INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 11, November 2026 | ISSN: 2456-4184 | IJNRD.ORG

Digital Twin For Jet Engine Failure Detection And Predictive Maintenance

[1]Chandana K, [2]Durga M, [3]Shreya R, [4] Sinchana M P, [5]Mrs.Champa C H
[1234]Don Bosco Institute of Technology Visvesvaraya Technological University, Bengaluru, India.
[5]Assistant Professor, Computer Science and Engineering, Don Bosco Institute of Technology, Bengaluru, India.

Abstract : Jet engines involve extreme operating conditions and can exhibit subtle changes in their condition as a result of wear and degradation. Monitoring such changes is essential in determining their remaining useful life and enabling effective maintenance decisions. This paper presents AEROTWIN – a digital twin-based application for jet engine health monitoring and predictive maintenance. The sensor-based prediction workflow uses the NASA C-MAPSS FD001 dataset. Fourteen sensor parameters are selected, and the data is preprocessed using a five-cycle rolling average and an RUL cap of 125 cycles for the corresponding preprocessing experiment. Eight machine learning models are compared using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), coefficient of determination (R^2), training, and prediction time. The LSTM model using a 30-cycle input window achieved the best reported prediction performance with an MAE of 11.0679 cycles, RMSE of 15.9482 cycles, and R^2 of 0.8416 in the supplied model-comparison output. The application also involves acoustic and vibration analysis features, sensor anomaly scoring, and dashboard visualization for engine monitoring, analytics, and digital twin visualization, explainable AI, and maintenance-related information. The current work highlights the implementation of a data-driven engine monitoring application. The reported results are limited to the supplied model-comparison outputs, and no certified aircraft engine fault diagnosis or real-time physical engine deployment can be claimed.

Keywords: Digital Twin, Jet Engine, Predictive Maintenance, Remaining Useful Life, NASA C-MAPSS, LSTM, Random Forest, Acoustic Analysis.

I. INTRODUCTION

Jet engines represent critical components subject to extreme pressures, temperature, and mechanical forces during operation. Over time, changes in the condition of the engine can occur as a consequence of normal wear and degradation. Detecting and interpreting such changes becomes vital in formulating maintenance-related decisions and avoiding unplanned operational disruptions. Conventional practices in maintenance operations involve the use of scheduled or corrective maintenance actions. While scheduled maintenance ensures that all potential failure modes are addressed, it might also lead to unnecessary servicing. On the other hand, corrective maintenance might respond adequately to failures but could leave the asset in a state of reduced performance until such servicing takes place. Predictive maintenance removes such limitations by analyzing the operational data of the equipment to estimate its condition and remaining useful life. Digital Twin technology offers the potential to represent a physical asset using a digital representation. Within the framework of a data-driven application, historical or real-time sensor information can be employed to emulate the operational characteristics of such an engine and enable its condition monitoring. A machine learning model can be leveraged within this digital twin framework to enable monitoring and remaining useful life prediction capabilities based on sensor readings and be visualized using an appropriate dashboard. AEROTWIN represents a digital twin-based application involving jet engine monitoring that is developed using the NASA C-MAPSS FD001 dataset and machine learning. AEROTWIN includes sensor data preprocessing, remaining useful life prediction, model-comparison visualization, and dashboard-based digital twin presentation. The application provides additional modules involving acoustic and vibration analysis for monitoring potential mechanical abnormality.

The contributions of the implementation include the following:

1. Dashboard-based digital twin application for jet engine monitoring;
2. Sensor data preprocessing and remaining useful life prediction workflow using NASA C-MAPSS FD001;
3. Comparative evaluation of eight machine learning models based on regression performance;
4. Acoustic and vibration analysis features for exploring potential mechanical abnormality;
5. Combination of monitoring and analytics, prediction, and maintenance-related information into a dashboard-based interface.

II. LITERATURE SURVEY

Igor Kabashkin proposed a digital twin framework for aircraft lifecycle management using data-driven models. The study focuses on how digital twin technology can support aircraft lifecycle activities by combining data analysis and computational models to improve decision-making and operational efficiency. It discusses the role of digital representations in understanding aircraft behaviour and supporting decisions throughout the lifecycle. The work highlights the importance of using data-driven approaches to make digital twin systems more adaptable to aircraft operational information. This research provides a broad foundation for AEROTWIN, which applies data-driven monitoring and predictive maintenance concepts to jet engine health assessment. [1]

Dong Zhong and colleagues reviewed the use of digital twin technology for predictive maintenance across several industrial domains, including aerospace, manufacturing, power systems, and shipbuilding. The paper explains how digital twins combine information from physical equipment with virtual models to support equipment monitoring, fault diagnosis, and remaining useful life prediction. It presents a predictive maintenance framework involving data acquisition, equipment-state assessment, and maintenance decision-making. The authors also discuss the benefits of integrating different data sources and the challenges associated with model accuracy, real-time data interaction, and practical implementation. This review is relevant to AEROTWIN because the project combines engine sensor information, degradation prediction, anomaly detection, and dashboard-based monitoring into a predictive maintenance workflow. [2]

Luning Li and colleagues presented a detailed review of digital twin technology and its potential applications in the aerospace industry. The paper explains the basic concepts, components, and development of digital twins, while also discussing misconceptions and limitations that can affect their implementation in safety-critical aerospace systems. It identifies important dimensions for future aerospace digital twins, including interaction between physical and virtual systems, standardisation, and cognitive capabilities. The authors emphasise that aerospace digital twins need to be developed with careful consideration of system requirements and operational constraints. This study provides background for AEROTWIN's approach to creating a digital monitoring system that uses engine data and predictive models to support engine health assessment. [3]

Wright and colleagues developed a digital twin of a small jet engine using reservoir computing, a machine-learning approach suited to modelling dynamic systems. The researchers first applied the method to a numerical simulation of a single-scroll jet engine and then developed a model of a JetCat P100-RX engine using experimental measurements. The experimental data included parameters such as thrust, exhaust gas temperature, shaft speed, and weather conditions collected under different engine test conditions. Their trained model predicted engine behaviour under previously unseen test conditions, with reported prediction accuracy better than 1.8% relative to the testing data. This work is closely related to AEROTWIN because it demonstrates the use of experimental engine data and machine learning to build a digital representation of jet engine behaviour. [4]

