EMBRY-RIDDLE Aeronautical University
SCHOLARLY COMMONS

International Journal of Aviation, Aeronautics, and Aerospace

Volume 8 | Issue 4 Article 7

2021

Aircraft Exhaust Gas Temperature Value Mining with Rough Set Method

Mustagime Tülin Yıldırım Asst. Prof.
Erciyes Universty, tulin@erciyes.edu.tr
Mehtap Taşcı
Erciyes University, mtasci@erciyes.edu.tr

Cover Page Footnote
This work was supported by Research Fund of the Erciyes University, Project no. FDK-2020-9981.

The aviation industry has great importance for transportation and commercial facilities of countries. In addition to being used only for transportation purposes, it is used in many different areas such as military forces, health, transportation, forest fires, agricultural spraying, and traffic control. Aircraft have the largest share among vehicles used in the aviation industry. From its first invention to the present day, it has constantly developed technologically and continues to develop. In line with the developments, the piston engines used in the first aircraft were replaced by gas turbine engines used today. Piston engines are still used in small aircraft, but gas turbine engines are used in commercial aircraft due to their high performance (Bureau of Transportation Statistics, 2021).

Over time, aircraft maintenance has also improved. Aircraft maintenance was progressing quite simply in the first years. The reason for this was that aircraft structures were also quite simple compared to today. Maintenance usually was accomplished within a short period of time after the flight. Planning of revisions and repairs were being done manually. Over time, aircraft became more complex and maintenance costs increased, making this practice unusable. As a result, the emergence of maintenance systems has become inevitable to reduce costs and increase efficiency (Van den Bergh et al., 2013).

Regular maintenance activities extend the life of the aircraft, provide safe flight, and contribute economically by reducing the duration of the aircraft on the ground. The safe flight is provided, as the flight will be carried out with healthy engines, and environmental protection is provided, as emission gases emitted into the atmosphere will be reduced. Several rules have been developed for maintenance operations by manufacturers, aviation companies, and the country's Civil Aviation Authority.

Maintenance activities have an important place in operating costs. Maintenance activities are divided into sub-branches as planned and unplanned maintenance. Planned maintenance ensures that the planes are constantly convenient to flight. Sudden or unplanned situations such as a hard landing, bird strike, and lightning strikes bring additional maintenance costs. Predictive maintenance methods have been developed to prevent unwanted additional maintenance. Predictive maintenance methods that emerged in the 1990s are based on the interpretation of data recorded while the aircraft is operating. With the predictions made before any malfunction occurs, it is aimed to perform maintenance before the problem grows (Demirci, 2009).

There are several studies on fault prediction and engine health monitoring in the literature. Gunetti et al. (2008) developed a simulation structure on gas turbine engine control and engine health monitoring. Mgaya et al. (2009) used fuzzy logic and artificial neural networks in the diagnosis and health management of turbine engines. In addition, Babbar et al. (2009) worked on the development of diagnostic and prognostic plans by using flight conditions data and condition parameters. With this study they determined engine behaviour and predicted engine performance. Babbar et al. stated in their study that decisions can be made about the current and future health of aircraft engines with data such as exhaust gas temperature (EGT), fuel flow (FF), engine fan speed (N1 and N2), and total air temperature (TAT).

Rong et al. (2010) introduced the concept of civil air engine health management according to the nature of civil air engine operation and maintenance and examined some basic methodologies in detail. Cui et al. (2010) conducted a study using acoustic emission technology to monitor the health of aircraft structural components. Yaşar et al. (2010) performed a hardware study for real-time health control of aircraft engine components using piezoelectric sensors. Similar to Yaşar et al., Sadough Vanini et al. (2014) developed a method for engine health monitoring. The method consists of dynamic neural networks (DNNs) and named fault detection and isolation (FDI).

Yang et al (2014) studied a similar method to estimate degraded engine component parameters using quantum-behaved particle swarm optimization (QPSO) algorithm. The new method was applied to turbine fan engine health status estimation and was compared with the other three representative methods. Tavakolpour-Saleh et al. (2015) focused on parametric and nonparametric system identification of an experimental turbojet engine and developed two parametric and nonparametric methods.

Li et al. (2015) applied clustering techniques to detect abnormal flights of unique data patterns. Daroogheh et al. (2015) proposed a novel hybrid structure for the development of health monitoring techniques of nonlinear systems by integration of model-based and computationally intelligent and data-driven techniques. In 2015, Lv et al. first analysed the types of motor failures, then studied on how to apply the motor failure prediction and the remaining useful life (RUL) estimation. Megatroika et al. (2016) developed two anomaly detection model: Self-Organizing Map Neural Network (SOM NN) and One-Class Support Vector Machine (SVM).

