

Received 19 July 2022, accepted 14 August 2022, date of publication 19 August 2022, date of current version 26 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3200472

RESEARCH ARTICLE

CRFormer: Complementary Reliability Perspective Transformer for Automotive Components Reliability Prediction Based on Claim Data

HYUN JOON PARK¹, TAEHYEONG KIM¹, YOUNG SEOK KIM¹, JINHONG MIN¹,
KI WOO SUNG², AND SUNG WON HAN¹

¹School of Industrial and Management Engineering, Korea University, Seongbuk-gu, Seoul 02841, Republic of Korea

²Automotive Research and Development Division, Hyundai Motor Group, Advanced Durability Development Team, Namyang 18280, South Korea

Corresponding author: Sung Won Han (swhan@korea.ac.kr)

This work was supported in part by the Korea Techno Complex Foundation under Grant R2112651 and Grant R2112652, in part by Korea University under Grant K2107521 and Grant K2202151, and in part by the Brain Korea 21 FOUR.

ABSTRACT Reliability prediction has been studied in many industries for managing stocks and reducing quality assurance costs and production costs. Particularly, in the automotive industry, reliability prediction is performed based on two automobile reliability perspectives, time and mileage. To maximize cost savings, researchers attempted reliability prediction with short-term inputs. However, limited information on short-term inputs resulted in unsatisfactory prediction results for the long warranty periods. Additionally, the overall evaluation metrics could not reflect the pattern-wise performance, such as the increasing failure patterns. This study proposes Complementary Reliability perspective Transformer (CRFormer) based on Transformer encoder to achieve enriched representations from a short-term input sequence. CRFormer fuses different automobile reliability perspective information and automobile features to compensate for the limited information on short-term input. The performance of CRFormer is evaluated based on automobile claim data accumulated over 16 years. Results showed that compared to previous methods in terms of overall, pattern-wise, and pattern similarity evaluation metrics, CRFormer achieved outstanding performance in time and mileage reliability prediction. Lastly, visualization results and survival analysis based on accurate model prediction can be used to support decision-making to reduce quality assurance costs and production costs.

INDEX TERMS Attention mechanism, automobile, reliability prediction, transformer.

I. INTRODUCTION

Automobiles have become the most popular mode of transport, and hence, automobile companies have been developing diverse strategies to provide competitive services. Among these services, quality assurance is a strategy to attract and retain customers. During the assurance period, customers can receive a repair or replacement from the company for a minimal charge [1]. Furthermore, to consolidate the automobile market dominance, automobile companies provide quality assurance service in two aspects, time and mileage, and

The associate editor coordinating the review of this manuscript and approving it for publication was Sajid Ali¹.

extend the assurance period. Although the service satisfies customers, automobile companies spend billions of dollars annually on the service [2]. Since the quality assurance service is significantly related to the reliability of the product [3], the use of reliability prediction has steadily increased to reduce the cost.

Reliability prediction has been studied in various industries to manage stocks, determine appropriate warranty periods, and reduce production costs [4], [5]. Also in the automotive industry, several researchers have utilized reliability prediction after sufficient data related to automobile failure accumulated [6], [7]. Among these data, the claim data is commonly used to predict automobile reliability. Researchers

use the claim sequence, which is failure counts per 1 month or 1,000 miles from claim data comprising information such as the date of failure, automotive code, components, and mileage. By predicting future failures or claims of automobiles based on historical claim sequences, researchers can estimate automobile reliability. However, the limited information in claim sequences, especially when targeting early prediction, makes it challenging to achieve accurate automobile reliability prediction results [8].

In the early days, researchers adopted parametric methods such as the Weibull distribution [9] and Lognormal distribution [10] for reliability prediction in the automotive industry using claim data [11], [12]. However, parametric methods require strict assumptions [13]. To address the limitations of parametric methods, [13] suggested the application of ARIMA model in reliability prediction. Although ARIMA model can capture sequential patterns, it results in unsatisfactory prediction results when predicting long-period failures with short-term inputs [14]. Recently, [14], [15] applied deep learning models to predict automobile reliability. Lee *et al.* [14] and Meng *et al.* [15] used 1-D Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), respectively, to represent time-series claim data. Lee *et al.* [14] showed that RNN outperformed the previous methods (i.e., parametric, time-series, and machine learning models) for remaining warranty period failure prediction with short-term inputs. However, problems such as limited representations made by short-term inputs still persist, due to which the model cannot predict increasing failure patterns. In addition, evaluation metrics that cannot reflect failure patterns, such as the failure-increasing pattern, provide only limited results, which are insufficient to analyze results in various aspects.

In short, the previous methods had the limitation of representing sequential information with short-term inputs, resulting in inaccurate prediction results. In addition, the evaluation metrics could not reflect the results depending on claim patterns. In this study, Complementary Reliability perspective Transformer (CRFormer) is proposed for predicting the time or mileage reliability of automotive components using claim data. By exploiting two reliability perspectives and claim feature information (i.e., car code, system, and components), CRFormer can find long-term failure patterns even with short-term inputs. CRFormer was evaluated on the automotive component claim data accumulated over 16 years. To compensate for the limitation of the previous evaluation metrics, this study provides both overall, pattern-wise, and pattern similarity evaluation results, thereby making it feasible for analyzing the prediction results depending on failure patterns. Experimental results show that CRFormer achieves outstanding performance in terms of both time and mileage reliability predictions.

II. RELATED WORK

This section addresses previous methods of reliability prediction in the automotive industry and their limitations.

Considering reliability prediction is usually regarded as a time-series problem to predict future failures or claims, deep learning models for time-series prediction are also investigated.

A. RELIABILITY PREDICTION IN THE AUTOMOTIVE INDUSTRY

Research on reliability prediction of automobiles has traditionally been conducted using parametric methods. Singpurwalla and Wilson [12] applied the bi-variate warranty forecasting method, which blends the perspectives of time and mileage. As statistical methods, linear models [16], Poisson distribution [11], and Weibull distribution [17] were proposed, but these methods had difficulty representing long remaining failure sequences with short initial failure sequences. In addition, reliability prediction based on distributions required a strict assumption [13] and showed unsatisfactory prediction results for new car models [18]. Although time-series models such as ARIMA and exponential smoothing alleviate the strict assumptions and capture sequential information of the failure sequence, the long-term forecasting performance is still disappointing [6], [19]. Meanwhile, several studies have been conducted based on machine learning using field claim data to predict the reliability of parts [20]. Support Vector Machines (SVM) [21] and gradient boosting models [22] showed better performance compared to statistical models. Owing to the expandability of deep learning models into various fields, it was possible to confirm the effects of deep learning models in reliability prediction [7]. Recently, [14], [15] suggested the 1-D CNN and RNN-based time-series models for automobile reliability prediction. In particular, [14] proved that deep learning models showed better performance in the reliability prediction of automotive components compared to existing statistical, time-series, and machine learning methods.