Yutong Zhang and colleagues proposed a joint-training approach for predicting the remaining useful life of turbofan engines. Their model combines a Deep Belief Network (DBN) for data fusion and degradation-feature learning with a Long Short-Term Memory (LSTM) network for RUL prediction. The authors designed the two components to be trained together, helping connect sensor-data fusion with the final life prediction rather than

treating them as entirely separate stages. The semi-supervised approach was intended to address limitations associated with insufficient labelled engine degradation data while providing information about the progression of engine health. The method was evaluated using the C-MAPSS dataset, making it particularly relevant to AEROTWIN's use of NASA C-MAPSS data and its LSTM-based RUL prediction model. [5]

Sharma and colleagues explored the use of Recurrent Neural Networks (RNNs) for predicting equipment degradation from sequential data. RNNs are designed to process time-dependent information by retaining information from previous time steps, allowing them to learn patterns that develop over time. In degradation prediction, this capability can be used to analyse historical measurements and identify changes in equipment condition. By learning from past operating data, an RNN-based model can estimate future degradation and support condition-based maintenance planning. This approach is relevant to AEROTWIN because the project uses sequential engine sensor data to predict remaining useful life and assess engine health. [6]

Ensarioglu and colleagues proposed a data-driven approach for estimating the remaining useful life of turbofan engines using change-point detection, feature engineering, and a hybrid deep-learning model. The researchers constructed difference-based features from sensor measurements and applied change-point detection to improve the piecewise-linear RUL labelling process. They then developed a hybrid model combining a one-dimensional Convolutional Neural Network (1D-CNN) and a Long Short-Term Memory (LSTM) network to learn degradation patterns from engine data. The proposed method was evaluated using the FD001 subset of the NASA C-MAPSS dataset, and the model using difference-based features achieved an RMSE of 16.1 and a score of 437.2. This study is highly relevant to AEROTWIN because it also uses the C-MAPSS FD001 dataset for turbofan RUL prediction and highlights the importance of feature engineering and temporal learning in improving prediction performance. [7]

The researchers investigated the application of Transformer-based models for remaining useful life (RUL) prediction. Transformer architectures use attention mechanisms to identify relationships between different time steps in sequential data, enabling them to capture long-term dependencies in equipment degradation patterns. In RUL prediction, this capability can help analyse multivariate sensor measurements and identify temporal patterns associated with changes in equipment health. Transformer-based approaches provide an alternative to recurrent neural networks such as LSTM, particularly when modelling long sequences of sensor data. This research is relevant to AEROTWIN because it explores another deep-learning approach that can be applied to turbofan engine degradation prediction. [8]

Wu and Li proposed a dynamic data-driven digital twin framework for complex engineering products, using aircraft engine health management as an example. The framework incorporates information collected from sensors and Industrial Internet of Things (IIoT) monitoring systems to represent the condition of the physical engine in a virtual environment. The researchers used sensor measurements selected from linear degradation models and developed an Long Short-Term Memory (LSTM) neural network to dynamically update the digital twin and estimate the engine's remaining useful life. The proposed method was compared with other machine-learning approaches, including similarity-based linear regression and feed-forward neural networks. This study is relevant to AEROTWIN because it demonstrates how engine sensor data and an LSTM model can be integrated into a digital twin for engine health monitoring and RUL prediction. [9]

Das and colleagues proposed an anomaly-detection method for aircraft engines based on survival data analysis. The approach estimates the hazard rate under the Cox regression model and uses a decision threshold to identify anomalous engine behaviour. By incorporating information about engine aging, the method accounts for changes in the likelihood of abnormal behaviour over an engine's operational lifetime. The researchers also considered the possibility of modelling differences between engines through frailty modelling. The proposed method was evaluated using data involving hundreds of flights from approximately one thousand aircraft engines and demonstrated better performance than standard classification- and clustering-based anomaly-detection methods. This research is relevant to AEROTWIN because it focuses on identifying abnormal engine behaviour to support predictive maintenance, complementing the project's anomaly-detection component. [10]

Tian and colleagues proposed a Spatial Correlation and Temporal Attention-based LSTM (SCTA-LSTM) framework for predicting the remaining useful life of turbofan engines. The model uses a spatial correlation attention mechanism to learn relationships between different sensor variables and assign importance weights to features at different time steps. An LSTM network is then used to extract temporal dependencies from the processed sensor data, while a temporal attention mechanism identifies important information from the LSTM hidden states. Finally, a fully connected network generates the RUL prediction. The proposed framework was evaluated using two NASA turbofan engine simulation datasets, including C-MAPSS, and the experimental results showed improved performance compared with other existing methods. This study is relevant to AEROTWIN because it uses turbofan engine sensor data and LSTM-based temporal modelling for RUL prediction, while also highlighting the importance of sensor relationships and attention mechanisms. [11]

III.FINDINGS

The performance of AEROTWIN was evaluated using two datasets: NASA C-MAPSS FD001 for Remaining Useful Life (RUL) prediction and DGEN 380 for acoustic condition classification. The evaluation focused on comparing machine learning models, studying prediction errors, identifying unusual sensor patterns, and checking the decision-level fusion mechanism.

A. Performance Comparison of Machine Learning Models

Different machine learning models were tested to identify the most suitable approach for predicting the Remaining Useful Life of turbofan engines. These included Linear Regression, Decision Tree, Random Forest, Extra Trees, Gradient Boosting, HistGradientBoosting, MLP Regressor, and LSTM (30-cycle). The models were compared using three evaluation metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score.

Among the tested models, LSTM achieved the best results. It recorded a validation MAE of 9.6272 cycles, RMSE of 14.3507 cycles, and R^2 score of 0.8870. Unlike traditional models that mainly depend on individual input records, LSTM can learn patterns from sensor readings collected over multiple operating cycles. This ability helped it achieve better RUL predictions.

TABLE I. Validation Performance Comparison of Machine Learning Models

Model	MAE	RMSE	R^2
LSTM (30-cycle)	9.6272	14.3507	0.8870
Random Forest + 5-cycle rolling	12.9940	19.0147	0.8017
Extra Trees	13.8078	19.0026	0.8019
MLP Regressor	13.8266	19.1029	0.7998
HistGradientBoosting	13.9062	19.0219	0.8015
Random Forest	14.0299	19.3919	0.7937
Gradient Boosting	14.6039	19.2574	0.7966
Linear Regression	15.5019	20.4442	0.7707
Decision Tree	18.6750	27.9716	0.5708

As shown in Table I, LSTM produced the lowest MAE and RMSE values and the highest R^2 score. Based on these results, the LSTM (30-cycle) model was selected for the RUL prediction module of AEROTWIN.