Kiakojoori and Khorasani (2016) studied the problem of health monitoring and prognosis of aircraft gas turbine engines by using computationally intelligent methods. In a similar manner, Amozegar and Khorasani (2016) proposed a new approach for Fault Detection and Isolation (FDI) of gas turbine engines by developing an ensemble of dynamic neural network identifiers. Yang et al. (2016) developed a new extreme learning machine optimized by quantum-behaved particle swarm optimization (QPSO) for a gas turbine fan engine diagnostic problem. Djaidir et al. (2017) proposed a fault monitoring system of a gas turbine engine based on vibration analysis technique using spectral analysis tools. In addition, Chen et al. (2018) studied Remaining Useful Life (RUL) estimation of an aircraft engine by using the whole lifecycle data and performance-deteriorated parameter data.

Benrahmoune et al. (2018) studied engine vibration with the aim of developing a failure detection system by using dynamic neural networks approach, whereas Giorgi et al. (2018) studied engine health condition estimation while Yıldırım and Kurt (2018) estimated the EGT parameter and the Low-Pressure Turbine Vibration (LPT) parameter (2019) using aircraft flight data to monitor the health of the aircraft. Balakrishnan et al. (2021) proposed a novel method for monitoring aircraft engine health using Whale Optimization Algorithm based Artificial Neural Network (WOANN).

There are some studies on aviation using data mining in the literature as follows. Letourneau et al. (1999) conducted studies to predict various errors in aircraft components using numerical and symbolic data obtained from aircraft. Skormin et al. (2002) studied the environmental impact of avionic systems using data mining methods while Lian and Liou (2004) used data mining on the problems of optimizing air foil and designing the turbopump of a cryogenic rocket engine. Foslien et al. (2004) used data mining to detect anomaly conditions that will occur in the space shuttle. Iverson (2008) handled the health status of mission operations of both the space shuttle and international space stations using the nearest neighbour algorithm in data mining. Using similar methods, Seichepine et al. (2011) made anomaly detection by data mining using flight data while Bastos et al. (2013) presented a method to forecast future failures based on data analysis by using Rapid Miner.

Pagels (2015) discussed data mining studies on aviation and investigated which algorithm gave better results on data sets by studying various data mining algorithms. Gharoun et al. (2019) used data mining techniques for fault detection. The proposed algorithm was applied on an aircraft turbofan engine using flight data. Musa et al. (2021) carried out a study on gas turbine engine health monitoring with big data method. However, these studies include data mining techniques, there are few studies conducted with the data mining method on the aircraft engine failure prediction. In order to fill the gap in this area, a study was carried out to predict aircraft engine failure using the data mining method.

In the literature there are some studies on exhaust gas temperature (EGT) for monitoring engine health condition. EGT temperature is one of the important operational parameters in an engine, it indicates the temperature of the gas leaving the turbine unit of the engine. European Aviation Safety Agency (EASA), approves the aircraft engine by placing a limit (red line) for EGT temperature. The limit values determined by the EASA for CFM 56 engine types are given in Table 1. In this paper, the EGT limit value of the gas turbine engine is 930° C. The difference between the EGT Redline value and the EGT Take - off value constitutes the EGT

Margin (EGTM) value. EGTM is an important for monitoring engine health and is calculated by Equation 1.

Table 1

CFM 56 Gas Turbine Motor EGT Limitation

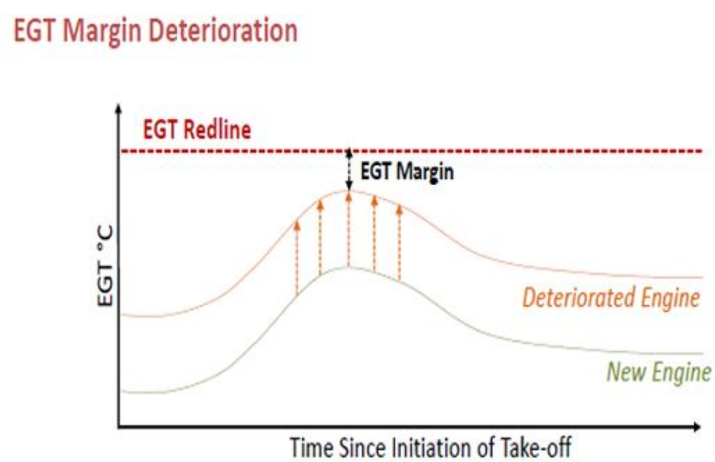
	Take Off	Take Off Transitory (20 seconds)	Maximum Continuous
CFM56-2, CFM56-2B	905	N/A	870
CFM56-2A	930	N/A	895
CFM56-3, CFM56-3B,	930	940	895

$$\text{EGT Margin } ^\circ\text{C} = \text{EGT Redline} - \text{EGT Take-off} \quad (1)$$

When the engine is new or renewed, EGTM is at the maximum level. As can be seen from the Figure 1, the EGTM decreases when the engine breaks down.

Figure 1

EGT Margin Value (Yildirim & Kurt, 2018).