Although the previous studies contributed to predicting the reliability of automobile parts based on various methods, their methods were not sufficient for extracting enriched representation from short failure sequences. The limitation has become more critical as the importance of early reliability prediction increases. In addition, the previous methods do not consider the difference of failure patterns in terms of both prediction and evaluation phases. The failure patterns appearing in the claim data are divided into decreasing failure patterns caused by quality issues at the beginning of the warranty period, and increasing failure patterns, such as durability problems that gradually increase failures over time. Although detecting increasing failure patterns is relatively more important than detecting decreasing failure patterns, they appear intermittently and rarely, making them almost unpredictable in previous studies. The success of early reliability prediction means not only naively predicting failures, but also detecting increasing failure patterns. Therefore, it is important to develop a model that can predict intermittent increasing failure patterns for practical reliability analysis, thereby indicating that improving the representation of short

input sequences is an important factor. Furthermore, given that the importance of failure patterns differs, it is necessary to provide evaluation metrics that consider the differences in failure patterns.

B. DEEP LEARNING MODELS FOR TIME-SERIES PREDICTION

Given that data availability and computing power have increased, researchers have adopted deep learning models for time-series prediction owing to the advantage of learning the representation of complex data without statistical assumptions [23]. Recurrent Neural Networks (RNNs) [24], [25], [26], [27] were developed to represent the previous time information at the current time step. Long Short-Term Memory (LSTM) [25] mitigates the gradient vanishing problem of RNN by designing cell states containing long-term information. Gated Recurrent Unit (GRU) [24] proposed an improved state update process with fewer gates than LSTM, and the state update process is as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (1)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (2)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \quad (3)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

where x_t and h_t are the current input and updated output of time step t , and σ is a sigmoid function. Each z_t and r_t denotes an update and reset gate, and a bias for b , respectively. In contrast, CNN is adopted for time-series models since it can determine the relationship between past and present information by applying a filter of a fixed size according to time sequence. Considering the advantages of RNNs and CNN, which can capture sequential information, researchers designed RNNs or CNN-based time-series models [28], [29], [30]. However, the problems of long-term dependence and incorrectly accumulated representation from the past steps still persist as the major problem.

Transformer [31] is an attention mechanism-based model used in machine translation that uses self-attention to learn the relationship between sequential information from each point of view. Transformer has the advantage of considering sequential information without the constraint of the sequence step distance via an attention mechanism. Owing to the added advantage of an attention mechanism that alleviates the vanishing gradient problem, the sequential feature extraction ability of Transformer helps achieve state-of-the-art performance in diverse domains such as pre-trained language models [32], image-relevant tasks [33], [34], [35], and a multi-modal task [36]. In the time-series prediction domain, [37], [38], [39] demonstrated that Transformer achieved better performance than the previous time-series models. Park et al. [39] suggested that Transformer models achieved higher performance than LSTM models in vessel fuel consumption prediction and [37], [38] proved the outstanding of Transformer by comparing their methods with both statistical methods and RNNs in

influenza and multi-horizon time-series prediction respectively. As previous studies have proven the success of Transformer in time-series prediction, researchers have proposed Transformer-based time-series models that satisfy both efficiency and performance [40], [41], [42], [43]. In particular, Informer [42] greatly reduced the amount of computation and improved the model performance by using the ProbSparse self-attention mechanism, self-attention distilling operation, and generative style decoder. Informer's ProbSparse attention utilizes Kullback-Leibler divergence to measure the importance of a query and calculates the dot product by sampling significant Top- u queries. Autoformer [43] reflected the time-series characteristics to the end-to-end model through the auto-correlation mechanism and series decomposition block and increased the computational efficiency. A series decomposition block divides time-series inputs into trend-cyclical and seasonal parts using the average pooling technique. Auto-correlation mechanism calculates auto-correlation by inverse fast Fourier transform after dot product using fast Fourier transform on query and key of self-attention mechanism. Then, the Top- k autocorrelations are used as attention weights. Although Informer and Autoformer show state-of-the-art performance in long-term sequence prediction, they are not optimized for reliability prediction which needs to predict long-term sequences with short failure sequences. Furthermore, the time-series characteristics of a short failure sequence do not last on a long-term sequence. Owing to differences in the domain and data, this study proposes a model to represent enriched representations from short-term inputs while taking advantage of Transformer.

III. METHOD

In this section, Complementary Reliability perspective Transformer (CRFormer) is described in detail. As shown in Figure 1, CRFormer consists of a nested sequence embedding module, Transformer encoder, Context Attention (CA), and a prediction layer. CRFormer is designed to predict the reliability of automotive components within warranty periods of *5-year* and *60000-mileage*, respectively. From the claim data, a nested sequence that contains both time and mileage-based information is generated and claim features are extracted. The nested sequence embedding module extracts sequential representations of the target reliability, time or mileage, from the nested sequence. Transformer encoder emphasizes the target sequential information. Finally, CA and a prediction layer merge claim features and each sequence step to predict future claims within the remaining warranty periods.

A. SEQUENCE PROCESSING

As shown in the input data processing in Figure 1, the claim frequencies and features are extracted depending on unique combinations of year, car code, system, and components from the claim data. Based on the claim frequencies, target claim sequences are generated for each target reliability (i.e., time and mileage). The time-based claim sequence contains claim counts per 1 month for 61 months which is the warranty