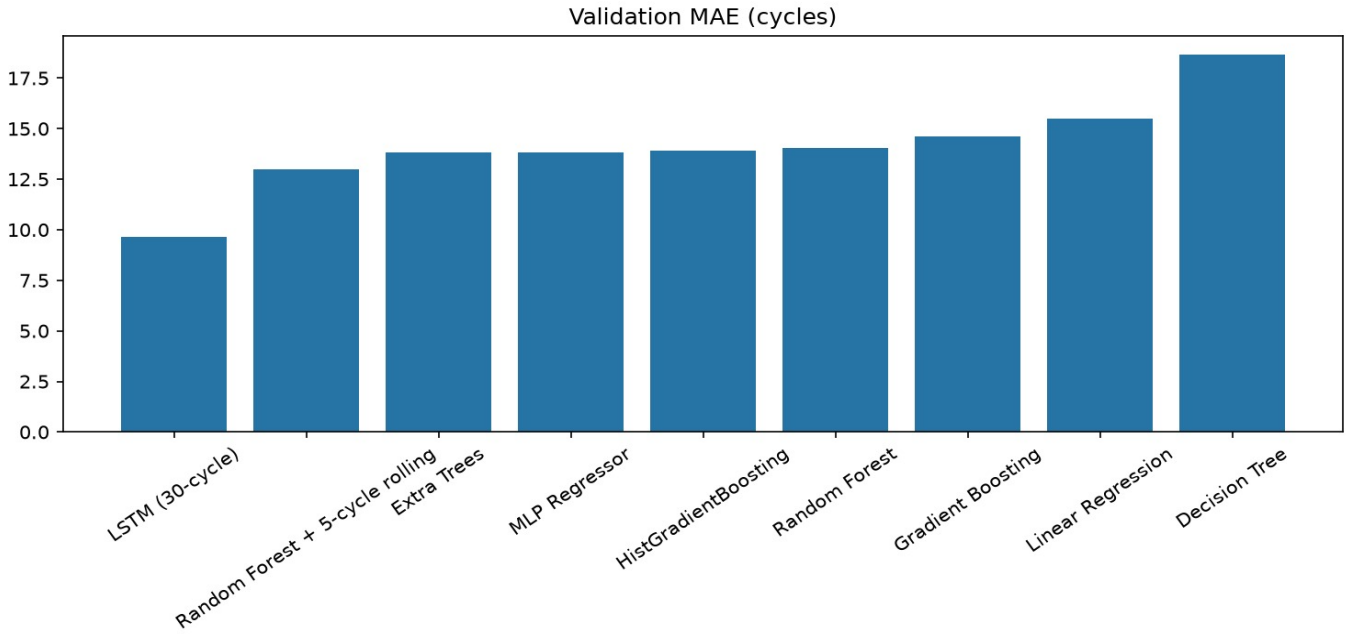

Figure. 1. Comparison of validation MAE values for the evaluated machine learning models.

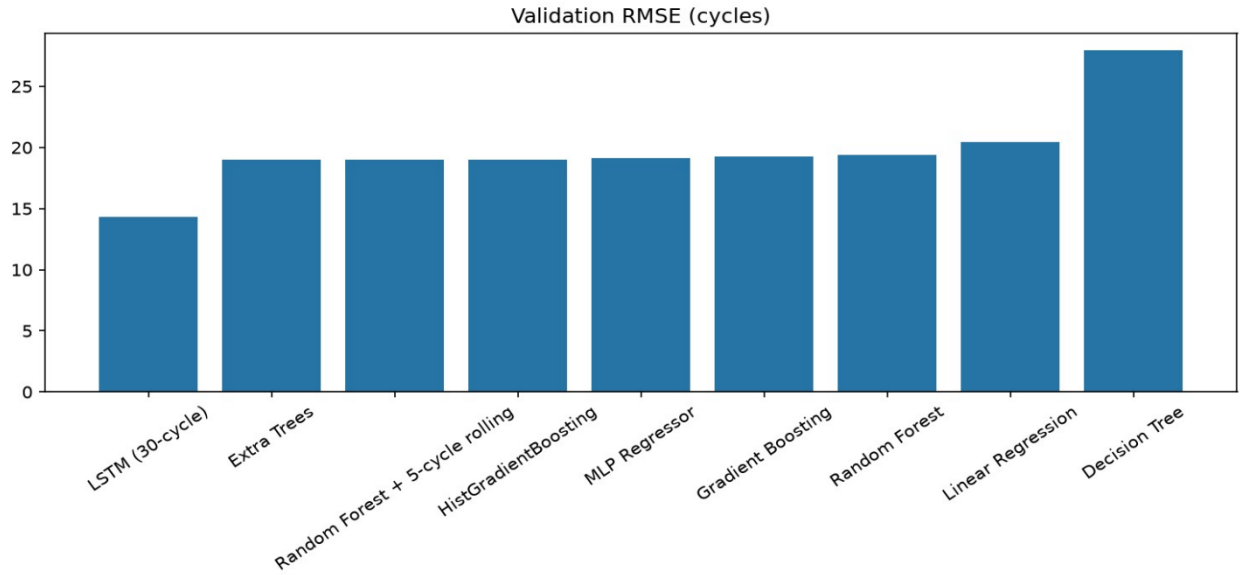

Figure. 2. Comparison of validation RMSE values for the evaluated machine learning models.