The EGTM is an indicator of how much deterioration in an engine. Over time, as a result of the fouling of the compressor, engine deterioration occurs and the efficiency of the engine decreases.

This situation causes the turbine that runs the compressor to run more. Consequently, the engine's EGT temperature becomes higher and EGTM decreases. Degradation in EGTM is the primary cause of engine removals, especially for engines operating in short-range missions (Technical Drivers of

Aircraft Engine Off-Wing MaintenanceSofema Aviation Services). High EGT value in flight may indicate engine stall, tail pipe fire, engine fire, and aging engine situations. That's why it's important. If the EGT of the gas turbine engine is higher than it should be, this will lead to further engine wear. Therefore, the performance of the engine is negatively affected. If the EGT value is close to the EGT redline value, the turbine blades will be damaged and the engine must be disassembled (Demirci, 2009; Yilmaz, 2009).

Although there are many different studies on EGT in the literature, as far as we know, there is no study on estimating the EGT value using the data mining method among these studies. Our study is aimed to fill this gap in the literature. In our paper, EGT parameter estimation was made using Knowledge Discovery in Database (KDD) data mining method, which has become very popular in the data processing. The engine health condition can be monitored with the estimation made, the malfunctions that may occur with the estimation are determined and preventive maintenance methods are applied before the malfunctions grow. Highly costly engine maintenance will be replaced by less costly protection maintenance, which will cause a decrease in the maintenance item of the operator. As the flight will be made with healthy engines, the safe flight will be ensured, emission gases emitted into the atmosphere will be reduced and environmental protection will be provided.

Method

Data Mining

Today, collecting data has gained great importance with the advancing technology. Only the collection and accumulation of data does not make sense in itself, as long as no action is taken on the collected data, it remains a meaningless data collection. It becomes meaningful if previously unknown and unpredictable information is obtained from the data. The process of retrieving information by processing this collected data with various software is called data mining. Data mining is a study between various disciplines such as statistics, machine learning, artificial intelligence, database management (Jiawei & Kamber, 2006; Prabhu & Venkatesan, 2006).

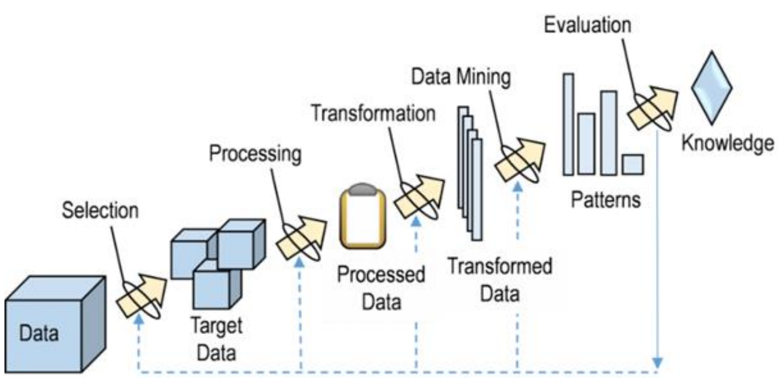
Data mining methods are used in customer relationship management, banking or finance sector, logistics and transportation, production and maintenance, insurance, computer software and hardware, engineering and science, defence industry, transportation sector, and health field (Olson & Delen, 2008). There are the most preferred and standardized data mining methods are SEMMA (Sample, Explore, Modify, Model and Assess) method, CRISP-DM (Cross-Industry Standard Process for Data Mining) method and KDD (Knowledge Discovery in Database) method. In this paper EGT estimation was performed using the KDD method.

Knowledge Discovery in Database (KDD) Method

KDD method consists of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and information presentation stages. In the KDD method, noisy and inconsistent data are extracted from the data set during data cleaning. If there is more than one data source, these data are combined during the data integration phase. During the data selection stage, the data related to the analysis to be made are determined. Then, the transformation of the data and the determination of the patterns in the data are performed sequentially. In the pattern evaluation phase, interesting patterns obtained according to some criteria are evaluated. Finally, the information obtained during the information presentation stage is presented to the users and the KDD process is completed (Azevedo & Santos, 2008). A schematic representation of the KDD method is given in Figure 2.

Data Pre-processing. As can be seen in the method mentioned above, data mining includes starting with the determination of the problem, performing various operations on the obtained data set, and reaching a result. For the data to be suitable for data mining, it must first go through a pre-processing stage. At this stage, data is cleaned from noisy data, missing data are completed. If the size of the data is large, size reduction is performed to be able to process.

Figure 2
KDD Method



Then, if there is any contrary data, it is detected and cleaned. In the data pre-processing stage of this study, the Rough Set Approach Method was used to reduce the attribute dimension of the data.

