Traffic Anomaly Detection in Intelligent Transport Applications with Time Series Data using Informer

Xinggan Peng[1], Yuxuan Lin[1], Qi Cao[2], Yigang Cen[3], Huiping Zhuang[1] and Zhiping Lin[1]

Abstract— **Multivariate time series traffic dataset is usually large with multiple feature dimensions for long time duration under certain time intervals or sampling rates. In applications such as intelligent transportation systems, some machine learning methods being applied to traffic anomaly detections are computed under certain assumptions and require further improvements. Transport traffic time series data may also suffer from unbalanced number of training data where large amount of labelled training data available for a few popular classes, but with very small amount of labelled data for corner cases. In this paper, based on the recent long sequences prediction method Informer, an anomaly detection algorithm with an anomaly score generator is proposed that does not require any assumptions of data. The encoder-decoder architecture is adopted in the anomaly score generator. The encoder consists of three stacking ProbSparse self-attention mechanisms that significantly reduce computing complexity. The decoder incorporates two multi-head attention layers and a fully connected layer to obtain an output of anomaly scores. Then a One-Class Support Vector Machines (OCSVM) is applied to be the anomaly classifier. The proposed algorithm is capable of detecting anomalies for both vehicle traffic flows and pedestrian flows. It has been verified by applying to a real-world dataset consisting of traffic flows recorded in 2021, as well as to a public anomaly detection dataset.**

I. INTRODUCTION

Time series data captures information over a duration of time frames with certain time intervals or sampling rate. Time series data could be univariant or multivariant over the time axis. Time series analysis is to study data patterns for predicting future patterns or detecting data anomalies [1]. Some examples of time series analysis can be found in transportation traffic [2], [3], financial market pricing, cyber activity logs, meteorology [1], etc.

In machine learning and neural networks systems, labelled training dataset are important to train the models. Under ideal cases, balanced training data are expected to be associated with each class, such that the models can perform well equally for all classes. In some circumstances, including some real-world applications of transport traffic data, number of available data for each class are largely unbalanced [4], [5]. This situation brings challenges on good performances to have the trained models on corner cases.

[1]Xinggan Peng, Yuxuan Lin, Huiping Zhuang and Zhiping Lin are with School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore {xinggan001; liny0108; huiping001}@e.ntu.edu.sg and ezplin@ntu.edu.sg.

[2]Qi Cao is with School of Computing Science, University of Glasgow, Singapore qi.cao@glasgow.ac.uk.

[3]Yigang Cen is with School of Computer Science Information Technology, Beijing Jiaotong University, Beijing, China ygcen@bjtu.edu.cn.

Research works on anomaly detection have been reported and studied in various applications to find out patterns not consistent with predicted behaviors [6], [7]. In time series analysis, a lot of useful information is able to be extracted from dataset. Unusual events or interesting insights may be obtained through time series analysis and anomaly detection.

Under the Smart City initiatives in various countries, intelligent transport systems (ITS) with different types of sensors or cameras have been deployed to monitor the transport traffic on roads. Relevant traffic flow data have been recorded for remote monitoring or analytics purpose. Such transport traffic flow data collected from various sensing sources in urban areas is a type of time series data [8]. The traffic flow time series data analysis is very important that could be classified into two major categories: traffic predictions and traffic anomaly detections. Traffic predictions mainly provide forecasting on future traffic information based on the historical and real-time traffic flow time series data, with several examples of prior research works being reported in [9]–[11]. While traffic anomaly detections mainly sense and flag out un-usual traffic patterns compared to past traffic flow under similar circumstances. Relevant examples of prior research works on traffic anomaly detections are introduced in [12]–[15].

The traffic anomalies may not happen frequently but may result in potential risks or consequences on urban traffic. The insights from the traffic anomaly detection in time series data are very useful to relevant agencies or organizations to make necessary arrangements or take actions to provide solutions for these anomalies. It could potentially reduce the traffic congestions at certain urban regions in certain time frames.