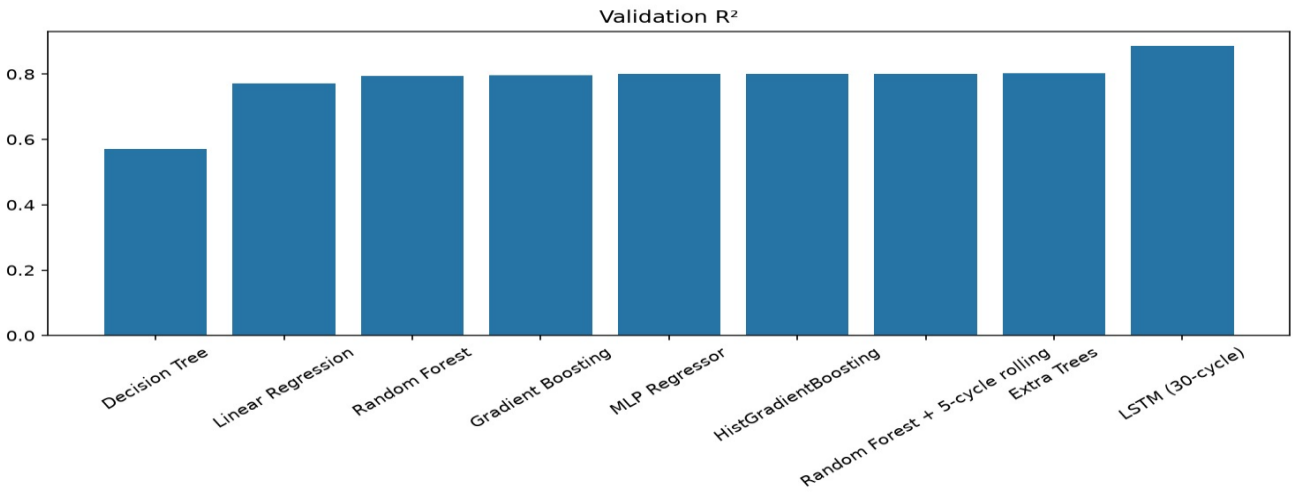

Figure. 3. Comparison of validation R² scores for the evaluated machine learning models.

The comparison shows that the LSTM model achieved the lowest MAE and RMSE and the highest R^2 score. Based on these results, the LSTM (30-cycle) model was selected for the RUL prediction component of the AEROTWIN system.

IV. IMPLEMENTATION

AEROTWIN was developed as a digital twin-based monitoring application to assess jet engine health and support predictive maintenance. The system brings together Remaining Useful Life (RUL) prediction, acoustic condition analysis, sensor anomaly detection, and decision-level fusion. The NASA C-MAPSS FD001 dataset was used to study engine degradation, while DGEN 380 recordings provided the data for acoustic analysis. These components were connected through a FastAPI backend, with a Streamlit dashboard used to display engine conditions, predictions, and monitoring results.

The implementation began with preparing the engine sensor data for RUL prediction. The NASA C-MAPSS FD001 dataset contains simulated sensor measurements collected from turbofan engines over successive operating cycles. Fourteen sensors were selected from the available measurements. The data was arranged by engine ID and operating cycle to maintain the correct sequence of readings. RUL values were calculated from the training data and capped at 125 cycles. After scaling the selected sensor readings, they were grouped into sequences of 30 operating cycles for the LSTM model. When an engine had fewer than 30 recorded cycles, its initial sensor readings were repeated to complete the required sequence length.

For RUL prediction, a Long Short-Term Memory (LSTM) neural network was implemented. LSTM was chosen because it can learn patterns from sensor readings collected over time, making it suitable for analysing changes in engine condition. The model takes 14 sensor features across 30 operating cycles as input and predicts the remaining useful life of the engine. During development, several machine learning approaches were tested, including Linear Regression, Decision Tree, Random Forest, Extra Trees, Gradient Boosting, HistGradientBoosting, MLP Regressor, and LSTM. After comparing their validation results, the LSTM (30-cycle) model was selected for the final prediction module. Its output is then passed to the monitoring interface, where users can view the estimated RUL and understand the engine's current condition.

In addition to sensor-based prediction, an acoustic analysis module was developed using DGEN 380 recordings. These recordings include microphone and accelerometer measurements, but the evaluated classification model uses features extracted from microphone signals only. The audio signals were divided into overlapping segments, and useful features were extracted from both the time and frequency domains. Mel-Frequency Cepstral Coefficients (MFCCs) were also included to capture important characteristics of the recorded engine sounds. A Random Forest classifier was then used to classify the extracted features into locally labelled normal and faulty acoustic conditions. The results are displayed in the application along with relevant signal information. Since the acoustic and sensor datasets were collected separately, their outputs are evaluated independently rather than treated as measurements from the same engine operating condition.

To identify unusual behaviour in the sensor readings, an Isolation Forest model was also included. This model analyses selected sensor measurements and assigns anomaly scores based on how much the readings differ from the patterns learned during training. Readings identified as unusual are flagged for further attention. This provides an additional way to monitor engine condition alongside the LSTM-based RUL predictions. However, these anomaly flags are not considered confirmed mechanical faults, as the evaluation data does not include independently verified physical fault labels.

A rule-based decision-level fusion mechanism was implemented to combine the sensor condition and acoustic classification outputs. The fusion module takes the status obtained from sensor analysis and the result of the acoustic classifier to determine an overall monitoring status. For example, if the acoustic module reports an abnormal condition, the system can raise the monitoring status to indicate that further inspection may be needed. Similarly, when the sensor analysis identifies a critical condition, the system retains that warning even if the

acoustic result is normal. The fusion mechanism also checks for missing or inconsistent inputs to avoid displaying misleading information. Its functionality was tested using seven controlled software cases, all of which produced the expected results.

The different machine learning components were integrated using a FastAPI backend. This backend manages access to engine sensor information, RUL predictions, acoustic classification results, and the combined monitoring status. A Streamlit dashboard was developed to make these results easier to access and understand. The dashboard provides sections for engine monitoring, digital twin visualisation, performance analytics, acoustic analysis, and maintenance-related information. Users can view sensor trends, predicted RUL values, and other condition indicators directly through the interface without working with the underlying datasets or machine learning models.

Overall, the implementation brings the LSTM-based RUL prediction model, Random Forest acoustic classifier, Isolation Forest anomaly detection, and rule-based fusion mechanism into a single monitoring application. By integrating these components, AEROTWIN provides a common platform for examining different aspects of jet engine condition. However, because the system uses simulated engine sensor measurements and separately collected acoustic recordings, it is best described as a data-driven digital twin research prototype rather than a fully synchronised digital twin connected to a real aircraft engine.