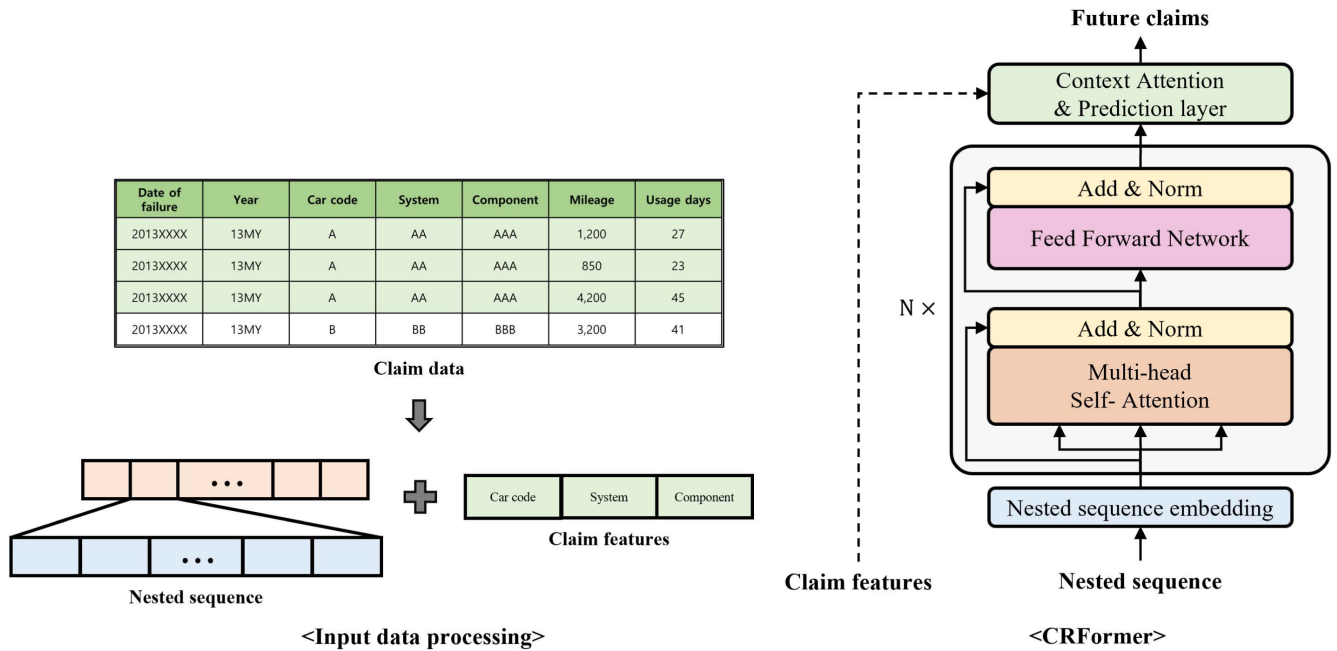


FIGURE 1. Pipeline of input data processing and overall architecture of CRFormer.

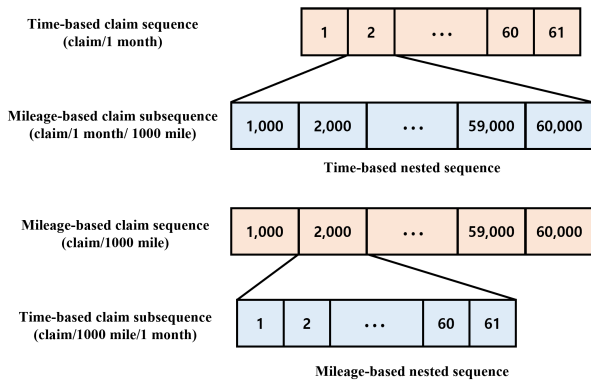


FIGURE 2. Nested sequence depending on target reliability, time and mileage.

period of time, 5-year. On the other hand, the mileage-based claim sequence represents claim counts per 1,000 miles for 60,000 miles which is the warranty period of mileage, 60000-mileage. Since the warranty period affects the claim observation period, it is important to set an appropriate warranty period. Considering the warranty periods of the previous works [11], [14] and the minimum warranty periods of the automobile company which provided the data, the warranty periods in this research are set as 5-year and 60000-mileage, respectively. The target claim sequence x is defined as $x \in \mathbb{R}^{L_r}$, where L_r is the total sequence length of each target reliability within warranty period. $r \in \{t, m\}$ is the set of automobile reliabilities, where t and m are the time and mileage. The target reliability parameter r is adopted to decide what reliability perspective the models predict, therefore the inputs

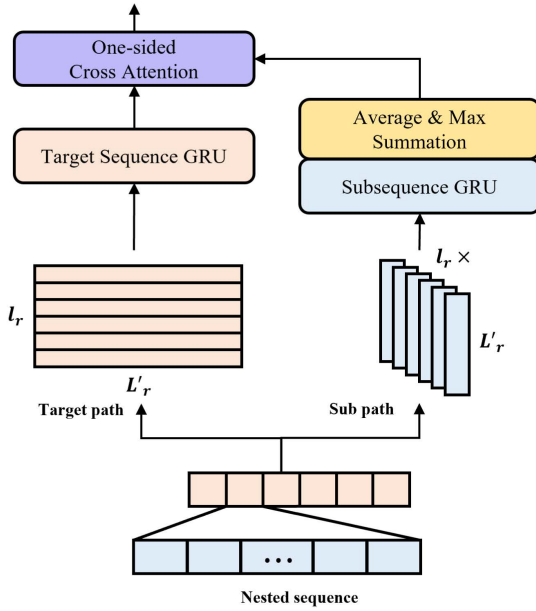
and outputs are defined by the parameter. Finally, the input sequence $x \in \mathbb{R}^{l_r}$ and the output sequence $y \in \mathbb{R}^{L_r - l_r}$ are defined when l_r is the input sequence length.

Since x contains only the claim frequencies for each target reliability, the problem of limited sequence information increases as the input sequence length decreases. To guarantee the enriched information even in a short input sequence, a nested sequence is proposed, containing different reliability information as in Figure 2. The nested sequence represents the claim counts per sub-reliability unit on each target reliability sequence step. Given L'_r is the total sequence length of each sub reliability within the warranty period, the input of the proposed model x is defined as $x \in \mathbb{R}^{l_r \times L'_r}$. For the detailed expressions, the inputs x for time and mileage reliability prediction are defined as $x \in \mathbb{R}^{l_t \times L'_m}$ and $x \in \mathbb{R}^{l_m \times L'_t}$.

B. NESTED SEQUENCE EMBEDDING

This study proposes a nested sequence embedding module that extracts enriched representations of the target reliability claim sequence by fusing with the subsequence made from a different reliability perspective. The nested sequence embedding module is composed of GRU and One-sided Cross Attention (OCA).

To consider the target sequential and subsequential information, the nested sequence $x \in \mathbb{R}^{l_r \times L'_r}$ is forwarded into two paths, as shown in Figure 3. In the target path, the target path input $x_T \in \mathbb{R}^{l_r \times L'_r}$ is defined while maintaining the original input dimension to consider the target sequential information. In contrast, the sub-path input is represented as


FIGURE 3. Nested sequence embedding.