In general, there are three types of anomaly detection approaches: supervised learning with training data been labelled as normal or anomaly; semi-supervised learning with only normal class in training data being labelled; and unsupervised learning without labelled training data. With the increasing amount of time series data being collected over the years from different cities or countries, it becomes challenging to manage big dataset and identify anomalies. Machine learning approaches have been employed in identifying meaningful insight and traffic anomaly detection tasks.

An incident detection algorithm based on Convolutional Neural Networks (CNN) is introduced to detect anomalous conditions in traffic simulation environment [16]. Its experiments are conducted on data generated by a city model in Simulation of Urban MObility (SUMO), with the comparisons between simulated traffic flow values and historical traffic patterns. A traffic anomaly detection method using Siamese neural network that contains the twin neural networks with the same weights in presented [15]. This model is used to measure the similarity and outputs of two vectors fed into the twin networks are compared. The anomaly scores are computed by a pre-trained machine learning model in vehicular traffic simulations of SUMO environments. A traffic anomaly detection method consisting of K-mean clustering component is introduced for univariate time-series analysis, which is evaluated on real traffic time series data collected through loop detectors locating at an urban road network in Europe [13]. Detecting road anomalies is formulated as a classification problem representing three sets of features for road conditions [17], where three deep learning approaches including Deep Feedforward Network, CNN, and Recurrent Neural Network (RNN) are evaluated and compared. A recurrent neural networks Long Short-Term Memory (LSTM) is introduced to detect traffic anomaly by identifying decrease or an excessive increase of vehicles flow on the highways compared to normal values [18]. Its model is evaluated using dataset generated by a multi-agent system representing the different time series data. Zhou et al. [19] present a long sequences prediction method named Informer to solve the long-sequence time-series forecasting (LSTF) problem, improving inference speed of long-sequence predictions observed from experiment results on four testing dataset.

Hybrid machine learning methods are also popular research topics in the traffic flow anomaly detection. A deep anomaly detection algorithm combining LSTM deep learning model and extreme value theory is reported that is able to provide recommendations to the ITS for better road traffic management [20]. The performance comparisons of CNN, LSTM and hybrid CNN- LSTM models for traffic congestion anomaly detection and prediction have been conducted by applying on 36.34 million data points collected on motorways in Sydney Australia [21]. It is observed from the experiment results that hybrid CNN-LSTM model underperformed than the LSTM model and applying anomaly detection to remove outlier data can significantly enhance traffic prediction results.

Due to the development based on Vanilla Transformer [22], Transformer models have shown better performance in capturing long-range dependency than RNN models in many applications [19]. However, the training of Transformer models heavily relies on the use on GPU [23], which pose a limit for applying such models in practical use. Recently, a novel improvement of Transformer model named Informer [19] has been introduced, which significantly reduce time and space complexity with outstanding prediction performances

In this work, we aim at improving on the existing machine learning solutions to identify traffic anomalies to uncover any meaningful insights of traffic behaviour patterns. By leveraging the Informer method for long sequence time-series forecasting, an anomaly detection algorithm for traffic time series data is proposed using the encoder-decoder architecture based anomaly score generator. Feature maps with different scales are extracted in the encoder structure.

TABLE I
DESCRIPTION OF THE COLLECTED TRAFFIC DATASET.

Dataset	# Features	# Train	# Validation	# Test
Vehicle Flow	5	4938	486	1162
Pedestrian Flow	4	4234	518	1386

Computations with two multi-head attention layers and a fully connected layer are conducted in the decoder structure. Anomaly scores of the target sequence at time t can be derived between predicted outputs and true values accordingly. The proposed machine learning approach has been evaluated on detecting traffic anomalies for long-sequence predictions, based on the real-world traffic dataset been recorded in Guiyang City, China in 2021. The dataset consists of traffic flows of different types of vehicles and pedestrians with one hour time intervals. Moreover, the proposed anomaly detection algorithm has been evaluated on a public dataset named the Skoltech anomaly benchmark (SKAB) [24].

