Prediction of Failures in Air Pressure System: A Semi-supervised Framework Based on Transformers

Shawqi Mohammed Farea,
Mehmet Emin Mumcuoglu, Mustafa Unel
Faculty of Engineering and Natural Sciences
Sabanci University
Istanbul, Turkey
munel@sabanciuniv.edu

Serdar Mise, Simge Unsal, Enes Cevik,
Metin Yilmaz, Kerem Koprubasi
Product Development
Ford OTOSAN
Istanbul, Turkey
kkopruba@ford.com.tr

***Abstract*—The air pressure system (APS) plays a prime role in pressurizing various subsystems of heavy-duty vehicles (HVDs). However, its reliability is crucial to ensure uninterrupted operation where failures in APS lead to HVDs being stranded on the road with the manufacturers and operators incurring associated high costs. This paper addresses the problem of predicting failures in APS using a semi-supervised transformer-based framework. The proposed framework commences with important preprocessing steps including data segmentation followed by sliding windows to handle the big raw data, and subsequent extraction of distinctive features. Using these features, the transformer model was trained to reconstruct data from healthy vehicles (i.e., vehicles without any APS failures) to capture the normal behavior of the healthy vehicles. At inference, the trained model distinguished the faulty vehicles with detected APS failure from the healthy ones based on their reconstruction errors. This semi-supervised formulation of APS failure detection overcomes limitations such as the imbalanced data issue and anomaly heterogeneity that are associated with the conventional supervised formulation. The model demonstrated robust performance with an F1 score of approximately 0.76, an accuracy of about 85%, and a high recall of 0.833, indicating successful detection of most faulty vehicles. Such advancements promise significant improvements in vehicle diagnostics and predictive maintenance.**

***Index Terms*—air pressure system (APS), semi-supervised transformer, fault detection, heavy-duty vehicles (HVDs), predictive maintenance.**

I. Introduction

Heavy-duty vehicles (HDVs) stand as the backbone of modern transportation and logistics industries. Despite their indispensable role, the reliability of HDVs remains a critical concern for manufacturers, fleet operators, and end-users alike. In HDVs, mechanical failures, downtime, and associated maintenance costs not only disrupt operations but also pose significant economic and safety risks. Therefore, there is a pressing need to enhance the reliability and efficiency of HDVs through advanced technologies such as predictive maintenance.

This study tackles the problem of predicting failures in the air pressure system (APS). APS provides the required pressurized air for HVDs' subsystems such as braking and suspension. Given its pivotal role, APS failures normally result in HVDs being stranded on the road. In general, a stranding incident leads to costly roadside assistance, operation disruption, and resultant customer dissatisfaction. Failures in APS can stem from sensor malfunctions or internal mechanical issues, e.g., valve failures. Early prediction of such issues decreases the chance of HVDs being stranded on the road and avoids the corresponding cost and downtime. However, prior detection of APS failures during regular maintenance inspections is still challenging and requires substantial domain knowledge.

Given their proven efficacy across different fields, machine learning (ML) techniques have been utilized in the literature for predicting APS failures as well. For instance, Prytz et al. [1], [2] used supervised classifiers based on k-nearest neighbors (KNN), decision trees, and random forest to detect APS failures in a fleet of Volvo trucks. In another study [3], the same research group implemented a fuzzy rule-based model to deal with APS failures in the same Volvo fleet. Fan et al. [4]–[6] employed different variants of a Consensus Self-Organizing Model (COSMO) to predict APS-related faults in a fleet of Volvo buses. On the other hand, multiple decision models were proposed in [7] to detect APS failures in medium-duty vehicles while utilizing the available diagnostic trouble codes and collected operational data.

Nonetheless, the existing literature formulated the problem of APS failure detection as a supervised classification. This inferior formulation of the problem suffers from the imbalanced data issue since the positive class (i.e., vehicles with APS failures) is small in size compared to the negative class (i.e., vehicles without APS failures). Additionally, it presumes that all failure types fall into a single class although failures are known to be heterogeneous. In accordance with its nature, APS failure detection should be formulated as a semi-supervised problem instead. In this alternative formulation, the training set includes exclusively healthy data (i.e., data from vehicles without any APS failures) to capture normal behavior. Accordingly, the trained model is expected to not be able to reconstruct the data from faulty vehicles as accurately as the data from healthy ones. Intuitively, the reconstruction error is then used as the anomaly score to distinguish between healthy and faulty data.