$x_S \in \mathbb{R}^{l_r \times l'_r \times 1}$ to reflect sequential information of the sub-sequence. Furthermore, a GRU layer is adopted to consider each sequential information. Given C is a hidden size, the outputs of each path, $x'_T \in \mathbb{R}^{l_r \times C}$ and $x'_S \in \mathbb{R}^{l_r \times l'_r \times C}$, are defined as $x'_T = GRU(x_T)$ and $x'_S = GRU(x_S)$, respectively. For x'_S , the summation of the average and max pooling on the subsequence aggregates the subsequential information. The output $x'_S \in \mathbb{R}^{l_r \times C}$ is defined as below:

$$x'_S = Average(x'_S) + Max(x'_S) \quad (5)$$

After extracting each sequential information, OCA based on self-attention [31] conveys the aggregated subsequence representations x'_S to the target sequence x'_T . For OCA, the query $Q \in \mathbb{R}^{l_r \times C}$ is represented by linear transformation with the weight $W_q \in \mathbb{R}^{C \times C}$ and x'_T . To convey the subsequence representation, the key and value, K and $V \in \mathbb{R}^{l_r \times C}$, are represented by linear transformations with each weight $W_{k,v} \in \mathbb{R}^{C \times C}$ and x'_S . Given d_k is the scaling size, the one-sided cross attention weight $\alpha_{Cr} \in \mathbb{R}^{l_r \times l_r}$ and the target sequence output $x'_T \in \mathbb{R}^{l_r \times C}$ are defined as follows:

$$\alpha_{Cr} = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (6)$$

$$x'_T = W(\alpha_{Cr}V) \quad (7)$$

where $W \in \mathbb{R}^{C \times C}$ is the weight for linear transformation on $\alpha_{Cr}V$. By considering each reliability perspective sequential information from the nested sequence and fusing them, the target reliability claim sequence contains representations of both automotive reliability perspectives. Since the embedding module represents the nested sequence into the target sequence, the represented target sequence can be forwarded into sequence models such as RNNs and Transformer.

C. ENCODER

CRFormer uses Transformer encoder [31] to emphasize the target claim sequential information represented by the nested sequence embedding module. Encoder is composed of the N number of encoder blocks. Each encoder block contains Multi-head Self-Attention and Feed-Forward Network, each is followed by residual connection and layer normalization.

Multi-head Self-Attention extracts representations by dividing the H number of heads and performing self-attention on each head to consider diverse views. Given $x_T \in \mathbb{R}^{l_r \times C}$ is the target claim sequence, Q_i , K_i , and $V_i \in \mathbb{R}^{l_r \times C/H}$ of each head is represented by a linear transformation with each weight $W_i^{q,k,v} \in \mathbb{R}^{C \times C/H}$ and x_T , where $i = \{1, 2, \dots, H\}$. The output of Multi-head Self-Attention $x'_T \in \mathbb{R}^{l_r \times C}$ is obtained by concatenating the self-attention outputs of each head, given as:

$$head_i = Attention(Q_i, K_i, V_i) = softmax\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \quad (8)$$

$$x'_T = MultiHead(Q, K, V) = W([head_1; \dots, head_H]) \quad (9)$$

$$x'_T = LayerNorm(x_T + x'_T) \quad (10)$$

d_k is defined as $d_k = C/H$ for the scaling size and $W \in \mathbb{R}^{C \times C}$ is used as a linear transformation for the output. Additionally, layer normalization and residual connection stabilize the learning process.

Then, the output of each block is obtained by Feed-Forward Network, residual connection, and layer normalization, which are applied to x'_T . The process is as follows:

$$FFN(x'_T) = W_{out}(ReLU(W_{in}(x'_T))) \quad (11)$$

$$x'_T = LayerNorm(x'_T + FFN(x'_T)) \quad (12)$$

where $W_{in} \in \mathbb{R}^{C \times C}$ and $W_{out} \in \mathbb{R}^{C \times C}$ denote the weights of Feed-Forward Network. The output of the encoder is obtained after passing through the N number of blocks.

D. CONTEXT ATTENTION AND PREDICTION LAYER

Instead of adopting Transformer decoder to predict future claims at each step, CRFormer contains Context Attention (CA) and a prediction layer. CRFormer adopts CA based on [44] and [45] to emphasize and aggregate sequential information using a context vector in the target claim sequence $x_T \in \mathbb{R}^{l_r \times C}$, which is the output of the encoder block. Given $C_{vector} \in \mathbb{R}^{C \times 1}$ is the learnable context vector for capturing important information of each sequence step, the context attention weight $\alpha_c \in \mathbb{R}^{l_r \times 1}$ and the aggregated output $x'_T \in \mathbb{R}^C$ which is the average of the sequence of the attention output are obtained as below:

$$\alpha_c = softmax(W_c(x_T)C_{vector}) \quad (13)$$

$$x'_T = Average(x_T \odot \alpha_c) \quad (14)$$

where $W_c \in \mathbb{R}^{C \times C}$ denotes the weight for linear transformation on x_T .

Regarding feature information of the claim sequence (i.e., car code, system, and components), each categorical feature

$x_{car,sys,part} \in \mathbb{R}^C$ is transformed by the weight $W_{car,sys,part} \in \mathbb{R}^{1 \times C}$. To merge the feature representation $x_{car,sys,part}$ and the aggregated representation x'_T , each $x_{car,sys,part}$ is concatenated, followed by a summation with x'_T and layer normalization. The process is defined as follows, where $W_f \in \mathbb{R}^{3C \times C}$ is used for the dimension reduction to the concatenated feature representation.

$$x_f = W_f([x_{car}; x_{sys}; x_{part}]) \tag{15}$$

$$x'_T = LayerNorm(x'_T + x_f) \tag{16}$$

The above equations are used when the feature information is exploited. Then, a linear prediction layer $W_p \in \mathbb{R}^{C \times L_r - l_r}$ on x'_T is applied to predict future claims $\hat{y} \in \mathbb{R}^{L_r - l_r}$. Since the outputs are arranged depending on the target reliability parameter r , time and mileage reliability prediction outputs are defined as $\hat{y} \in \mathbb{R}^{L_t - l_t}$ and $\hat{y} \in \mathbb{R}^{L_m - l_m}$, respectively.