The main contributions of this research are as follows:

1. By improving the long sequences prediction method, Informer, an anomaly detection algorithm with an anomaly score generator is proposed to solve anomaly detection problem on both vehicles and pedestrian flow dataset.

2. The proposed anomaly detection algorithm has been evaluated by the real-world time series dataset containing various types of vehicle traffic flows and pedestrian flows with one hour sampling rate.

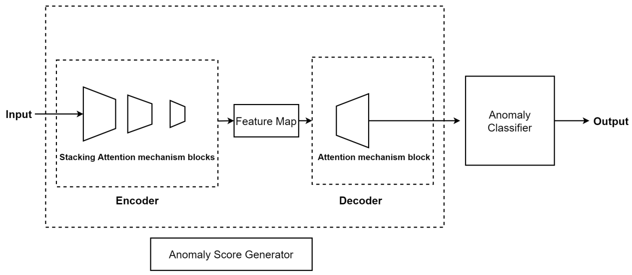
3. The proposed anomaly detection algorithm has been evaluated on the public dataset SKAB. It achieved competitive results compared with other baseline methods.

The organization of the remaining parts of this paper is shown next. Section II describes our proposed methodology and algorithm. Section III presents the experiments and evaluations on the real-world traffic dataset and the public dataset SKAB. Section IV concludes this paper.

II. Methodology and Proposed Algorithm

As shown in Fig. 1, there are two primary components of the proposed algorithm: an anomaly score generator and an anomaly classifier. We adopt the structure of Informer to be the first component of our method to generate anomaly scores of the samples. The One-Class Support Vector Machines (OCSVM) is applied to be the classifier. The anomaly score generator applies the encoder-decoder architecture, aiming to encode input sequences $\mathcal{X}^t$ into the encoded representation $\mathcal{H}^t$. Next, the decoder reconstructs this representation to generate the output $\mathcal{Y}^t$ from $\mathcal{H}^t$.

The encoder consists of three stacking *ProbSparse* self-attention mechanism [19] based blocks to generate feature maps with different scales. Comparing with canonical self-attention mechanism [22], which performs the scaled dot-product on tuple inputs (i.e., query ($\mathbf{Q}$), key ($\mathbf{K}$) and value ($\mathbf{V}$)), the *ProbSparse* [19] self-attention mechanism allows each $\mathbf{K}$ to only attend to the top-m dominant of $\mathbf{Q}$. It significantly reduces the computing complexity as in Eq. (1).



(a)

Fig. 1. Network block diagram of the proposed algorithm.

$$\mathcal{A}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{Softmax}(\frac{\overline{\mathbf{Q}}\mathbf{K}^{\mathbf{T}}}{\sqrt{d}})\mathbf{V} \tag{1}$$

where $\overline{\mathbf{Q}}$ is the sparse matrix that only contains top-m dominant of $\mathbf{Q}$ under the query sparsity measurement [19] and d is the dimension of the input.

Self-attention distilling operations [19] are performed in the layers of the encoder to reduce the size of the network. Next, feature maps generated from three stacked blocks are concatenated as the output of the encoder.

The decoder applies the standard structure mentioned in [22] with two multi-head attention layers and a fully connected layer to obtain an output of the decoder.

Based on the network of the anomaly score generator, at time t an input sequence with length L_x can be represented as Eq. (2).

$$\mathcal{X}^t = \{x_1^t, ..., x_{L_x}^t \quad | \quad x_i^t \in \mathbb{R}^{d_x}\} \tag{2}$$

where x_i^t is the observed inputs at the i^{th} time stamp and d_x is the dimension of the input.