Various semi-supervised techniques – including one-class support vector machine (SVM), support vector data description (SVDD), and autoencoders – have been used in the automotive literature for failure detection in general. For example, a one-class SVM classifier was used for predicting failures related to an internal combustion engine in [8] and braking system failures in [9]. In addition, an SVDD was employed in [10] to detect lithium battery failures. On the other hand, autoencoders were used for detecting faulty sensors in [11] and powertrain faults in [12]. Semi-supervised models based on transformers have also been proposed to detect lithium battery faults [13], [14]. Yet, the application of semi-supervised techniques to APS failure detection is still lacking.

In this paper, a semi-supervised framework based on a transformer model was proposed for failure prediction in APS. The main contributions are as follows:

- A dataset containing 30-day operational data from 77 healthy HDVs and 30 faulty HDVs – i.e., vehicles that have experienced APS failures and subsequent replacements in their APS.
- Multiple preprocessing steps in order to process the big data and extract relevant, meaningful features facilitating the distinction between healthy and faulty vehicles and the development of an effective ML model.
- A transformer-based model is proposed to tackle the problem of APS failure detection as semi-supervised anomaly detection.

The rest of this paper is organized as follows: Section II provides a brief description of the collected data and introduces the multiple preprocessing steps needed to prepare the data for the subsequent model. Section III explains in detail the transformer model while Section IV reports and discusses the experimental results. Finally, Section V concludes the paper.

II. Data Description and Preprocessing

A. Data Description

The collected dataset comprises 30-day operational data from 77 healthy vehicles and 30 faulty vehicles. These single-brand vehicles are Ford F-Max heavy-duty trucks operating in Turkey and Europe. The faulty vehicles have experienced APS failure and subsequent replacements of their electronic air pressure units (E-APU), the core unit in APS. The data from these faulty vehicles correspond to the last 30 days just before failure occurrence. In contrast, the data from healthy vehicles, with clean maintenance records in terms of APS, represent historical 30-day driving sequences from different times throughout the year.

A summary of the collected data is presented in Table I. The overall data includes 2730 daily records from 107 Ford trucks in total. The signals related to APS are listed in Table II. While most signals were acquired periodically, Air Compressor Status and Brake Pedal Position were acquired only when a change in their values occurred. This discrepancy in the acquisition type and sampling rate leads to a significant amount of missing values in all signals.

TABLE I
Summary of Collected Data

Details	Healthy	Faulty
Number of vehicles	77	30
Number of daily records	1,959	771
Number of signals	9	
Nominal sampling rate	1 Hz	

TABLE II
List of APS-related Signals

No.	Signal name	Sampling period
1	Air Compressor Status	on change
2	Brake Pedal Position	on change
3	Service Brake Circuit 1 Air Pressure	one second
4	Service Brake Circuit 2 Air Pressure	one second
5	Parking and/or Trailer Air Pressure	one second
6	Engine Speed	one second
7	Vehicle Speed	one second
8	Total Traveled Distance	10 seconds
9	Engine Total Hours of Operation	300 seconds

B. Data Preprocessing

The first step in preprocessing data is to delete periods when vehicles were logging data but they were dynamically stationary (i.e., their engines were off). After discarding these irrelevant periods, the daily driving records were split into drive cycles where the time gap between consecutive drive cycles is at least five minutes. However, the collected dataset still includes a significant amount of missing data due to connectivity issues or the inconsistent sampling types and rates of the different signals. To circumvent that issue, data interpolation tailored according to the interpolated signal was conducted. Then, moving statistics (e.g., mean, standard deviation, and minimum) were extracted from the interpolated drive cycles using a sliding window with a length of 20 minutes and a shift of 10 minutes. This downsampling process via sliding windows has threefold advantages: it helps in dealing with the big dataset, in extracting meaningful features (e.g., the duty cycle), and in smoothing noisy data. The preprocessing workflow is depicted in Fig. 1 where 'W' and 'S' represent the length and shift of the sliding windows and 'd' represents the time gap between consecutive drive cycles.

The moving statistics extracted from sliding windows are listed in Table III. In total, 11 features were extracted from each sliding window. These features act as distinctive indicators in terms of healthy and faulty APS. Any failure in APS normally manifests itself in some of these features – e.g., duty cycle, compressor on/off count, and minimums of pressure signals – while the remaining features – e.g., mean and standard deviation of vehicle speed – serve as the context summarizing the dynamical status of the vehicle.