V. METHODOLOGY

The AEROTWIN system follows a data-driven approach to monitor jet engine health and support predictive maintenance. It brings together Remaining Useful Life (RUL) prediction, acoustic analysis, sensor anomaly detection, and decision-level fusion. Two datasets were used for this purpose: NASA C-MAPSS FD001 to study engine degradation and DGEN 380 recordings to analyse acoustic conditions. Both datasets were processed separately, and their results were brought together through a software-based monitoring framework. The methodology covers data preprocessing, feature extraction, model training, performance evaluation, and the presentation of results through the dashboard.

The process began with preparing the sensor data from the NASA C-MAPSS FD001 dataset, which contains simulated turbofan engine readings recorded over multiple operating cycles. Fourteen sensors were selected for RUL prediction: sensors 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, and 21. The readings were arranged according to engine ID and operating cycle to maintain their chronological order. For each training engine, RUL was calculated by subtracting the current operating cycle from the engine's maximum recorded cycle. The calculated values were then capped at 125 cycles to maintain a consistent prediction range.

Once the data was prepared, the selected sensor readings were scaled and arranged into sequences of 30 operating cycles. These sequences formed the input to the Long Short-Term Memory (LSTM) model. For engines with fewer than 30 available cycles, the first sensor reading was repeated until the required sequence length was reached. This ensured that all inputs had the same dimensions, regardless of the number of recorded cycles. A separate experiment was also conducted using Random Forest with five-cycle rolling averages to study whether smoothing the sensor readings could improve prediction performance. However, these rolling features were not used in the final LSTM model.

The LSTM model was developed to predict the remaining useful life of an engine by learning patterns from its sensor readings over time. The model uses 14 input features, a hidden size of 32, and a fully connected regression layer to produce the RUL estimate. Before training, the target RUL values were divided by 125 to normalize them. The model was trained for 25 epochs using the Adam optimizer, Mean Squared Error (MSE) loss function, a learning rate of 0.001, and a batch size of 256. After prediction, the output was converted back into engine operating cycles and limited to values between 0 and 125.

To identify the most suitable model, several machine learning algorithms were evaluated, including Linear Regression, Decision Tree, Random Forest, Extra Trees, Gradient Boosting, HistGradientBoosting, MLP Regressor, and LSTM (30-cycle). The dataset was divided based on engine identity to prevent readings from the same engine appearing in both training and validation sets. Eighty engines were used for training, while the remaining 20 engines were used for validation, with 120 observations selected for the validation evaluation. Model performance was compared using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score. Among the evaluated approaches, the LSTM model achieved the lowest validation MAE and was therefore selected for the RUL prediction module.

The selected LSTM model was further tested using the official NASA C-MAPSS FD001 test records. This evaluation considered the final recorded operating cycle of each of the 100 test engines along with its corresponding RUL value. The predicted values were compared with the actual RUL values to determine prediction errors. Residual analysis was performed to examine the differences between actual and predicted values, while engine-wise absolute error plots were used to identify variations in prediction accuracy across individual engines. The validation and official-test results were reported separately because they were obtained using different evaluation observations. Since the official test labels had also been used in earlier project comparisons, this was taken into account when interpreting the results.

For acoustic analysis, DGEN 380 recordings were used to examine differences between locally labelled normal and faulty acoustic conditions. Although the recordings contain both microphone and accelerometer measurements, only microphone signals were used for the evaluated classification model. The experiment included four recordings, with two labelled as normal and two as faulty. Each microphone signal was divided into two-second segments with 50% overlap. A total of 37 numerical features were extracted from these segments, including time-domain features, frequency-domain characteristics, and Mel-Frequency Cepstral Coefficients (MFCCs). These features helped represent different characteristics of the recorded engine sounds.

A Random Forest classifier containing 100 decision trees was trained using the extracted acoustic features. To evaluate the model, recording-grouped validation was used so that windows from the same recording would not appear in both training and testing data. The evaluation was carried out in two paired folds, with one normal-labelled and one faulty-labelled recording used for testing in each fold. The classifier's performance was measured using accuracy, precision, recall, F1-score, ROC-AUC, and a confusion matrix. However, because the recordings were collected under different operating configurations associated with their labels, the results were considered an initial acoustic classification experiment rather than reliable evidence of physical engine fault detection.

An Isolation Forest model was also implemented to identify unusual patterns in the engine sensor readings. The model was trained using sensor observations from the project-defined healthy operating region, where the capped RUL was equal to 125 cycles. It analysed the sensor values and assigned anomaly scores based on how much each observation differed from the patterns learned during training. A threshold was then used to flag unusual readings. The proportion of flagged observations was compared between the healthy-region and degradation-stage groups. Since these groups were defined using RUL values rather than independently verified fault labels, the resulting anomaly rates were not treated as physical fault-detection accuracy.

The methodology also includes a rule-based decision-level fusion mechanism that combines sensor condition information with acoustic classification results. The mechanism uses predefined rules to determine the overall engine monitoring status. When the acoustic result is normal or unavailable, the sensor-based condition is maintained. If suspicious or abnormal acoustic activity is reported, certain monitoring states may be raised to indicate the need for further inspection. Critical sensor conditions remain unchanged regardless of a normal acoustic result. Missing or inconsistent inputs are handled separately to avoid producing misleading information. The fusion mechanism was tested using seven controlled software cases to confirm that the implemented rules produced the expected outputs.

Finally, all the analysis components were connected through a FastAPI backend, and their results were presented using a Streamlit dashboard. The dashboard allows users to view predicted RUL values, engine sensor trends, acoustic analysis results, anomaly indicators, and maintenance-related information in one place. Since the NASA C-MAPSS and DGEN 380 datasets contain measurements from different sources rather than synchronized recordings of the same engine, their outputs are combined at the software decision level instead of being used in a single multimodal prediction model. Through this approach, AEROTWIN demonstrates how different machine learning techniques can be integrated into a common platform for jet engine health monitoring and predictive maintenance research.

VI. RESULT ANALYSIS

The AEROTWIN system was evaluated to understand how well its different components performed, including Remaining Useful Life (RUL) prediction, acoustic condition classification, sensor anomaly detection, and decision-level fusion. The NASA C-MAPSS FD001 dataset was used to study engine degradation, while DGEN 380 recordings were used for acoustic analysis. The performance of each component was examined using evaluation metrics, prediction errors, anomaly scores, and software test cases. The analysis also considered how these components worked together within the engine monitoring application.