IV. EXPERIMENT

A. DATASET

This study uses the claim data provided by an automobile company. The claim data were collected for 16 years, comprising 951,170 claims, from 2006 to 2021. The claim data are acquired when the claim or failure is reported, and they contain some information about the claim. The status and meta information of the car in failure are recorded as the claim information. As shown in Table 1, the claim data contain three types of information (i.e., automotive, usage, and sales). Among the automotive information, the system and components indicate the automotive factors which caused the failure. Usage information includes the periods and mileage up to the failure. Sales information refers to the total number of car sales, where the car is distinguished by car code and year. To target the prediction of reliability within the warranty of 5-year and 60000-mileage, the claim data that is over five years from the release date (i.e., between 2006 to 2016) are used. In the selected data, the number of unique cars, systems, and components is 11, 8, and 567, respectively. The failure proportion of 11 car codes is described in Table 2.

To transform the claim data into claim sequences for model inputs, the sequence processing described in Section III-A is applied. After the transformation, the total number of claim sequences from 2006 to 2016 is 4,663, where the year represents the release year of the cars. To train and evaluate

TABLE 1. Features of claim data.

Category	Features	Type	Description
Automotive information	Car code	str	Unique code of car
	Year	int	Release year of car
	System	str	System causing failure
	Component	str	Component causing failure
Usage information	Date of failure	UTC	Date of failure occurred
	Usage days	int	Period up to failure
	Mileage	int	Mileage up to failure
Sales information	Number of sales	int	Total number of car sales

TABLE 2. Failure proportion per car code. The car codes are alphabetically masked.

Car code	Failure proportion	Car code	Failure proportion
A	0.059	G	0.190
B	0.084	H	0.013
C	0.082	I	0.139
D	0.056	J	0.095
E	0.034	K	0.139
F	0.109		

the models, the train, validation, and test dataset are set as 2006 to 2014, 2015, and 2016 year, respectively. Therefore, there are 3,440 claim sequence data from 2006 to 2014 as train data, 676 claim sequence data in 2015 as validation data, and 574 claim sequence data in 2016 as test data.

B. BENCHMARK MODELS

Benchmark models from previous studies are set for comparison with the proposed models. The benchmark models contain the Weibull distribution and ARIMA for the parametric and time-series models, respectively. In particular, their parameters (i.e., shape and scale for Weibull distribution and order of auto-regressive model, differencing, and moving-average model for ARIMA) are determined by validation scores. Additionally, considering that [14] showed the superiority of LSTM in claim prediction compared to the previous methods, RNN, LSTM, and GRU are also adopted. Finally, Transformer is used, which is the base model of the proposed models, and Transformer-based time-series models, namely, Informer and Autoformer, for comparison.

C. IMPLEMENTATION DETAILS

For training, a batch size of 128 and an epoch of 200 are used. In addition, 12 loss and Adam optimizer are adopted to optimize the proposed models. During training, the best model weights are maintained based on the validation score. Regarding model hyperparameters, CRFormer has a hidden dimension size $C = 256$, the number of encoder blocks $N = 6$, and the number of heads in Self-Attention $H = 4$. Given the input sequence length l_r is 6, the experiments for comparing the proposed models with benchmark models are conducted. To verify the complementarity of the proposed models, the target reliability perspective r is set as each time t and mileage m . That is, two experiments show each result of predicting time and mileage reliability. Given the warranty periods of 5-year and 60000-mileage, each period is divided by 1 month and 1,000 miles respectively, thus, each total sequence length L_t and L_m are 61 and 60. In addition, the proposed models are distinguished by CRFormer and CRFormer-F (i.e., CRFormer with features) to verify the influence of the features on the model performance gain. CRFormer-F uses the automotive information features except for year information, and sale information is added as features in the further experiment. Finally, the same level of hyperparameters is used for the benchmark models.

TABLE 3. Comparison results of time reliability prediction in terms of claim prediction and pattern metrics. 0, 1, and 2 in RMSE and MAE denote irregular, decreasing, and increasing patterns, respectively.

Model	RMSE					MAE					Pattern	
	ACC	OVR	0	1	2	ACC	OVR	0	1	2	Accuracy	F1
Weibull	161.23	3.87	1.57	3.22	13.72	62.55	1.41	1.00	2.10	6.08	0.91	0.57
ARIMA	413.06	8.08	1.26	26.42	11.40	85.45	1.86	0.75	10.96	5.64	0.87	0.43
RNN	132.47	4.34	1.65	5.71	12.55	80.72	1.99	1.31	4.61	5.36	0.91	0.61
LSTM	159.67	4.63	1.06	4.17	15.27	53.03	1.17	0.44	2.13	6.69	0.92	0.59
GRU	115.49	4.03	1.22	4.44	12.68	52.43	1.33	0.67	3.46	5.41	0.91	0.60
Transformer	121.93	4.01	1.30	5.79	11.92	55.81	1.38	0.66	4.05	5.24	0.91	0.59
Informer	118.45	3.61	1.11	4.04	11.35	84.00	1.31	0.81	2.57	4.99	0.92	0.61
Autoformer	242.57	6.44	1.73	24.11	11.85	78.64	2.25	1.25	9.29	5.70	0.82	0.35
CRFormer	98.48	3.46	1.18	4.12	10.83	47.89	1.21	0.57	3.26	5.19	0.91	0.61
CRFormer-F	139.01	4.13	1.39	6.13	12.10	53.51	1.17	0.55	3.21	5.44	0.90	0.63

D. EVALUATION METRICS

The proposed and benchmark models are evaluated in terms of claim prediction and pattern similarity. To evaluate future claim predictions, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are adopted. The MAE and RMSE, which calculate the average of errors in each step during the predicted warranty period, are defined as the overall MAE and RMSE. Given l_r is the input sequence length and L_r is the total warranty period, the metrics are calculated as below:

$$MAE_{OVR} = \frac{1}{n(L_r - l_r)} \sum_{i=1}^n \sum_{t=l_r}^{L_r} |y_{i,t} - \hat{y}_{i,t}| \quad (17)$$

$$RMSE_{OVR} = \sqrt{\frac{1}{n(L_r - l_r)} \sum_{i=1}^n \sum_{t=l_r}^{L_r} (y_{i,t} - \hat{y}_{i,t})^2} \quad (18)$$

Above the equations, n and t denote the number of claim sequences and each step in the claim sequences, respectively. Furthermore, accumulated MAE and RMSE are adopted to evaluate the accumulation of predicted claims over warranty periods. The accumulated evaluation metrics are calculated as below:

$$MAE_{ACC} = \frac{1}{n} \sum_{i=1}^n \left| \sum_{t=l_r}^{L_r} y_t - \sum_{t=l_r}^{L_r} \hat{y}_t \right|_i \quad (19)$$

$$RMSE_{ACC} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\sum_{t=l_r}^{L_r} y_t - \sum_{t=l_r}^{L_r} \hat{y}_t \right)_i^2} \quad (20)$$

The claim sequences can be divided into several patterns (i.e., irregular, decreasing, and increasing), and the previous metrics can not reflect the pattern-wise evaluation results. Therefore, to evaluate the models pattern-wisely, Dynamic Time Warping (DTW) clustering [46] is applied to the claim sequences, dividing the claim sequences into irregular, decreasing, and increasing patterns, and MAE and RMSE are measured depending on the patterns. The pattern-wise MAE and RMSE are calculated by the Equations 17 and 18 depending on their claim patterns.