The network will predict the corresponding target sequence with length L_y shown in Eq. (3).

$$\hat{\mathcal{Y}}^t = \{\hat{y}_{L_x+\tau}^t, ..., \hat{y}_{L_x+\tau+L_y}^t \quad | \quad \hat{y_i}^t \in \mathbb{R}^{d_y}\} \tag{3}$$

where $\hat{y}_i^t$ is the predicted value of the target sequence at the i^{th} time stamp; τ is the time delay and d_y is the dimension of the target sequence.

Next, the anomaly score of the target sequence at time t can be computed as the Mean Squared Error (MSE) between the predicted output $\hat{\mathcal{Y}}^t$ and the true value $\mathcal{Y}^t$ as:

$$Score_t = MSE(\hat{\mathcal{Y}}^t, \mathcal{Y}^t) \tag{4}$$

Hence, it is straightforward to determine that the target sequence $\mathcal{Y}^t$ is anomalous if the anomaly score of $Score_t$ is larger than a defined threshold. But such operations take much time to find the optimal value of the threshold. In addition, some prior works also apply the encoder-decoder scheme to detect time-series anomalies where the value of the threshold is determined by assuming the distribution of the anomaly score follows multivariate Gaussian distribution [25], [26]. However, this assumption may not be possible in many practical applications. Therefore, we aim to overcome such disadvantages by an anomaly classifier that does not require any assumption of data. Since the OCSVM has been widely applied to different anomaly detection tasks, it is adopted as the anomaly classifier in this work. Specifically, the anomaly score of $Score_t$ is sent to a trained OCSVM based on the training data to detect anomalies automatically.

TABLE II

PERFORMANCE COMPARISONS OF THE PROPOSED ALGORITHM WITH OTHER FIVE METHODS ON THE VEHICLE FLOWS TASK.

Method	Precision (%)	Recall (%)	F_1 (%)
LOF [27]	38.98	44.23	41.44
COPOD [28]	**81.08**	57.69	67.42
OCSVM [29]	69.64	75.00	72.22
ECOD [30]	70.18	76.92	73.39
Informer [19]	76.92	76.92	76.92
Ours	70.59	**92.31**	**80.00**

TABLE III

PERFORMANCE COMPARISONS OF THE PROPOSED ALGORITHM WITH OTHER FIVE METHODS ON THE PEDESTRIAN FLOWS TASK.

Method	Precision (%)	Recall (%)	F_1 (%)
LOF [27]	18.25	71.88	29.11
COPOD [28]	**88.24**	46.88	61.22
OCSVM [29]	52.83	**87.50**	65.88
ECOD [30]	84.21	50.00	62.75
Informer [19]	73.33	68.75	70.97
Ours	77.78	65.63	**71.19**

III. EXPERIMENTS

The effectiveness of the proposed anomaly detection algorithm is evaluated on a real-world traffic dataset including different types of vehicle flows and pedestrian flows is also constructed to evaluate the performances of the proposed algorithm to detect anomalous vehicle pedestrian flows. The time series dataset is collected in Guiyang City, China, with an hourly sampling rate from March 2021 to December 2021. The time series dataset is constructed to record the number of different types of vehicles and number of pedestrians per hour. The dataset contains one class for pedestrians and three classes for vehicles: large vehicles with multiple axles including multi-wheel heavy trucks; medium vehicles with two axles including buses, pickup trucks, and lorries, etc.; small vehicles including sedans and SUV cars. The time series dataset exhibits largely unbalanced number of data, as the number of small vehicles are about 10 – 20 times higher than those of the medium vehicles. While the number of medium vehicles in the dataset are about 5 – 10 times higher than those of the large vehicles.

With the constructed time series dataset, there are two subsets for the anomaly detection tasks: one to detect anomalous vehicle flows, the other for anomalous pedestrian flows detection. We split each subset into train, validation and test folds. The detailed descriptions of the collected traffic dataset are listed in TABLE I.