Rough Set Approach Method

The Rough Set Approach Method was developed by Pawlak in the 1980s. In this approach, sets do not have clear boundaries, which are similar to fuzzy sets. These structures, which do not have strict boundaries, allow us to act flexibly on

DOI: https://doi.org/10.15394/ijaaa.2021.1660

the data. If our data is incomplete, insufficient, or uncertain, it allows us to organize it and make it suitable for analysis (Pawlak, 1982).

Aircraft Exhaust Gas Temperature Value Mining with Rough Set Method

N1, N2, altitude, airspeed, ground speed, pitch, angle of attack, roll, heading, vertical acceleration, latitude acceleration, longitude acceleration, speed break, total air temperature, EGT data were used in our study from the data obtained by flight data recorder during the flight of a commercial aircraft. In this study, EGT value is used for failure prediction with the KDD data mining method.

Data Pre-Processing with Rough Set Approach

The data, N1 (Fan Speed), N2 (Core Speed), altitude, airspeed, ground speed, pitch, angle of attack (AOA), roll, heading, vertical acceleration, latitude acceleration, longitude acceleration, speed break, total air temperature (TAT), and EGT parameters (attributes) are used in this pre-processing section. Our data set consists of 1248 lines. The rough set approach was used to reduce the size of the data. And it makes higher accuracy estimation possibility. The size reduction process was carried out with the ROSETTA program and a genetic algorithm was used during reduction. In this section, 20 subsets obtained with the size reduction process, but we took only the 7 subsets which are including EGT value and these subsets are given below:

Set 1:{altitude, ground speed, vertical acceleration, latitude acceleration, longitude acceleration, EGT}

Set 2:{altitude, pitch, roll, vertical acceleration, latitude acceleration, EGT}

Set 3:{altitude, roll, vertical acceleration, latitude acceleration, TAT, EGT}

Set 4:{altitude, vertical acceleration, latitude acceleration, longitude acceleration, TAT, EGT}

Set 5:{altitude, roll, airspeed, vertical acceleration, latitude acceleration, longitude acceleration, EGT}

Set 6:{N1, altitude, roll, vertical acceleration, latitude acceleration, longitude acceleration, EGT}

Set 7:{N1, N2, altitude, airspeed, ground speed, vertical acceleration, latitude acceleration, longitude acceleration, EGT}

Pre-Processing with Outlier Data Detection

In the modelling step, for the algorithms to work better, outlier data should be cleaned from the data sets. Each of the sets obtained as a result of data reduction was cleaned from the outliers using the RAPIDMINER program. While performing the data reduction process, the K-Nearest Neighbour Algorithm (KNN) and in this KNN Algorithm, the Euclidean Distance is used. Euclidean Distance equation is given in Equation 2.

$$d(i,j) = \sqrt{\sum_{k=1}^{p}(X_{ik} - X_{jk})^2} \qquad (2)$$

Outlier data analysis was performed for Set 1, Set 2, Set 3, Set 4, Set 5, Set 6, and Set 7. According to the analysis results, the outlier data graphs of the Set 7 shows the most successful result and the Set 3 shows the most unsuccessful result. The outlier data graph of Set 7 is shown in Figure 3 and the outlier data graph of Set 3 is shown in Figure 4. In these figures the red dots indicate outlier data.

Figure 3
The Set 7 Outlier Data Analysis Chart

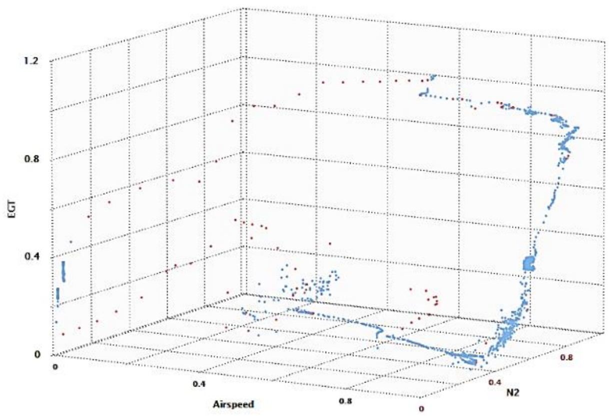
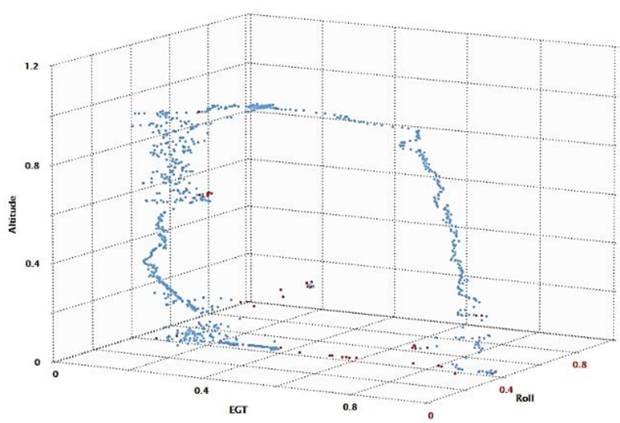


Figure 4
The Set 3 Outlier Data Analysis Graph



Modelling

After the data pre-processing stages, data sets are ready for modelling. Algorithm selection has great importance in the modelling phase. In this study, model is developed on estimation algorithms and according to the data type. In categorical data type, classification process is used and in continuous data type, regression analysis is used. In this study continuous data type was used and the regression process was performed on the data sets.