In claim prediction, it is important that the predicted claim sequence has a pattern similar to the real pattern, especially

when the claim sequence has an increasing pattern. Therefore, to evaluate the predicted claims in terms of pattern similarity, the DTW which is trained on claim sequences to divide claim sequence patterns is used to classify the pattern of the predicted claims, as in [47]. Given c_i is the claim sequence pattern of y_i , the claim pattern prediction \hat{c}_i is defined as $\hat{c}_i = DTW(\hat{y}_i)$. Based on the classification results of the DTW, the accuracy and f1-score are measured for pattern similarity evaluation.

E. EXPERIMENTAL RESULTS

The proposed model is verified by conducting experiments depending on the target reliability. Since the automobile reliability prediction can be performed from a time and mileage reliability perspective, the experiments are designed to compare reliability prediction results based on each perspective. The proposed models and benchmark models are compared in terms of the RMSE, MAE, and pattern similarity. The RMSE and MAE are divided into 0, 1, and 2 depending on the claim sequence patterns (i.e., irregular, decreasing, and increasing) to reflect pattern-wise evaluation results. In addition, the results contain the accumulated RMSE and MAE which are the evaluation metrics for accumulated predicted claims within warranty periods, whereas the overall RMSE and MAE indicate the average errors of each step during warranty periods. Each accumulated metric and overall metric is denoted as ACC and OVR, respectively, in the result table. Regarding the RMSE and MAE of each pattern, the scores are calculated in the same way as the overall metrics (OVR) which average errors of each step during warranty periods. To evaluate the models in terms of pattern similarity, evaluation metrics include the accuracy and f1 score. Since the decreasing and increasing patterns account for a relatively small portion, it is necessary to focus more on the f1 score. Furthermore, the prediction results are investigated by visualizations.

Table 3 presents the experimental results for predicting future claim sequences when the target reliability is time. The results suggest that the proposed model, CRFormer, generally achieves improved performance compared to the previous methods in terms of the RMSE, MAE, and f1 score.

TABLE 4. Comparison results of mileage reliability prediction in terms of claim prediction and pattern metrics. 0, 1, and 2 in RMSE and MAE denote irregular, decreasing, and increasing patterns, respectively.

Model	RMSE					MAE					Pattern	
	ACC	OVR	0	1	2	ACC	OVR	0	1	2	Accuracy	F1
Weibull	163.20	3.29	1.51	2.38	10.47	59.83	1.31	0.91	1.76	5.70	0.90	0.53
ARIMA	383.36	7.25	1.12	31.38	8.88	66.07	1.15	0.64	10.16	5.27	0.89	0.46
RNN	139.61	3.36	1.44	4.80	8.98	59.18	1.40	0.79	3.82	5.30	0.91	0.62
LSTM	167.34	3.87	1.11	3.18	11.60	56.25	1.24	0.49	1.87	6.71	0.92	0.59
GRU	122.44	3.04	1.12	4.27	8.63	51.71	1.12	0.51	3.26	5.58	0.92	0.63
Transformer	118.15	2.89	1.12	5.64	7.74	48.03	1.17	0.61	4.07	4.90	0.91	0.57
Informer	134.70	3.05	1.71	6.92	6.95	87.51	1.89	1.52	3.90	4.14	0.91	0.56
Autoformer	310.68	5.64	1.44	29.32	9.72	70.58	1.89	1.02	10.34	5.59	0.89	0.50
CRFormer	113.34	2.78	1.19	3.04	8.14	51.43	1.04	0.51	1.93	5.57	0.92	0.62
CRFormer-F	151.58	3.58	1.58	5.19	9.55	54.42	1.25	0.64	2.64	5.90	0.92	0.70