We consider three widely-used metrics for evaluating the effectiveness of the proposed anomaly detection algorithm,

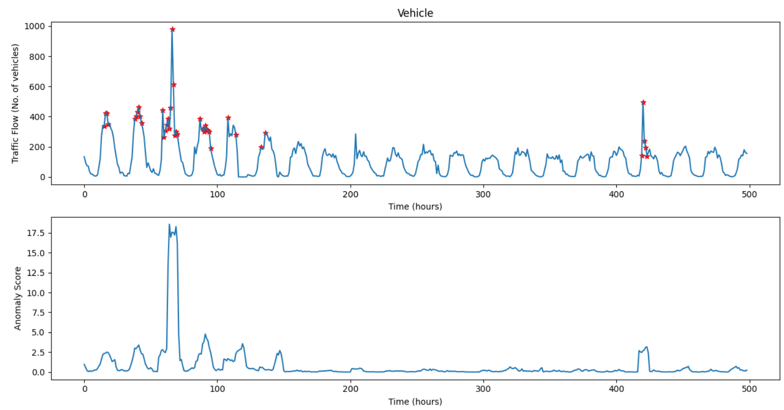


Fig. 2. Example of vehicle flow data with point anomaly and continuous group anomaly (top), and the anomaly scores generated by the proposed algorithm (bottom). Anomalous data are in red.

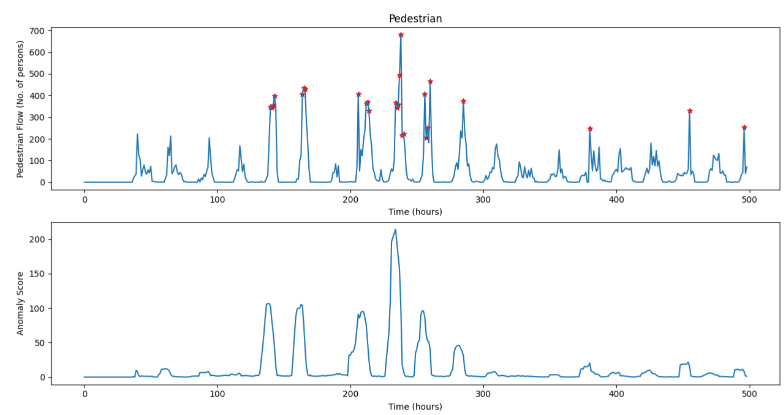


Fig. 3. Example of passenger flow data with point anomaly and continuous group anomaly (top), and the anomaly scores generated by the proposed algorithm (bottom). Anomalous data are in red.

including the *Precision*, *Recall* and F_1 *score*:

$$\begin{aligned} Precision &= \frac{TP}{FP+TP}, \\ Recall &= \frac{TP}{FN+TP}, \\ F_1 &= \frac{2 \times Precision \times Recall}{Precision + Recall}, \end{aligned} \tag{5}$$

where TP denotes the number of correct predictions; FP and FN denote the number of false positive and false negative predictions, respectively. Better performances are indicated by higher values in the F_1 score, precision and recall.

All experiments are run on a PC with an Intel Core$^{\mathrm{TM}}$ i7-8750H 2.20 GHz processor and 16 GB RAM. Other comparison methods are the Informer [19] and other four commonly used methods implemented using frameworks provided by [31], including Local Outlier Factor (LOF) [27], Copula-Based Outlier Detection (COPOD) [28], OCSVM [29] and Unsupervised Outlier Detection using Empirical Cumulative Distribution Functions (ECOD) [30] . The values of the hyper-parameters in the experiments are set as follows: time delay $\tau = 1$ hour, input length $L_x = 32$ hours and target length $L_y = 8$ hours for the traffic dataset.

Fig. 2 and Fig. 3 present examples of two types of anomalies (i.e., the point anomalies and continuous group anomalies) in the vehicle and pedestrian flow dataset. It is observed that the anomaly scores generated by the proposed algorithm are stable for normal samples. But values are relatively high for both samples belonging to the point anomalies and continuous group anomalies. The results illustrate that the proposed algorithm is able to capture characteristics of two types of anomalies.