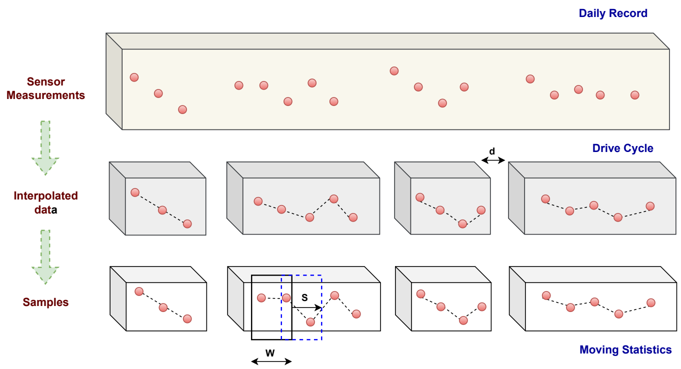

Fig. 1. Data preprocessing workflow: segmenting daily driving records into drive cycles, applying data interpolation, and finally conducting sampling via moving statistics.

TABLE III
LIST OF EXTRACTED FEATURES OVER SLIDING WINDOWS

No.	Extracted feature
1	duty cycle
2	compressor on/off count
3	min. of service brake circuit 1 air pressure
4	min. of service brake circuit 2 air pressure
5	min. of parking and/or trailer air Pressure
6	traveled distance
7	mean of brake pedal position
8	mean of engine speed
9	std. of engine speed
10	mean of vehicle speed
11	std. of vehicle speed

III. TRANSFORMER-BASED FAILURE DETECTION

Transformers were proposed in [15], and they have gained growing popularity since then. These deep learning models have achieved state-of-the-art results in fields ranging from natural language processing (NLP) [16] to computer vision [17] to speech recognition [18] and beyond. Time series anomaly detection is another field in which transformers showed promising results [19], [20]. In the automotive sector, transformers have also proved their efficacy in applications like autonomous parking [21], collision avoidance [22], and path planning [23].

Transformers' architecture generally consists of an encoder and decoder where multi-head attention layers serve as the core unit of both. These attention layers enable the architecture to attend to parts at different locations in the input sequence and model global representations of this sequence accordingly. Therefore, transformers, unlike recurrent neural networks (RNN) and their variants, have the ability to model long-term temporal dependencies of the input sequence without suffering from the problem of vanishing gradients. Furthermore, all the elements (i.e., tokens) of the input sequence can be processed simultaneously in transformers. This enhances the parallelism of these architectures and accelerates the underlying training and inference processes [15].

The transformer architecture used in this work is depicted in Fig. 2. Both the encoder and decoder consist of a single layer each. As the input sequence was normalized to fall into the [0,1] range, the sigmoid function is used as the activation function of the output layer. This normally yields better results as suggested in [20]. The model was trained using only data from healthy vehicles in a semi-supervised fashion. The main aim of the model is to reconstruct the input sequences based on the latent representations extracted by the multi-head attention layers in the encoder and decoder. Each input sequence consists of 10 sample data points where each data point contains the moving statistics of a sliding window – the features listed in Table III. According to the semi-supervised learning scheme, the trained model is expected to reconstruct data from healthy vehicles more accurately than those from faulty vehicles as the trained model has already seen and learned the general behavior of healthy data. Hence, the reconstruction error can be used as the anomaly score where high anomaly scores indicate potential faulty instances. The overall loss function, which is also defined as the anomaly score, is as follows:

$$Loss = ||\hat{S}_t - S_t|| \tag{1}$$

where S_t and $\hat{S}_t$ are the input sequence and reconstructed sequence, respectively.

Accordingly, each input sequence will be associated with an anomaly score based on its reconstruction error. Nevertheless, the anomaly score should be vehicle-level instead of sequence-level. In other words, each vehicle has to have a single anomaly score to predict whether the APS in this vehicle is faulty or not. To address this issue, the anomaly score per vehicle is defined as the median of the anomaly scores of the sequences belonging to that vehicle. The median is used since it is more robust against noisy scores as opposed to the mean. To convert the anomaly scores into binary labels, an optimal threshold in terms of the F1 score is found using a grid search.