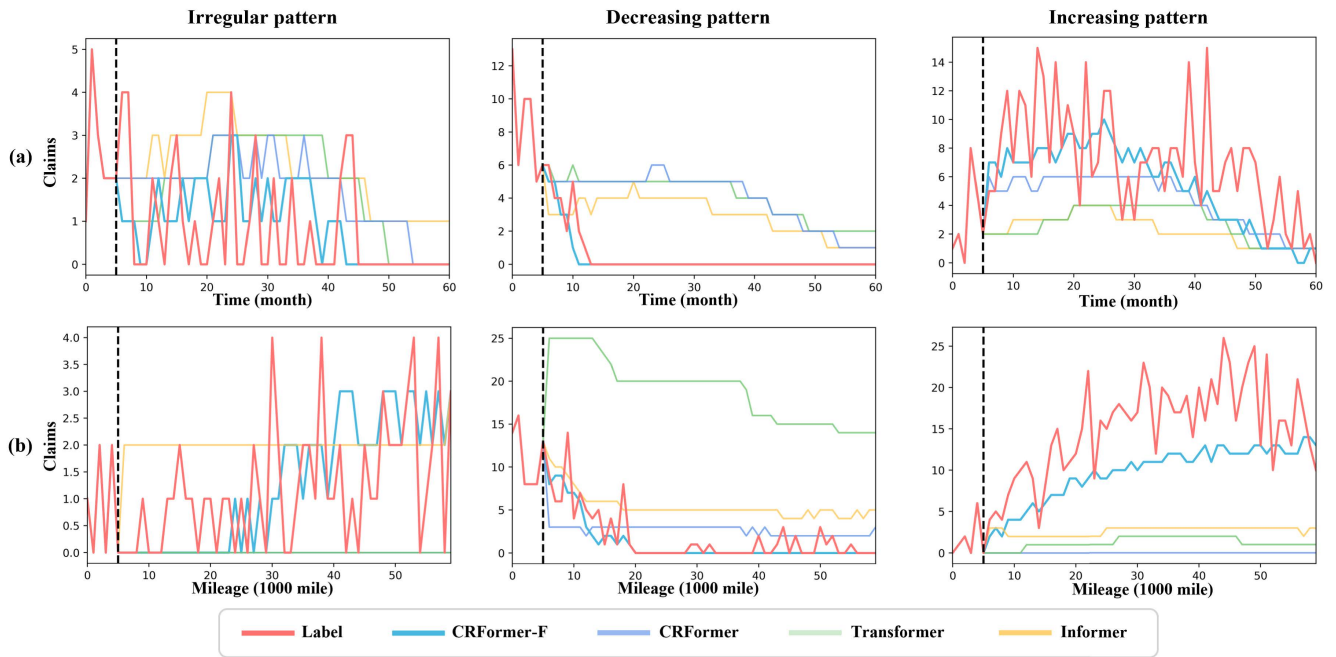


FIGURE 4. Prediction results visualization depending on claim patterns and reliability perspectives when the input sequence length is 6. The black vertical dashed lines show the given input sequence length. (a) Prediction results for time reliability. (b) Prediction results for mileage reliability.

CRFormer shows improved performance in terms of the overall RMSE and MAE, which are the average errors of each step over warranty periods, and the superiority is also revealed through the evaluation result of accumulated predicted claims. Furthermore, compared to Transformer which does not apply the nested sequence, CRFormer improves the performance, and it indicates the effect of the nested sequence on the performance improvement. Although CRFormer-F achieves a lower RMSE than Transformer and Informer, the f1 score is the highest, which indicates CRFormer-F can follow the actual pattern of the claim sequence. Interestingly, Autoformer and ARIMA show relatively lower performance than other models, which implies that the time-series characteristics of the short claim sequence do not last on the remaining long future claim sequence. In contrast, although the Weibull distribution and RNNs can capture decreasing

patterns, they showed poor performance in increasing patterns.

Meanwhile, the comparison result for the mileage reliability prediction demonstrates a similar result. Table 4 suggests that CRFormer is superior to the previous methods also in mileage reliability prediction. This indicates that the nested sequence fusing different reliabilities operates complementarily. As addressed in the experiment results of Table 3, the comparison between CRFormer and Transformer shows the effect of the nested sequence. By adopting both reliability perspectives, CRFormer can predict accurate future claims with enriched representations. Furthermore, the f1 score of CRFormer-F increases significantly, whereas the f1 scores of other models generally decrease compared to time reliability prediction. By comparing the mileage and time reliability prediction results, it is observed that the feature information

is more effective in determining the pattern information of the mileage-based sequence.

From the above experiments, the results suggested that the proposed models can extract enriched representations even with short-term input sequences. Additionally, the nested sequences contributed to performance improvement regardless of the reliability type, and the feature information guides the capture of the pattern information of the claim sequences. To investigate the prediction results of the proposed models, the predicted claims are visualized depending on the reliability perspectives and claim patterns. As shown in Figure 4, the prediction results of the proposed models capture the real claim patterns better, especially in the decreasing and increasing patterns, compared to other Transformer-based models. Regarding increasing patterns in the mileage reliability prediction, CRFormer-F can even capture the increasing patterns, whereas the others can not react at all. The poor performance of the other models in increasing patterns was also revealed through the low f1 scores.

F. FURTHER EXPERIMENTS

In the further experiment, the experiments examine the effect of the proposed methods, scaling, and input sequence length on the model performance, respectively.

1) ABLATION EXPERIMENT

The ablation experiment investigates the effect of each proposed method on the model performance. Table 5 lists the prediction results of the diverse model versions based on Transformer as the base model, depending on the proposed methods. Each proposed method is accumulated from the base model. Among the model versions in Table 5, CA & linear refers to the Context Attention and a prediction layer that this study proposed instead of Transformer decoder. Nested sequence represents the nested sequence concept and the proposed embedding module. Feature represents the usage of categorical features in the Context Attention and a prediction layer. As shown in Table 5, CA & linear and nested sequence generally improve the model performance in terms of the RMSE, MAE, and f1 score. Furthermore, the nested sequence contributes to improving the performance regardless of the

TABLE 5. Comparison results of reliability prediction for time and mileage depending on the proposed methods.

Obj.	Ver.	RMSE	MAE	Pattern
				F1
Time	Base	4.01	1.38	0.59
	+ CA & linear	3.54	1.17	0.59
	+ Nested sequence	3.46	1.21	0.61
	+ Feature	4.13	1.17	0.63
Mileage	Base	2.89	1.17	0.57
	+ CA & linear	3.09	1.12	0.60
	+ Nested sequence	2.78	1.04	0.62
	+ Feature	3.58	1.25	0.70

target reliability, which indicates that the nested sequence and its embedding operate in a complementary manner. Although the use of features decreases the performance of RMSE and MAE, it significantly improves the f1 score compared to other versions, thereby indicating that categorical features are effective in finding hidden representations of patterns.

2) SCALING EXPERIMENT

The scaling experiment analyzes the effect of the scaling technique on the performance of the proposed models. For scaling, sales information is used in Table 1. Given that i is the unique car made in a specific year and n_i is the number of sales of i , the input sequence $x_i \in \mathbb{R}^{l_r \times L_r}$ and output sequence $y_i \in \mathbb{R}^{L_r - l_r}$ is defined as $x_i = Constant \times x_i/n_i$ and $y_i = Constant \times y_i/n_i$, respectively. *Constant* is used as the sales unit adjusting scaling and is set as *Constant* = 10,000. The scaled input and output are used in the training phase, and the prediction results are transformed into the original scale for a fair comparison. In addition, to investigate the possibility of using sales information as a feature, not as a scaling factor, CRFormer-FS uses sales information as a feature while keeping the claim features. As shown in Table 6, the results suggest that the performance of the proposed models with scaling is similar to or worse than that of the models without scaling in terms of RMSE and MAE. However, when the scaling technique is applied, there is a significant performance decrease in terms of the f1 score, and a similar tendency of performance change occurs also in CRFormer-FS. From the experiment result, it is reasonable to deduce that the scaling technique can not rather reflect the original scaling difference of each claim sequence, thereby the effect appears in a lower f1 score.

TABLE 6. Comparison results of reliability prediction for time and mileage depending on scaling.

