVM-based approach with feature selection and outliers detection for classification problems. *Expert Systems with Applications*, *178*, 115017.

Barnett, V., Lewis, T., & Abeles, F. (1979). *Outliers in statistical data*. American Institute of Physics.

Bertolini, M., Mezzogori, D., Neroni, M., & Zammori, F. (2021). Machine Learning for industrial applications: A comprehensive literature review. *Expert Systems with Applications*, *175*, 114820.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Grobler, A. G. and J., Layton, R., VanderPlas, J., Joly, A., Holt, B., &

Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 108–122.

Carvalho, T. P., Soares, F. A., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, *137*, 106024.

Centobelli, P., Cerchione, R., & Ertz, M. (2021). Food cold chain management: what we know and what we deserve. *Supply Chain Management: An International Journal*, *26*(1), 102–135.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, *41*(3), 1–58.

Claesen, M., & De Moor, B. (2015). Hyperparameter search in machine learning. *ArXiv Preprint ArXiv:1502.02127*.

Ding, J. F., Weng, J. H., & Chou, C. C. (2023). Assessment of key risk factors in the cold chain logistics operations of container carriers using best worst method. *International Journal of Refrigeration*, *153*, 116–126. https://doi.org/10.1016/j.ijrefrig.2023.06.013

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Advances in Neural Information Processing Systems*, *9*.

Escalante, H. J. (2005). A comparison of outlier detection algorithms for machine learning. *Proceedings of the International Conference on Communications in Computing*, 228–237.

Fadil, I., Helmiawan, M. A., & Sofiyan, Y. (2021). Optimization Parameters Support Vector Regression using Grid Search Method. *2021 9th International Conference on Cyber and IT Service Management (CITSM)*, 1–5. https://doi.org/10.1109/CITSM52892.2021.9589028

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques, 3rd Edition* (3rd Edition). Morgan Kaufmann Pub Inc340 Pine Str, 6th Flr, San Francisco, CA 94104-3205 USA.

Hernandez-Jaimes, M. L., Martinez-Cruz, A., & Ramirez-Gutiérrez, K. A. (2024). A Machine Learning approach for anomaly detection on the Internet of Things based on Locality-Sensitive Hashing. *Integration*, *96*, 102159.

Joshi, R., Banwet, D. K., & Shankar, R. (2009). Indian cold chain: modeling the inhibitors. *British Food Journal*, *111*(11), 1260–1283.

Kang, Z., Catal, C., & Tekinerdogan, B. (2020). Machine learning applications in production lines: A systematic literature review. *Computers & Industrial Engineering*, *149*, 106773.

Konovalenko, I., & Ludwig, A. (2021). Comparison of machine learning classifiers: A case study of temperature alarms in a pharmaceutical supply chain. *Information Systems*, *100*, 101759. https://doi.org/https://doi.org/10.1016/j.is.2021.101759

Liao, H., Hu, Z., Zhang, Z., Tang, M., & Banaitis, A. (2023). Outranking-based failure mode and effects analysis considering interactions between risk factors and its application to food cold chain management. *Engineering Applications of Artificial Intelligence*, *126*. https://doi.org/10.1016/j.engappai.2023.106831

Lin, H. L., Ma, Y. Y., & Lin, C. T. (2023). An Evaluation System for COVID-19 Vaccine Transportation Quality Based on Fuzzy Analytic Hierarchy Process. *Mathematics*, *11*(18). https://doi.org/10.3390/math11183914

Mangini, C. G., Lima, N. D. da S., & Nääs, I. de A. (2024). Risk Prediction Score for Thermal Mapping of Pharmaceutical Transport Routes in Brazil. *Logistics*, *8*(3). https://doi.org/10.3390/logistics8030084

Mascali, L., Schiera, D. S., Eiraudo, S., Barbierato, L., Giannantonio, R., Patti, E., Bottaccioli, L., & Lanzini, A. (2023). A machine learning-based Anomaly Detection Framework for building electricity consumption data. *Sustainable Energy, Grids and Networks*, *36*, 101194.