IV. RESULTS AND DISCUSSION

The proposed model was trained using 70% of the healthy data while the remaining 30% of the healthy data were used for validating the model. Fig. 3 shows the convergence curves of the training and validation data. Apparently, the model succeeded in reconstructing both training and validation data where reconstruction errors became very small as the training process proceeded. For instance, the reconstruction errors for the training and validation data became below 0.002 just after the fifth iteration (i.e., epoch). The model succeeded in reconstructing the validation data accurately although it did not see them during training. This indicates that the normal behavior exhibited by healthy vehicles was learned by the model. Furthermore, the model convergence is fast; the reconstruction error fell below 0.007 just after the first iteration. Given this quick convergence, the total number of iterations was set as 20, enabling a rapid training process.

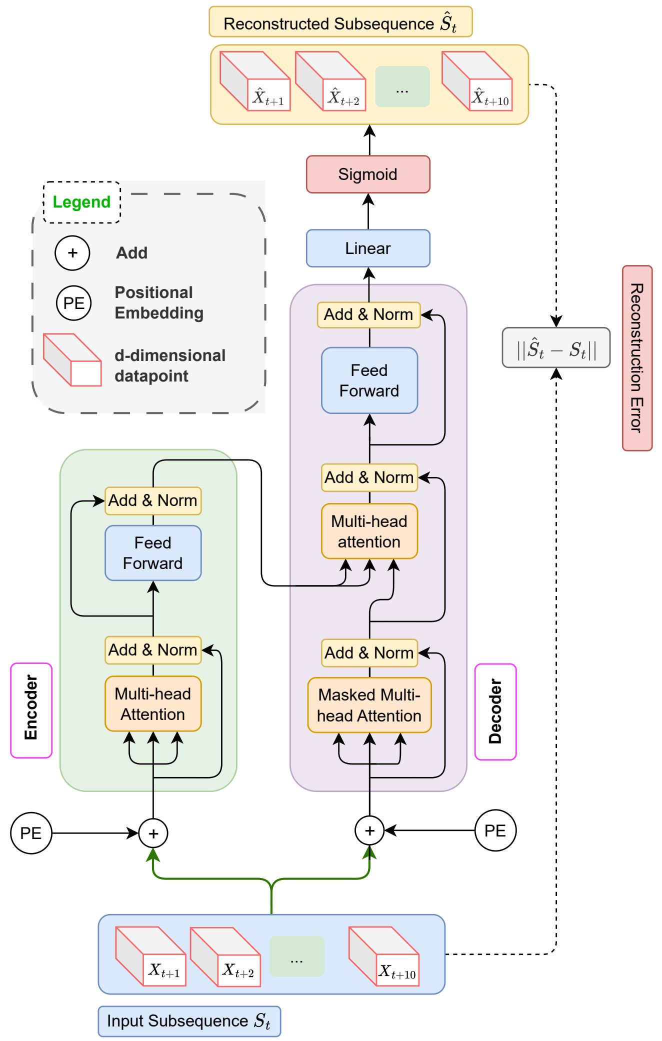

Fig. 2. Transformer architecture (the input and output of the model are a sequence of 10 d-dimensional data points where d is 11 in this work).

Using the trained model, all the healthy and faulty vehicles were assigned anomaly scores as previously explained. An optimal threshold was obtained using the grid search method, and each vehicle was predicted as healthy or faulty according to its anomaly score with respect to this threshold. The obtained confusion matrix is shown in Fig. 4. The model correctly predicted 66 healthy vehicles whilst it incorrectly predicted the remaining 11 healthy vehicles as faulty; that is, the model has 11 false positives. In contrast, the model correctly predicted 25 faulty vehicles as faulty and incorrectly predicted the remaining five faulty vehicles as healthy; that is, the model has five false negatives.

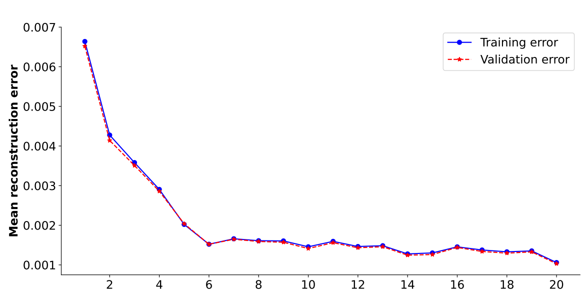

Fig. 3. Convergence curves of the transformer model (both the training and validation data are from the healthy vehicles).