The RUL prediction analysis focused on the LSTM model, which uses 14 engine sensor parameters collected over sequences of 30 operating cycles. During validation, the model achieved a Mean Absolute Error (MAE) of 9.6272 cycles, a Root Mean Squared Error (RMSE) of 14.3507 cycles, and an R^2 score of 0.8870. These were the best validation results among the main machine learning models compared in the project. The lower prediction errors suggest that the LSTM was able to identify useful patterns in the sequential sensor data, making it a suitable choice for estimating the remaining operating life of turbofan engines.

The selected LSTM model was then evaluated using the official NASA C-MAPSS FD001 test records. This evaluation included the final available operating cycle of each of the 100 test engines. The model achieved an MAE of 11.0679 cycles, an RMSE of 15.9482 cycles, and an R^2 score of 0.8416. A comparison of actual and predicted RUL values showed that many predictions were reasonably close to the expected values, although some engines had larger prediction errors. This indicates that the model learned important degradation patterns from the dataset, but its performance differed across individual engine operating histories.

To understand these differences, residual analysis was carried out by examining the gap between predicted and actual RUL values. Both positive and negative residuals were observed, showing that the model sometimes predicted a longer remaining life and sometimes a shorter one. The engine-wise absolute error analysis also revealed that prediction errors were not equally distributed among the test engines. Some engines had relatively accurate predictions, while others showed noticeably larger deviations. These findings show why it is important to examine individual predictions in addition to overall metrics such as MAE and RMSE. Another limitation is that the official FD001 test labels had been used during earlier stages of project development, so the reported test performance should not be considered an evaluation on completely untouched data.

The acoustic classification module was evaluated separately using microphone recordings from the DGEN 380 dataset. Four recordings were used in the experiment and divided into 1,716 overlapping signal windows. A Random Forest classifier was trained using features extracted from the microphone signals to distinguish between locally labelled normal and faulty acoustic conditions. The evaluation resulted in an accuracy of 51.22%, precision of 0.5092, recall of 0.6760, F1-score of 0.5809, and ROC-AUC of 0.4625. These results indicate that the model had difficulty distinguishing the two acoustic conditions with the available data. Since the recordings were collected under different operating configurations associated with their labels, the results do not provide sufficient evidence that the classifier can reliably detect actual mechanical faults.

For sensor anomaly detection, an Isolation Forest model was used to identify unusual patterns in engine sensor measurements. The evaluation considered two groups based on their RUL values. In the group with capped

RUL equal to 125 cycles, the model flagged 10 out of 1,570 records as anomalous, resulting in a rate of 0.637%. In comparison, 1,174 out of 2,500 records were flagged in the group with RUL below 125 cycles, giving a rate of 46.960%. This difference suggests that unusual sensor patterns became more common in observations associated with engine degradation. However, these rates were calculated using RUL-based groups rather than independently confirmed fault labels. Therefore, they indicate how frequently the model identified anomalies and should not be treated as verified fault-detection accuracy.

The decision-level fusion mechanism was tested to check how the system handled different combinations of sensor and acoustic information. Seven controlled software test cases were used, covering healthy engine conditions, moderate degradation, abnormal acoustic results, critical sensor conditions, missing sensor or acoustic information, and inconsistent responses. All seven cases produced the expected outputs. The tests showed that critical sensor warnings were maintained, normal or unavailable acoustic information did not unnecessarily change the sensor status, and inconsistent inputs were handled appropriately. These results confirm that the implemented decision rules functioned as intended. However, the tests did not establish whether combining sensor and acoustic information improved prediction or fault-detection accuracy.

The AEROTWIN dashboard was used to bring the results of the different analysis components into one interface. It displays predicted RUL, engine health status, sensor measurements, acoustic classification results, anomaly indicators, and maintenance-related information. A FastAPI backend connects the machine learning components to the Streamlit dashboard, allowing users to view the available engine condition information without directly accessing the individual models. This integration makes it easier to examine different aspects of engine health in one place. However, the sensor and acoustic datasets were collected independently and do not contain synchronized measurements from the same physical engine.

Overall, the evaluation shows that the LSTM (30-cycle) model achieved the best RUL prediction results among the approaches tested in AEROTWIN. Residual and engine-wise error analyses revealed differences in prediction accuracy across engines. The acoustic classification model showed limited performance, while the Isolation Forest identified a higher proportion of unusual sensor readings in observations associated with lower RUL values. The decision-level fusion tests also confirmed that the implemented software rules worked as expected. Together, these results demonstrate the integration of multiple analysis techniques into a single engine monitoring and predictive maintenance research prototype. Further testing with independently verified fault data, synchronized sensor and acoustic measurements, and physical engine experiments would be necessary to determine its reliability and suitability for real-world aircraft maintenance.