The stages of the regression process are respectively as follows. In the first stage, the normalization process was performed on the data set, which was made ready for modelling with data pre-processing, and it was ensured that the data took values between 0 and +1. The normalization process has been carried out because the data sizes are different and the differences between them make it difficult to make accurate predictions in our model. For the normalization process, the min-max normalization equation was used and given in Equation 3.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{3}$$

After the normalization process, the EGT data were turned into a Label with the Set Role operator. Making the label means data that is desired to be estimated in supervised learning. After this process, 70% of the data set was separated as training data and 30% as test data with the help of Split Data operator. In this part of the modelling for Linear Regression, Gradient Boosted Trees, Random Forest, Deep Learning Algorithms, the same architecture is applied for all algorithms. After separating our data as training and test data as described above, the Linear Regression operator was used for the implementation of the Linear Regression Algorithm, the Gradient Boosted Trees operator was used for the Gradient Boosted Trees Algorithm, the Random Forest operator was used for the Random Forest Algorithm, and the Deep Learning operator was used in the Deep Learning Algorithm. The regression analysis results, and graphical representation of the Set 7, are given in the Results and Conclusion section.

Each of the steps described above was carried out separately on 7 data sets obtained with the help of a Rough Set. Analysis was made on 7 data sets with Linear Regression, Gradient Boosted Trees, Random Forest, Deep Learning Algorithms. For comparison of these algorithms analysis results, the Root Mean Squared Error (RMSE) criteria was used. The RMSE value is a measure of the performance of the estimation process and it is given in Equation 4.

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n} e_j^2}{n}} \tag{4}$$

Comparison of Root Mean Squared Error (RMSE) values of algorithms are given in the Results and Conclusion section.

Results

In this paper, data mining methods were applied to the data obtained from aircraft flight data recorder (FDR) and the EGT parameter was estimated. In the data mining application, firstly the data reduction process and outlier data cleaning process were performed, and the pre-processing process was completed and the data was made ready for modelling. By using the data reduction rough set method, 20 sub-sets were obtained. Then for estimation the EGT parameter purpose, 7 data sets containing the EGT parameter were separated from these subsets and these 7 data sets were used in the following applications. In order to clean the outlier data in the new sets obtained, the K-Nearest Neighbour Algorithm was applied to the 7 data sets and the outliers were cleaned. Regression analysis was performed using Linear Regression Algorithm, Random Forest Algorithm, Gradient Boosted Trees Algorithm, and Deep Learning Algorithm to estimate the EGT parameter and RMSE values were calculated. These algorithms are used in data mining applications for modelling. The comparisons of RMSE values obtained as a result of the analysis on 7 data sets are presented in Table 2.

Table 2

Comparison of Root Mean Squared Error (RMSE) Values of Algorithms

Algorithms/Sets	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7
Linear Regression Algorithm	0.191	0.323	0.503	0.215	0.213	0.209	0.136
Random Forest Algorithm	0.043	0.091	0.401	0.048	0.047	0.040	0.015
Gradient Boosted Trees Algorithm	0.201	0.213	0.446	0.199	0.197	0.194	0.192
Deep Learning Algorithm	0.098	0.150	0.500	0.081	0.075	0.065	0.101

According to the results in Table 2, the Random Forest Algorithm took **0.015** RMSE value in the Set 7 and estimated the EGT value with the lowest error. In Table 3, the lowest and the highest RMSE values of the sets with these algorithms are given.

DOI: https://doi.org/10.15394/ijaaa.2021.1660

Table 3
The Lowest and the Highest RMSE Values of Sets

Sets/Values	The lowest value	The highest value
Set 1	0.043 (Random Forest Algorithm)	0.201 (Gradient Boosted Trees Algorithm)
Set 2	0.091 (Random Forest Algorithm)	0.213 (Gradient Boosted Trees Algorithm)
Set 3	0.401 (Random Forest Algorithm)	0.503 (Linear Regression Algorithm)
Set 4	0.048 (Random Forest Algorithm)	0.215 (Linear Regression Algorithm)
Set 5	0.047 (Random Forest Algorithm)	0.213 (Linear Regression Algorithm)
Set 6	0.040 (Random Forest Algorithm)	0.209 (Linear Regression Algorithm)
Set 7	0.015 (Random Forest Algorithm)	0.192 (Gradient Boosted Trees Algorithm)