The model was evaluated using the following metrics: precision, recall, F1 score, and accuracy. These evaluation metrics for the proposed model were reported in Table IV. The total accuracy of the model was approximately 85%. Nonetheless, the accuracy alone cannot be relied upon since the positive class representing faulty vehicles has a smaller size as compared to the negative class representing healthy vehicles. Thus, the other metrics should be considered as well to assess the model more accurately. The model achieved a recall of roughly 0.83 and a precision of almost 0.69. This high recall is due to the fact that most of the faulty vehicles (25 of them) were predicted correctly as faulty and the model missed only five of the faulty vehicles. On the other hand, the model predicted 36 vehicles in total as faulty; however, only 25 of these vehicles are really faulty. Accordingly, the precision was nearly 0.69. The F1 score, defined as the harmonic mean of the precision and recall, was approximately 0.76 which is a relatively high value indicating the effectiveness of the proposed model.

The transformer architecture used in this work is relatively simple with a single-layer encoder and a single-layer decoder. Yet, the model showed promising results as stated above. In particular, the model started to successfully reconstruct healthy data and even the unseen validation data at the early stage of the training process (see Fig. 3). In addition, the model predicted most of the faulty vehicles and achieved good results in terms of F1 score, recall, and prediction accuracy. Therefore,

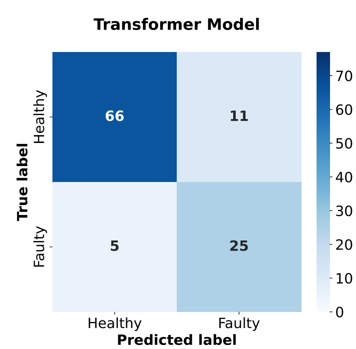

Fig. 4. Confusion matrix of Transformer model (Faulty represents the positive class while Healthy represents the negative class).

TABLE IV
RESULTS OF TRANSFORMER MODEL

Precision	Recall	F1 Score	Accuracy
0.694	0.833	0.758	85.1%

the complexity of the proposed transformer model is deemed sufficient for this problem given its good performance. The preceding preprocessing steps played a major role in enabling this relatively simple architecture to achieve good results.

V. Conclusion

The reliability of heavy-duty vehicles (HDVs) raises concerns among their manufacturers and operators alike. The air pressure system (APS) is considered a key component of HVDs that supplies pressure to various subsystems such as brakes and suspension systems. Accordingly, the early prediction of APS failures ensures the uninterrupted operation of HVDs and helps avoid the high cost and customer dissatisfaction associated with such failures. This paper proposed a semi-supervised framework to predict failures in APS. The framework is based on transformer, an efficient deep-learning model. The framework includes crucial preprocessing steps to handle the large-scale raw data by first segmenting the raw data into distinct driving subsections (called drive cycles), then interpolating missing values, and eventually applying sliding windows to extract meaningful, distinctive features. Based on these features, the transformer model reconstructed a subset of healthy data in a semi-supervised fashion to learn the general behavior of healthy data. Therefore, when reconstructed by the trained model, faulty data generally exhibited higher reconstruction errors as opposed to healthy data. The reconstruction errors were used as the anomaly scores, accordingly. The model achieved an F1 score of approximately 0.76 with a corresponding accuracy of nearly 85%. The model also succeeded in accurately predicting most faulty vehicles leading to a high recall of 0.833.

In future work, more advanced transformer architectures are to be investigated and compared against the current architecture. Another research direction for improving the proposed framework is the potential application of explainable artificial intelligence (XAI) on top of the proposed model to provide explanations for the model predictions. This helps improve the trustworthiness and transparency of the current black-box model. Also, it helps in discovering any biasedness involved with the model.

Acknowledgment

The funding provided by Ford Otosan is gratefully acknowledged.

References

[1] R. Prytz, S. Nowaczyk, T. Rögnvaldsson, and S. Byttner, "Analysis of truck compressor failures based on logged vehicle data," in *9th International Conference on Data Mining, Las Vegas, Nevada, USA, July 22–25, 2013*. CSREA Press, 2013.

[2] R. Prytz, S. Nowaczyk, T. Rögnvaldsson, and S. Byttner, "Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data," *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 139–150, 2015.