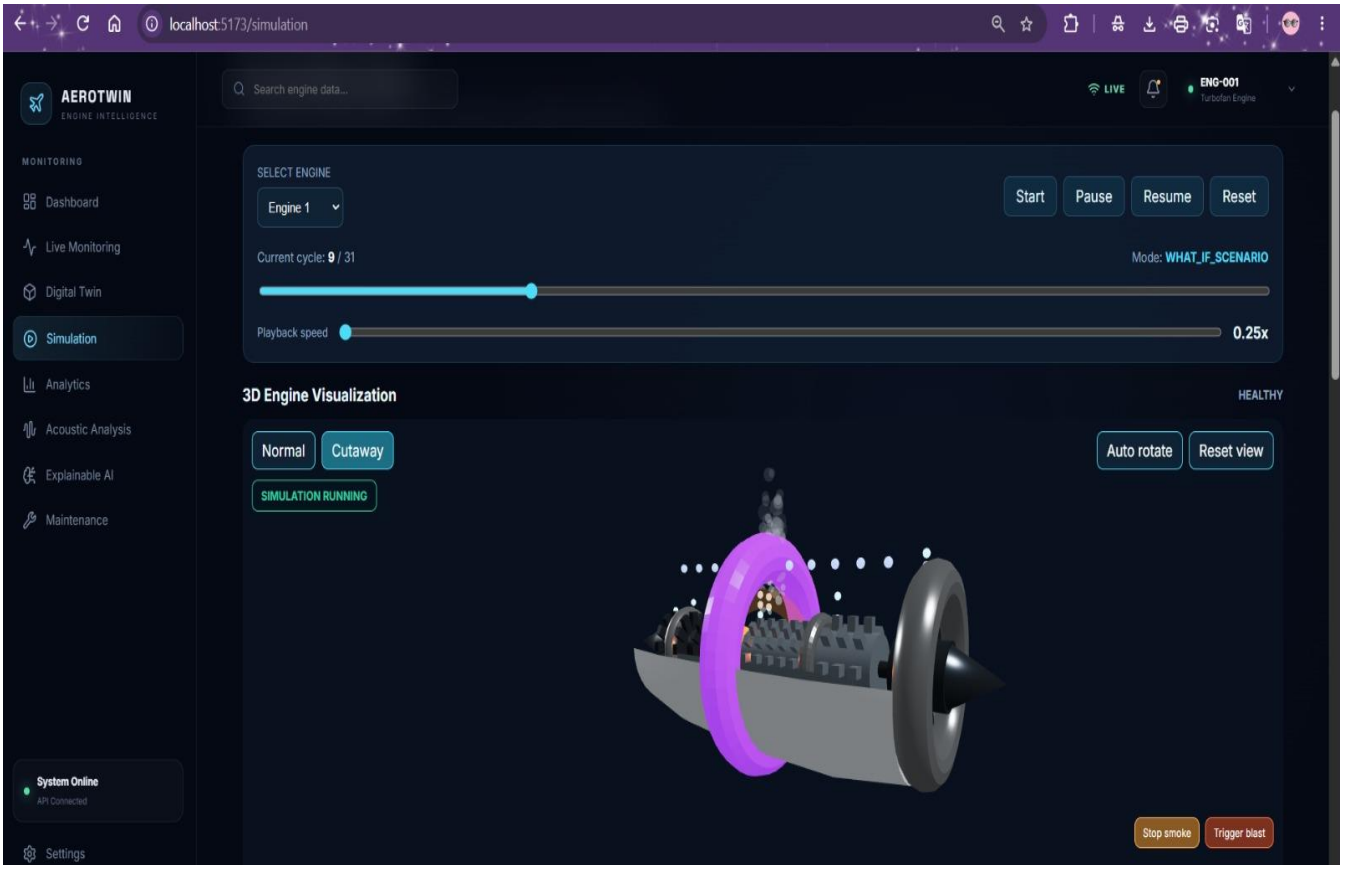

Figure 4: Simulation Results and Engine Health Analysis

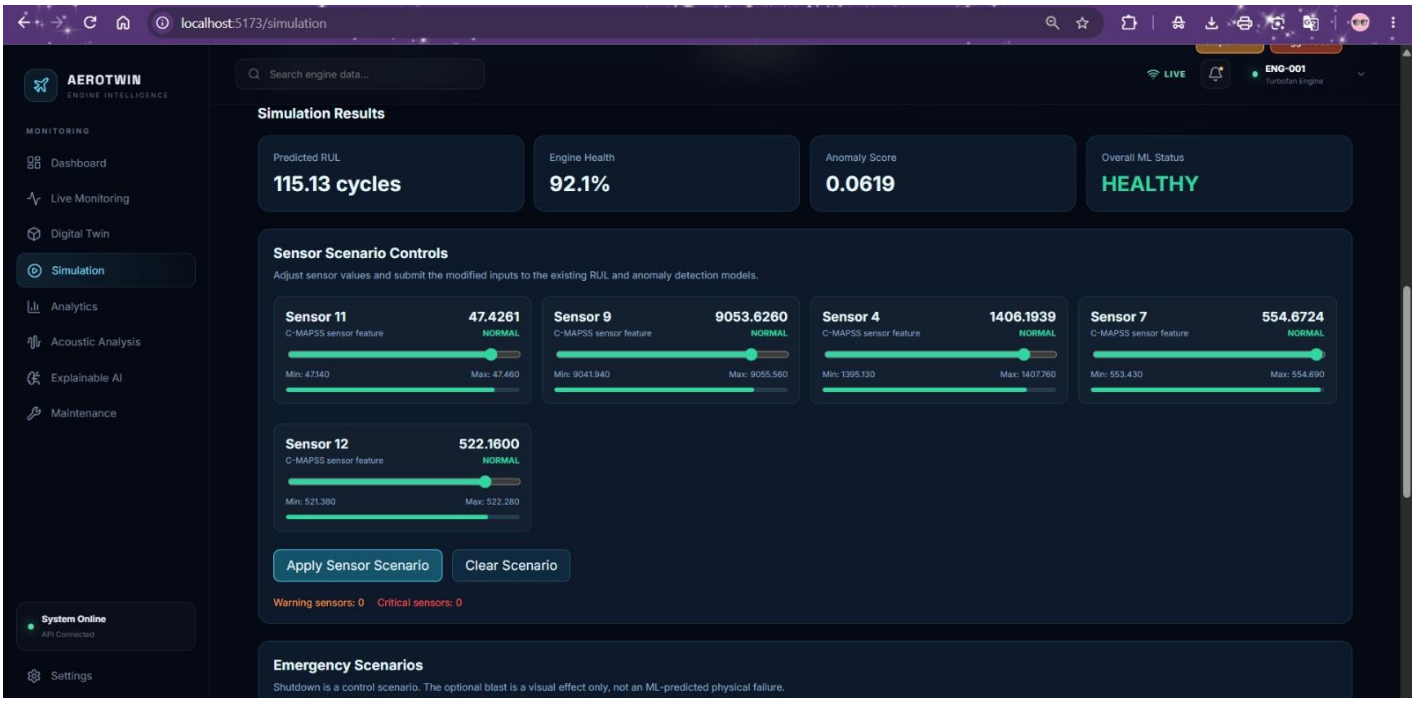

Figure 5: 3D Jet Engine Digital Twin Simulation

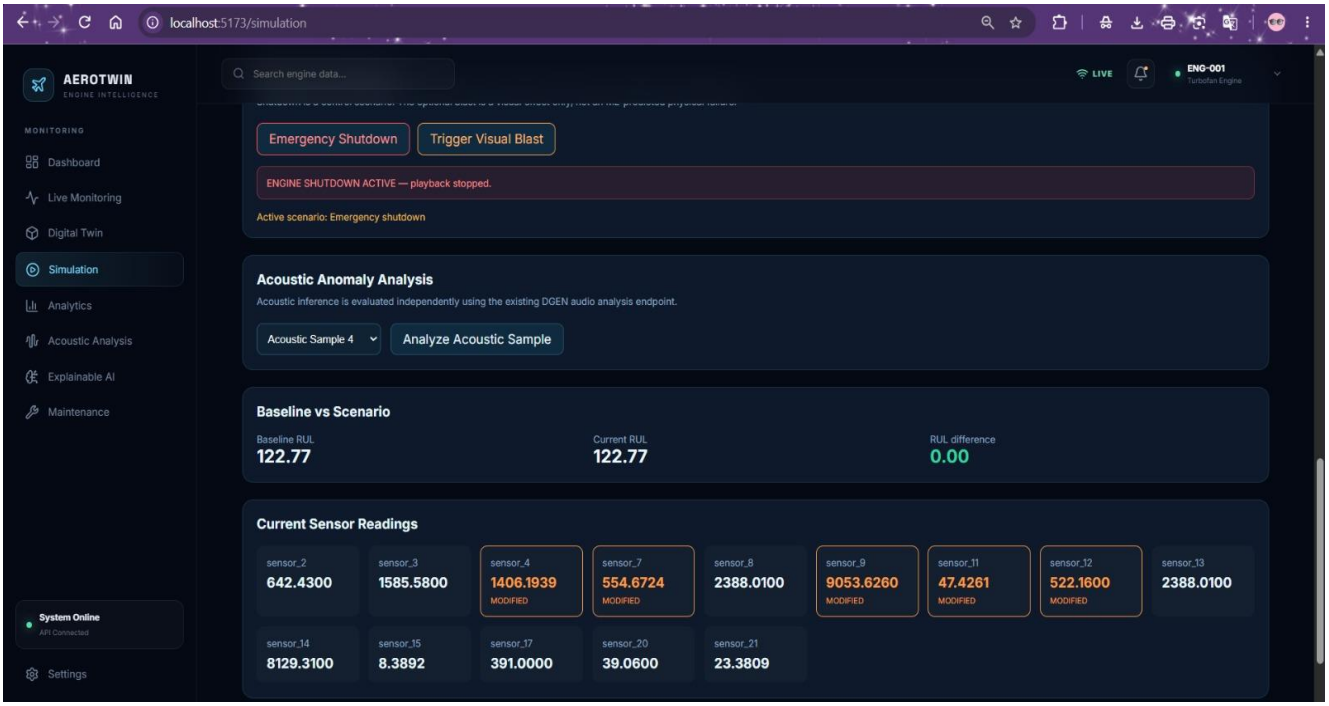

Figure 6: Emergency Shutdown and Sensor Anomaly Detection

V. CONCLUSION

The AEROTWIN project demonstrates how digital twin technology and machine learning can be used to monitor jet engine health and support predictive maintenance. The system combines Remaining Useful Life (RUL) prediction, acoustic analysis, sensor anomaly detection, and a monitoring dashboard. The NASA C-MAPSS FD001 dataset was used to study engine degradation, while DGEN 380 recordings were used for acoustic analysis.

Several machine learning models were evaluated, and the LSTM (30-cycle) model achieved the best RUL prediction results. On the reported FD001 test evaluation, it obtained an MAE of 11.0679 cycles, RMSE of 15.9482 cycles, and R^2 score of 0.8416. Random Forest was used for acoustic classification, Isolation Forest for detecting unusual sensor patterns, and rule-based fusion for combining the monitoring results.

The developed dashboard presents engine health information, predicted RUL, and maintenance-related indicators in a single interface, making the results easier to understand. While the RUL prediction model showed promising performance, the acoustic classification module needs further improvement and testing. Overall, AEROTWIN serves as a research prototype that brings together different machine learning techniques for jet engine monitoring. With further testing using real engine data, the system could provide a foundation for more reliable predictive maintenance solutions.

REFERENCES

[1] I. Kabashkin, "Digital Twin Framework for Aircraft Lifecycle Management Based on Data-Driven Models," *Mathematics, 2024.*