As can be seen from Table 3, Linear Regression Algorithm estimated the RMSE value as 0.503 in the Set 3. This value is the farthest predicted value obtained. The Set 3, which consists of altitude, roll, vertical acceleration, latitude acceleration, total air temperature, and the EGT, took an RMSE value between 0.401-0.503 and formed the set with the weakest success in EGT estimation. Random Forest Algorithm, which consists of N1, N2, altitude, airspeed, ground speed, vertical acceleration, latitude acceleration, longitude acceleration, and EGT, has achieved the most successful result with a value of 0.015. In the Set 7, unlike the Set 3, there are N1, N2, airspeed, ground speed, longitude acceleration attributes. Depending on the results of the analysis, we can say that these features are more effective in estimating the EGT parameter. In Table 4, the lowest and the highest RMSE value of the Linear Regression Algorithm, the Random Forest

Algorithm, Gradient Boosted Trees Algorithm, and Deep Learning Algorithm are given.

Table 4

The Lowest and the Highest RMSE Value of the Algorithms

Used Algorithm/Value	The lowest value	The highest value
Linear Regression	0.136 (Set 7)	0.503 (Set 3)
Random Forest	0.015 (Set 7)	0.401 (Set 3)
Gradient Boosted Trees	0.192 (Set 7)	0.446 (Set 3)
Deep Learning	0.065 (Set 6)	0.500 (Set 3)

When all algorithms are evaluated, the Random Forest Algorithm seems to be the most successful algorithm. Algorithms other than Deep Learning Algorithm yielded more successful prediction results in the Set 7. Deep Learning Algorithm gave the most successful result in the Set 6. Algorithms all made unsuccessful predictions in Set 3. When all these evaluations are made together, it is seen that the best estimation result is realized in the Set 7.

The regression analysis results are given in Table 5 and Table 6. In Table 5, Model Summary is shown and in Table 6, Set 7 correlation values are shown. In Figure 5 graphical representation of the predicted EGT and the real EGT values.

Table 5

The 7 Set Model Summary Table

Change Statistics				
R Square Change	F Change	df1	df2	Sig. F Change
,902a	1521,993	8	1326	,000

DOI: https://doi.org/10.15394/ijaaa.2021.1660

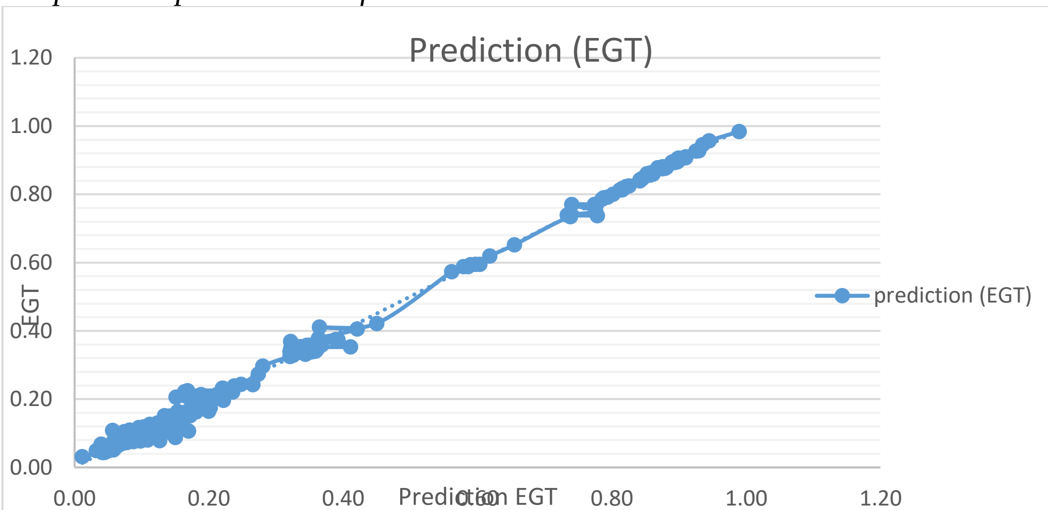
Table 6

The Set 7 Correlation Tables.

Correlations										
		EGT	N1	N2	Altitude	Air speed	Ground speed	Vertical acceleration	Latitude acceleration	Longitude acceleration
	EGT	1,00	,684	,600	,087	-,001	,052	-,121	,131	,859
	N1	,684	1,00	,957	,703	,603	,695	-,127	,289	,645
	N2	,600	,957	1,00	,622	,660	,691	-,107	,306	,626
	Altitude	,087	,703	,622	1,00	,810	,941	-,101	,266	,037
	Air speed	-,001	,603	,660	,810	1,00	,958	-,060	,341	,082
	Ground speed	,052	,695	,691	,941	,958	1,00	-,082	,323	,075
	Vertical acceleration	-,121	-,127	-,107	-,101	-,060	-,082	1,00	-,083	-,091
	Latitude acceleration	,131	,289	,306	,266	,341	,323	-,083	1,00	,165
	Longitude acceleration	,859	,645	,626	,037	,082	,075	-,091	,165	1,000
Sig	EGT	.	,000	,000	,001	,483	,029	,000	,000	,000
	N1	,000	.	,000	,000	,000	,000	,000	,000	,000
	N2	,000	,000	.	,000	,000	,000	,000	,000	,000
	Altitude	,001	,000	,000	.	,000	,000	,000	,000	,087
	Air speed	,483	,000	,000	,000	.	,000	,014	,000	,001
	Ground speed	,029	,000	,000	,000	,000	.	,001	,000	,003
	Vertical acceleration	,000	,000	,000	,000	,014	,001	.	,001	,000
	Latitude acceleration	,000	,000	,000	,000	,000	,000	,001	.	,000
	Longitude acceleration	,000	,000	,000	,087	,001	,003	,000	,000	.