TABLE II and TABLE III present the comparison results among the proposed detection algorithm and other methods on the vehicle flow and pedestrian flow dataset, respectively. It is observed from TABLE II, the proposed detection algorithm obtains the highest performance in both results of *Recall* and F_1 *score* (i.e., 92.31% and 80.00%, respectively). It illustrates the developed detection algorithm is able to recognize most of the anomalies in the dataset and has better overall detection performances compared with other widely used anomaly detection algorithms. Similarly, in the anomaly detection task on pedestrian flows shown in Table III, the proposed algorithm has the best performance in the result of F_1 *score* (71.19%). These experiment results show that the proposed algorithm has better overall detection performance and makes good balances on false alarms and missing alarms compared with other five types of methods. Although the OCSVM method obtains the highest *Recall* value at 87.50%, the results of *Precision* and F_1 *score* are much lower, which means it generates a lot of false alarms. Additionally, performances of the five comparing methods achieved in the pedestrian flow task are much lower than those obtained in the vehicle flow task. This performance drop is because the change of pedestrian flows is much faster and more unpredictable. On the contrary, the performances of the proposed detection algorithm are stable in both vehicle flow and pedestrian flow tasks. It demonstrates that the proposed algorithm is more robust in dealing with fast-changing time-series samples and making it more suitable for real-world applications. Comparing results obtained in both vehicle flow and pedestrian flow tasks, the proposed algorithm, which combines the Informer and OCSVM, exhibits a higher F_1 *score* than individually using Informer or OCSVM. This performance improvement shows that the proposed algorithm uses Informer and OCSVM to achieve better anomaly detection performances.

In addition, a public dataset named the SKAB [24], which has 34 subsets of multivariate time-series data is also used to evaluate the proposed anomaly detection algorithm. By clicking the link[1], ten commonly used anomaly detection methods are listed as the baselines for comparison, including the multivariate state estimation technique (MSET) [32], Hotelling's T-squared statistic + Q statistic (SPE index) based on PCA (T-squared + Q (PCA)) [33], LSTM [34], multi-scale convolutional recurrent encoder-decoder (MSCRED) [35], Hotelling's T-squared statistic (T-squared) [36], Isolation Forest [37], feed-forward neural network with autoencoder (Autoencoder) [38] and three variants of it. We adopt the performance of these reported methods as the benchmark for comparison.

The values of the hyper-parameters in the experiments of for the SKAB dataset are set as follows: time delay $\tau = 1$ second, input length $L_x = 32$ seconds and target length $L_y = 8$ seconds. False alarm rate (FAR), missing alarm rate (MAR) and F_1 score (in two decimal places) are posted as evaluation metrics in the SKAB dataset website, so we keep this setting:

$$\begin{aligned} FAR &= \frac{FP}{FP+TN}, \\ MAR &= \frac{FN}{FN+TP}, \end{aligned} \tag{6}$$

where TN denotes the number of correct predictions on normal samples. TP, FP, FN and F_1 score are referring to definitions in Eq. (5). Better performances are indicated by lower values in the FAR and MAR.

[1]https://github.com/waico/SKAB

TABLE IV

PERFORMANCE COMPARISONS OF THE PROPOSED METHOD WITH OTHER TEN METHODS ON THE SKAB DATASET.