[3] S. Nowaczyk, R. Prytz, T. Rögnvaldsson, and S. Byttner, "Towards a machine learning algorithm for predicting truck compressor failures using logged vehicle data," in *12th Scandinavian Conference on Artificial Intelligence, Aalborg, Denmark, November 20–22, 2013*. IOS Press, 2013, pp. 205–214.

[4] Y. Fan, S. Nowaczyk, and T. Rögnvaldsson, "Evaluation of self-organized approach for predicting compressor faults in a city bus fleet," *Procedia Computer Science*, vol. 53, pp. 447–456, 2015.

[5] Y. Fan, S. Nowaczyk, and T. S. Rögnvaldsson, "Incorporating expert knowledge into a self-organized approach for predicting compressor faults in a city bus fleet." in *SCAI*, 2015, pp. 58–67.

[6] Y. Fan, S. Nowaczyk, E. A. Antonelo *et al.*, "Predicting air compressor failures with echo state networks," in *PHM Society European Conference*, vol. 3, no. 1, 2016.

[7] C. Panda and T. R. Singh, "Ml-based vehicle downtime reduction: A case of air compressor failure detection," *Engineering Applications of Artificial Intelligence*, vol. 122, p. 106031, 2023.

[8] D. Jung, "Data-driven open-set fault classification of residual data using bayesian filtering," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 5, pp. 2045–2052, 2020.

[9] J. Sang, J. Zhang, T. Guo, D. Zhou, M. Chen, and X. Tai, "Detection of incipient faults in emu braking system based on data domain description and variable control limit," *Neurocomputing*, vol. 383, pp. 348–358, 2020.

[10] J. Zhang, Y. Wang, B. Jiang, H. He, S. Huang, C. Wang, Y. Zhang, X. Han, D. Guo, G. He, and M. Ouyang, "Realistic fault detection of li-ion battery via dynamical deep learning," *Nature Communications*, vol. 14, no. 1, p. 5940, 2023.

[11] H. Min, Y. Fang, X. Wu, X. Lei, S. Chen, R. Teixeira, B. Zhu, X. Zhao, and Z. Xu, "A fault diagnosis framework for autonomous vehicles with sensor self-diagnosis," *Expert Systems with Applications*, vol. 224, p. 120002, 2023.

[12] A. Geglio, E. Hedayati, M. Tascillo, D. Anderson, J. Barker, and T. C. Havens, "Deep convolutional autoencoder for assessment of drive-cycle anomalies in connected vehicle sensor data," in *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2022, pp. 743–749.

[13] D. Wang, P. Ruan, D. Xu, W. Xie, X. Chen, and H. Li, "Tranad: A deep transformer model for fault diagnosis of lithium batteries," in *2023 International Conference on Smart Electrical Grid and Renewable Energy (SEGRE)*, 2023, pp. 133–139.

[14] J. Zhao, X. Feng, J. Wang, Y. Lian, M. Ouyang, and A. F. Burke, "Battery fault diagnosis and failure prognosis for electric vehicles using spatio-temporal transformer networks," *Applied Energy*, vol. 352, p. 121949, 2023.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[16] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[17] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[18] S. Kim, A. Gholami, A. Shaw, N. Lee, K. Mangalam, J. Malik, M. W. Mahoney, and K. Keutzer, "Squeezeformer: An efficient transformer for automatic speech recognition," *Advances in Neural Information Processing Systems*, vol. 35, pp. 9361–9373, 2022.

[19] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy," in *International Conference on Learning Representations*, 2021.

[20] S. Tuli, G. Casale, and N. R. Jennings, "Tranad: deep transformer networks for anomaly detection in multivariate time series data," *Proceedings of the VLDB Endowment*, vol. 15, no. 6, pp. 1201–1214, 2022.

[21] L. Wang, X. Zhang, W. Zeng, W. Liu, L. Yang, J. Li, and H. Liu, "Global perception-based robust parking space detection using a low-cost camera," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1439–1448, 2023.

[22] C. Lin, Y. Cheng, X. Wang, J. Yuan, and G. Wang, "Transformer-based dual-channel self-attention for uuv autonomous collision avoidance," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 3, pp. 2319–2331, 2023.

[23] Z. Yu, M. Zhu, K. Chen, X. Chu, and X. Wang, "Lf-net: A learning-based frenet planning approach for urban autonomous driving," *IEEE Transactions on Intelligent Vehicles*, pp. 1–14, 2023.