Figure 5

Graphical Representation of the Predicted EGT and the Real EGT Values



Conclusions

Although there are many studies on engine health monitoring in the literature, there is a few studies on estimating the EGT value using the data mining methods among these studies. EGT parameter is one of the most used parameters to monitor engine health. In this study the EGT parameter was estimated with high accuracy. For this purpose, Linear Regression Algorithm, Random Forest Algorithm, Gradient Boosted Trees Algorithm, and Deep Learning Algorithm were applied on 7 different sets in the estimation process. This study has great importance for maintenance costs, safe flight, and environmental protection.

Funding

This work was supported by Research Fund of the Erciyes University, Project No. FDK-2020-9981.

Conflict Of Interest

There is no conflict of interest in this study.

References

Bureau of Transportation Statistics. (2021). *Active U.S. air carrier and general aviation fleet by type of aircraft*. (2021). https://www.bts.gov/content/active-us-air-carrier-and-general-aviation-fleet-type-aircraft

Amozegar, M., & Khorasani, K. (2016). An ensemble of dynamic neural network identifiers for fault detection and isolation of gas turbine engines. *Neural Networks*, *76*, 106–121. https://doi.org/10.1016/j.neunet.2016.01.003

Babbar, A., Ortiz, E. M., & Syrmos, V. L. (2009). *Fuzzy clustering based fault diagnosis for aircraft engine health management*. 17th Mediterranean Conference on Control & Automation, 199–204. https://doi.org/10.1109/med.2009.5164539

Babbar, A., Ortiz, E. M., Syrmos, V. L., & Arita, M. M. (2009). *Advanced diagnostics and prognostics for engine health monitoring*. 2009 IEEE Aerospace Conference, 1–10. https://doi.org/10.1109/AERO.2009.4839657

Balakrishnan, N., Devasigamani, A. I., Anupama, K. R., & Sharma, N. (2021). Aero-engine health monitoring with real flight data using whale optimization algorithm based artificial neural network technique. *Optical Memory and Neural Networks (Information Optics), 30*(1), 80–96. https://doi.org/10.3103/S1060992X21010094

Bastos, P., Lopes, I., & Pires, L. (2013). Application of data mining in a maintenance system for failure prediction. *Safety, Reliability and Risk Analysis*, 933–940. https://doi.org/10.1201/b15938-138

Benrahmoune, M., Hafaifa, A., Guemana, M., & Chen, X. Q. (2018). Detection and modeling vibrational behavior of a gas turbine based on dynamic neural networks approach. *Journal of Mechanical Engineering*, *68*(3), 143–166. https://doi.org/10.2478/scjme-2018-0032

Chen, Z., Cao, S., & Mao, Z. (2018). Remaining useful life estimation of aircraft engines using a modified similarity and supporting vector machine (SVM) approach. *Energies*, *11*(1). https://doi.org/10.3390/en11010028

Cui, J., Wang, Y., Li, Z., Liu, L., & Xu, G. (2010). *A health diagnosis method for aeroplane structural component based on fault inference engine*. 2010 Chinese Control and Decision Conference, 3928–3931. https://doi.org/10.1109/CCDC.2010.5498448

Daroogheh, N., Baniamerian, A., Meskin, N., & Khorasani, K. (2015). *A hybrid prognosis and health monitoring strategy by integrating particle filters and neural networks for gas turbine engines*. 2015 IEEE Conference on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHAf Technology and Application, PHM 2015. https://doi.org/10.1109/ICPHM.2015.7245020

Djaidir, B., Hafaifa, A., & Kouzou, A. (2017). Faults detection in gas turbine rotor using vibration analysis under varying conditions. *Journal of Theoretical and Applied Mechanics (Poland), 55*(2), 393–406. https://doi.org/10.15632/jtam-pl.55.2.393

Gharoun, H., Keramati, A., Nasiri, M. M., & Azadeh, A. (2019). An integrated approach for aircraft turbofan engine fault detection based on data mining techniques. *Expert Systems, 36*(2). https://doi.org/10.1111/exsy.12370