Method	F_1	FAR (%)	MAR (%)
Ours	**0.82**	13.89	**14.01**
Conv-AE [24]	0.79	13.69	17.77
MSET [32]	0.73	20.82	20.08
LSTM-AE [24]	0.68	14.24	35.56
T-squared + Q (PCA) [33]	0.67	13.95	36.32
LSTM [34]	0.64	15.4	39.93
MSCRED [35]	0.64	13.56	41.16
LSTM-VAE [24]	0.56	9.13	55.03
T-squared [36]	0.56	12.14	52.56
Autoencoder [38]	0.45	7.56	66.57
Isolation Forest [37]	0.40	**6.86**	72.09

Table IV presents the results of the proposed algorithm compared with others. It can be observed that the proposed algorithm achieves the best performances in both F_1 and MAR (i.e., 0.82 and 14.01%, respectively). Although Isolation Forest achieves the lowest FAR with 6.86%, its MAR is too large (72.09%). This phenomenon shows that such a method can not recognize most of the anomalies. Compared with these methods, the proposed algorithm can sufficiently balance false alarms and miss alarms, leading to better overall detection performances.

In summary, the algorithm can correctly tackle two kinds of anomaly detection tasks on time series traffic flow. This is quite promising since many existing methods only focus on vehicle flow detection instead of detecting both vehicle and pedestrian flows. In addition, the proposed algorithm achieves competitive results on the public SKAB dataset.

IV. CONCLUSIONS

In this paper, the anomaly detection algorithm for time series vehicle and pedestrian flows is proposed. In the conducted experiments, the proposed algorithm can detect two kinds of anomalous flows with high accuracy. The analysis also shows that the proposed algorithm obtains competitive results compared with other widely used anomaly detection algorithms. Moreover, the prosed anomaly detection algorithm has been evaluated on the public SKAB dataset. Moving forward, we will extend the proposed approach to the more general cases of time series anomaly detections. Our discovery could motivate further research to improve time series traffic anomaly detection performances. In the future works, we could also exploit our anomaly detection algorithm on data collected from different urban areas, cities or countries.

ACKNOWLEDGMENT

This work was supported in part by the National Key R&D Program of China under Grant 2021YFE0110500; in part by the National Natural Science Foundation of China under Grant 61872034, Grant 62011530042, and Grant 62062021; in part by the Beijing Municipal Natural Science Foundation under Grant 4202055. The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg).

REFERENCES

[1] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: a systematic review," *IEEE Access*, 2021.

[2] J. Monteil, A. Dekusar, C. Gambella, Y. Lassoued, and M. Mevissen, "On model selection for scalable time series forecasting in transport networks," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[3] C. Ma, G. Dai, and J. Zhou, "Short-term traffic flow prediction for urban road sections based on time series analysis and lstm_bilstm method," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[4] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 9268–9277.

[5] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit, and S. Kumar, "Long-tail learning via logit adjustment," *International Conference on Learning Representations*, 2021.

[6] X. Wang, A. Fagette, P. Sartelet, and L. Sun, "A probabilistic tensor factorization approach to detect anomalies in spatiotemporal traffic activities," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 1658–1663.

[7] Z. Liu, Z. Zhu, J. Gao, and C. Xu, "Forecast methods for time series data: a survey," *IEEE Access*, vol. 9, pp. 91 896–91 912, 2021.

[8] S. Zhao, S. Lin, Y. Li, J. Xu, and Y. Wang, "Urban traffic flow forecasting based on memory time-series network," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–6.

[9] H. Yuan and G. Li, "A survey of traffic prediction: from spatio-temporal data to intelligent transportation," *Data Science and Engineering*, vol. 6, no. 1, pp. 63–85, 2021.

[10] J. Zheng and M. Huang, "Traffic flow forecast through time series analysis based on deep learning," *IEEE Access*, vol. 8, pp. 82 562–82 570, 2020.

[11] A. Avila and I. Mezić, "Data-driven analysis and forecasting of highway traffic dynamics," *Nature communications*, vol. 11, no. 1, pp. 1–16, 2020.

[12] M. Bawaneh and V. Simon, "Anomaly detection in smart city traffic based on time series analysis," in *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, 2019, pp. 1–6.

[13] M. R. Alam, I. Gerostathopoulos, S. Amini, C. Prehofer, and A. Attanasi, "Adaptable anomaly detection in traffic flow time series," in *2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, 2019, pp. 1–9.