Oliveira, R. M. A., Sant'Anna, Â. M. O., & da Silva, P. H. F. (2024). Explainable machine learning models for defects detection in industrial processes. *Computers & Industrial Engineering*, *192*, 110214.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel V. and Thirion, B., Grisel, O., Blondel, M., Prettenhofer P. and Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Ravinder, M., & Kulkarni, V. (2023). Performance Analysis of Machine Learning Techniques for Anomaly Detection in Indian Electricity Consumption Data. *Procedia Computer Science*, *230*, 287–296.

Ruiz-Garcia, L., & Lunadei, L. (2010). Monitoring cold chain logistics by means of RFID. *Sustainable Radio Frequency Identification Solutions*, *2*, 37–50.

Salin, V., & Nayga Jr, R. M. (2003). A cold chain network for food exports to developing countries. *International Journal of Physical Distribution & Logistics Management*, *33*(10), 918–933.

Seo, J., Kim, Y., Ha, J., Kwak, D., Ko, M., & Yoo, M. (2024). Unsupervised anomaly detection for earthquake detection on Korea high-speed trains using autoencoder-based deep learning models. *Scientific Reports*, *14*(1), 639.

Shashi, Cerchione, R., Singh, R., Centobelli, P., & Shabani, A. (2018). Food cold chain management: From a structured literature review to a conceptual framework and research agenda. In *International Journal of Logistics Management* (Vol. 29, Issue 3, pp. 792–821). Emerald Group Holdings Ltd. https://doi.org/10.1108/IJLM-01-2017-0007

Shen, Y., & Qian, Y. (2023). Expert system and entropy-ordered weighted average method in cold supply chain risk evaluation. *Expert Systems*, *40*(5). https://doi.org/10.1111/exsy.13163

Thongkam, J., Xu, G., Zhang, Y., & Huang, F. (2008). Support vector machine for outlier detection in breast cancer survivability prediction. *Advanced Web and Network Technologies, and Applications: APWeb 2008 International Workshops: BIDM, IWHDM, and DeWeb Shenyang, China, April 26-28, 2008. Revised Selected Papers 10*, 99–109.

Tian, H., Kong, H., Lu, S., & Li, K. (2025). Unsupervised anomaly detection of multivariate time series based on multi-standard fusion. *Neurocomputing*, *611*, 128634.

Olimov, B. A. U., Veluvolu, K. C., Paul, A., & Kim, J. (2022). UzADL: Anomaly detection and localization using graph Laplacian matrix-based unsupervised learning method. *Computers & Industrial Engineering*, *171*, 108313.

Yildiz Ozenc, S. C., Er, M., & Firat, S. U. (2023). Risks in Supply Chain 4.0: A Literature Review Study. *Global Joint Conference on Industrial Engineering and Its Application Areas*, 163–177.

Yu, H., & Kim, S. (2012). SVM Tutorial-Classification, Regression and Ranking. *Handbook of Natural Computing*, *1*, 479–506.

Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications. *ArXiv Preprint ArXiv:2003.05689*.

Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. In *Machine learning* (pp. 123–140). Elsevier.

Zhang, Y., Meratnia, N., & Havinga, P. J. M. (2013). Distributed online outlier detection in wireless sensor networks using ellipsoidal support vector machine. *Ad Hoc Networks*, *11*(3), 1062–1074.

Zhao, W., Dai, W., & Zhou, S. (2013). Outlier detection in cold-chain logistics temperature monitoring. *Elektronika Ir Elektrotechnika*, *19*(3), 65–68.

Zipfel, J., Verworner, F., Fischer, M., Wieland, U., Kraus, M., & Zschech, P. (2023). Anomaly detection for industrial quality assurance: A comparative evaluation of unsupervised deep learning models. *Computers & Industrial Engineering*, *177*, 109045.

Zsidisin, G. A., & Ritchie, B. (2009). Supply chain risk management–developments, issues and challenges. In *Supply chain risk: A handbook of assessment, management, and performance* (pp. 1–12). Springer.