Giorgi, M. G. De, Campilongo, S., & Ficarella, A. (2018). A diagnostics tool for aero-engines health monitoring using machine learning technique. *Energy Procedia, 148*, 860–867. https://doi.org/10.1016/j.egypro.2018.08.109

Gunetti, P., Mills, A., & Thompson, H. (2008). *A distributed intelligent agent architecture for gas-turbine engine health management.* 46th AIAA Aerospace Sciences Meeting and Exhibit 7 - 10 January 2008, Reno, Nevada, (January), 883. https://doi.org/10.2514/6.2008-883

Kiakojoori, S., & Khorasani, K. (2016). Dynamic neural networks for gas turbine engine degradation prediction, health monitoring and prognosis. *Neural Computing and Applications, 27*(8), 2157–2192. https://doi.org/10.1007/s00521-015-1990-0

Letourneau, S., Famili, F., & Matwin, S. (1999). Data mining for prediction of aircraft component replacement. *IEEE Intelligent Systems and Their Applications*, *14*(6), 59–66.

Li, L., Das, S., Hansman, R. J., Palacios, R., & Srivastava, A. N. (2015). Analysis of flight data using clustering techniques for detecting abnormal operations. *Journal of Aerospace Information Systems, 12*(9), 587–598. https://doi.org/10.2514/1.I010329

Lv, Z., Wang, J., Zhang, G., & Jiayang, H. (2015). *Prognostics health management of condition-based maintenance for aircraft engine systems*. 2015 IEEE Conference on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHAf Technology and Application, PHM 2015, 1–6. https://doi.org/10.1109/ICPHM.2015.7245055

Megatroika, A., Galinium, M., Mahendra, A., & Ruseno, N. (2016). *Aircraft anomaly detection using algorithmic model and data model trained on FOQA data*. Proceedings of 2015 International Conference on Data and Software Engineering, ICODSE 2015, 42–47. https://doi.org/10.1109/ICODSE.2015.7436969

Mgaya, R. H., McCurry, C. D., & Zein-Sabatto, S. (2009). *Intelligent methodology for turbine engine diagnosis and health management*. IEEE Southeastcon 2009, 410–415. https://doi.org/10.1109/SECON.2009.5174115

Musa, G., Alrashed, M., & Muhammad, N. M. (2021). development of big data lean optimisation using different control. *Energy Reports, 7*, 4872–4881. https://doi.org/https://doi.org/10.1016/j.egyr.2021.07.071

Rong, X., Zuo, H., & Chen, Z. (2010). *Civil aero-engine health management integrating with life prediction and maintenance decision-making*. 2010 Prognostics and System Health Management Conference, PHM '10, 803(29). https://doi.org/10.1109/PHM.2010.5414570

Sadough Vanini, Z. N., Khorasani, K., & Meskin, N. (2014). Fault detection and isolation of a dual spool gas turbine engine using dynamic neural networks and multiple model approach. *Information Sciences, 259*, 234–251. https://doi.org/10.1016/j.ins.2013.05.032

Tavakolpour-Saleh, A. R., Nasib, S. A. R., Sepasyan, A., & Hashemi, S. M. (2015). Parametric and nonparametric system identification of an experimental turbojet engine. *Aerospace Science and Technology, 43*, 21–29. https://doi.org/10.1016/j.ast.2015.02.013

Yang, X., Pang, S., Shen, W., Lin, X., Jiang, K., & Wang, Y. (2016). Aero engine fault diagnosis using an optimized extreme learning machine. *International Journal of Aerospace Engineering, 2016*. https://doi.org/10.1155/2016/7892875

Yang, X., Shen, W., Pang, S., Li, B., Jiang, K., & Wang, Y. (2014). A novel gas turbine engine health status estimation method using quantum-behaved particle swarm optimization. *Mathematical Problems in Engineering, 2014*. https://doi.org/10.1155/2014/302514

Yasar, M., Purekar, A., & Sheth, D. (2010). *Real-time hardware implementation of symbolic health monitoring for aircraft engine components*. 2010 American Control Conference, 2224–2229. https://doi.org/10.1109/acc.2010.5531250

Yildirim, M. T., & Kurt, B. (2018). Aircraft gas turbine engine health monitoring system by real flight data. *International Journal of Aerospace Engineering, 2018*(i). https://doi.org/10.1155/2018/9570873

Abbreviations

AOA	Angle of Attack
CRISP-DM	Cross-Industry Standard Process for Data Mining
EGT	Exhaust Gas Temperature
KDD	Knowledge Discovery in Database
KNN	K-Nearest Neighbor
N1	Engine Fan Speed
N2	Engine Core Speed
RMSE	Root Mean Square Error
SEMMA	Sample, Explore, Modify, Model and Assess

TAT	Total Air Temperature

DOI: https://doi.org/10.15394/ijaaa.2021.1660