[2] D. Zhong, Z Xia, Y. Zhu, and J. Duan," Overview of Predictive Maintenance Based on Digital Twin Technology," *Heliyon, vol. 9, no. 4, Art. no. e14534, 2023.*

[3] L. Li, S. Aslam, A. Wileman, and S. Perinpanayagam, "Digital Twin in Aerospace Industry: A Gentle Introduction," *IEEE Access, vol. 10, pp. 9543–9562, 2022.*

[4] C. J. Wright, N. Biederman, B. Gyovai, D. J. Gauthier, and J. P. Wilhelm, "Small Jet Engine Reservoir Computing Digital Twin," *arXiv preprint arXiv:2312.09978, 2023.*

[5] Y. Zhang, X. Tang, and X. Zhang, "Remaining Useful Life Predictions for Turbofan Engine Using Semi-supervised DBN-LSTM Joint Training Model," *in Proceedings of the 2022 International Conference on Control, Decision and Information Technologies (CoDIT), pp. 94–99, 2022.*

[6] S. Sharma, A. K. Pandit, and S. Sharanya, "Predicting Degradation with RNN," *2024.*

[7] K. Ensarioğlu, T. İnkaya, and E. Emel, "Remaining Useful Life Estimation of Turbofan Engines with Deep Learning Using Change-Point Detection Based Labeling and Feature Engineering," *Applied Sciences, vol. 13, no. 21, Art. no. 11893, 2023.*

[8] Q. Ma, M. Zhang, Y. Xu, J. Song, and T. Zhang, "RUL Prediction Using Transformers," *2021.*

[9] Z. Wu and J. Li, "A Framework of Dynamic Data Driven Digital Twin for Complex Engineering Products: The Example of Aircraft Engine Health Management," *Procedia Manufacturing, vol. 55, pp. 139–146, 2021.*

[10] S. Das, D. Sengupta, R. Chandrasekharan, and S. Mitra, "Detection of Anomalous Behaviour of Aircraft Engines for Predictive Maintenance," *in Proceedings of the 2023 7th International Conference on System Reliability and Safety (ICSRS), pp. 423–427, 2023.*

[11] H. Tian, L. Yang, and B. Ju, "Spatial Correlation and Temporal Attention-Based LSTM for Remaining Useful Life Prediction of Turbofan Engine," 2023.

Copyright & License:

© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.