[14] K. Kalair and C. Connaughton, "Anomaly detection and classification in traffic flow data from fluctuations in the flow–density relationship," *Transportation Research Part C: Emerging Technologies*, vol. 127, p. 103178, 2021.

[15] S. Sabour, S. Rao, and M. Ghaderi, "Deepflow: Abnormal traffic flow detection using siamese networks," in *2021 IEEE International Smart Cities Conference (ISC2)*. IEEE, 2021, pp. 1–7.

[16] L. Zhu, R. Krishnan, A. Sivakumar, F. Guo, and J. W. Polak, "Traffic monitoring and anomaly detection based on simulation of luxembourg road network," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 382–387.

[17] D. Luo, J. Lu, and G. Guo, "Road anomaly detection through deep learning approaches," *IEEE Access*, vol. 8, pp. 117 390–117 404, 2020.

[18] S. M. Snineh, N. E. A. Amrani, M. Youssfi, O. Bouattane, and A. Daaif, "Detection of traffic anomaly in highways by using recurrent neural network," in *2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS)*. IEEE, 2021, pp. 1–6.

[19] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of AAAI*, 2021.

[20] N. Davis, G. Raina, and K. Jagannathan, "A framework for end-to-end deep learning-based anomaly detection in transportation networks," *Transportation research interdisciplinary perspectives*, vol. 5, p. 100112, 2020.

[21] A.-S. Mihaita, H. Li, and M.-A. Rizoiu, "Traffic congestion anomaly detection and prediction using deep learning," *arXiv preprint arXiv:2006.13215*, 2020.

[22] 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.

[23] 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.

[24] I. D. Katser and V. O. Kozitsin, "Skoltech anomaly benchmark (skab)," https://www.kaggle.com/dsv/1693952, 2020.

[25] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "Lstm-based encoder-decoder for multi-sensor anomaly detection," *arXiv preprint arXiv:1607.00148*, 2016.

[26] P. Malhotra, L. Vig, G. Shroff, P. Agarwal *et al.*, "Long short term memory networks for anomaly detection in time series," in *Proceedings*, vol. 89, 2015, pp. 89–94.

[27] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.

[28] Z. Li, Y. Zhao, N. Botta, C. Ionescu, and X. Hu, "Copod: copula-based outlier detection," in *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2020, pp. 1118–1123.

[29] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[30] Z. Li, Y. Zhao, X. Hu, N. Botta, C. Ionescu, and G. H. Chen, "Ecod: Unsupervised outlier detection using empirical cumulative distribution functions," *arXiv preprint arXiv:2201.00382*, 2022.

[31] Y. Zhao, Z. Nasrullah, and Z. Li, "Pyod: A python toolbox for scalable outlier detection," *Journal of Machine Learning Research*, vol. 20, no. 96, pp. 1–7, 2019. [Online]. Available: http://jmlr.org/papers/v20/19-011.html

[32] N. Zavaljevski and K. C. Gross, "Sensor fault detection in nuclear power plants using multivariate state estimation technique and support vector machines." Argonne National Lab., Argonne, IL (US), Tech. Rep., 2000.

[33] S. Joe Qin, "Statistical process monitoring: basics and beyond," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 17, no. 8-9, pp. 480–502, 2003.

[34] P. Filonov, A. Lavrentyev, and A. Vorontsov, "Multivariate industrial time series with cyber-attack simulation: Fault detection using an lstm-based predictive data model," *arXiv preprint arXiv:1612.06676*, 2016.

[35] C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, and N. V. Chawla, "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 1409–1416.

[36] H. Hotelling, "Multivariate quality control-illustrated by the air testing of sample bombsights," 1947.

[37] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, pp. 1–39, 2012.

[38] J. Chen, S. Sathe, C. Aggarwal, and D. Turaga, "Outlier detection with autoencoder ensembles," in *Proceedings of the SIAM international conference on data mining*, 2017, pp. 90–98.