Obj.	Model	Scaling	RMSE	MAE	Pattern
					F1
Time	CRFormer		3.46	1.21	0.61
	CRFormer	✓	3.75	1.28	0.57
	CRFormer-F		4.13	1.17	0.63
	CRFormer-F	✓	3.89	1.13	0.63
	CRFormer-FS		3.86	1.12	0.59
Mileage	CRFormer		2.78	1.04	0.62
	CRFormer	✓	3.32	1.17	0.61
	CRFormer-F		3.58	1.25	0.70
	CRFormer-F	✓	3.52	1.20	0.66
	CRFormer-FS		3.35	1.15	0.65

3) SEQUENCE LENGTH EXPERIMENT

In this experiment, different sequence length inputs are used to verify the proposed models. The input sequence length is set to 6, 9, and 12 for the proposed models, and examine the difference in performance achievement depending on the input sequence length. Table 7 shows that the performance generally increases as the input sequence length increases. In particular, a significant performance improvement is

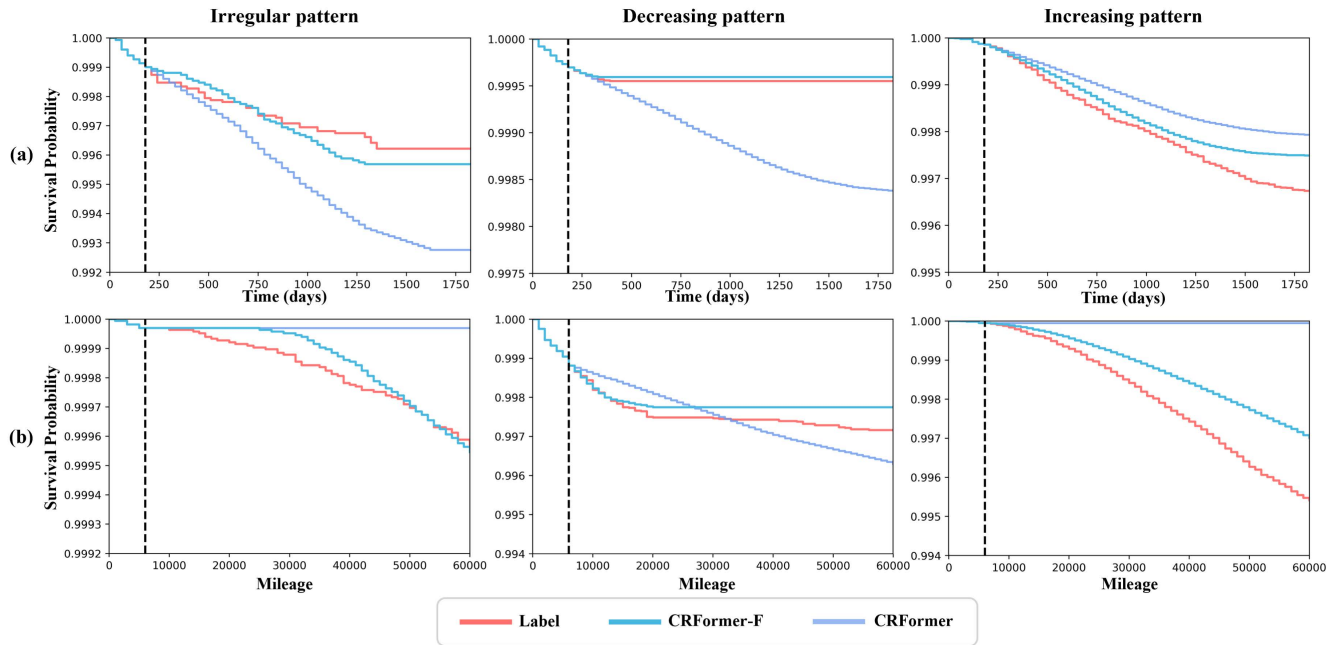


FIGURE 5. Survival curves visualization depending on claim patterns and reliability perspectives when the input sequence length is 6. The black vertical dashed lines show the given input sequence length. (a) Estimated survival curve for time reliability. (b) Estimated survival curve for mileage reliability.

TABLE 7. Comparison results of reliability prediction for time and mileage depending on input sequence length.

Obj.	Model	Length	RMSE	MAE	Pattern
					F1
Time	CRFormer	6	3.46	1.21	0.61
	CRFormer	9	3.27	1.05	0.71
	CRFormer	12	3.26	0.96	0.76
	CRFormer-F	6	4.13	1.17	0.63
	CRFormer-F	9	3.95	1.10	0.70
	CRFormer-F	12	3.52	1.02	0.72
Mileage	CRFormer	6	2.78	1.19	0.62
	CRFormer	9	3.10	1.11	0.66
	CRFormer	12	3.16	1.11	0.68
	CRFormer-F	6	3.58	1.58	0.70
	CRFormer-F	9	2.90	1.50	0.74
	CRFormer-F	12	2.84	1.25	0.73

observed in the f1 score, which indicates the proposed models can better find the claim patterns when the sequence is guaranteed to have sufficient length. Thus, the experimental result suggests that the proposed models can exploit more sequential information to predict claims and patterns of the sequence as the input sequence length increases.

V. RELIABILITY ANALYSIS

In this section, reliability analysis is conducted by the claim predictions. Since the proposed models, designed for time-series prediction, only consider the uncensored data within warranty periods of 5-year or 6000-mileage, the proposed models do not provide the reliability aspects while

considering the censored data. Although the predicted automobile component future claims and their patterns can be used for decision making, considering censored data is necessary to analyze the reliability of automobile components within warranty periods.

Reliability analysis means analyzing whether the objects can perform their functions under given conditions. In the automotive industry, reliability analysis is conducted over time or mileage. This can be analyzed through the survival probabilities of automotive components under given conditions by considering both censored and uncensored data [9]. To conduct the reliability analysis, it is necessary to determine the number of censored data [48], [49]. Given the sales information of the unique car released in the specific year in Section IV, the claim frequencies and sales information within warranty periods can be used for obtaining the censored and uncensored data.

To conduct a reliability analysis, the Kaplan-Meier estimator is adopted, which is a useful non-parametric method to estimate the reliability curve in the presence of the censored and uncensored data [50]. The survival function $\hat{S}(t)$ is estimated as follows:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \tag{21}$$

where t_i is the i^{th} point of time or mileage when failure occurs and d_i is the number of failures that occurs at t_i . n_i is the individuals known to have survived up to t_i .

Based on the above equation, the survival curves of real and predicted claims are estimated. Figure 5 shows the visualization results of the estimated survival curves depending

on claim patterns and reliability perspectives. Each case is the same as that used in Figure 4 for the predicted claim sequence visualizations. As shown in Figure 5, the proposed models can follow the actual survival curve. Furthermore, the ability of CRFormer-F to accurately predict increasing claims is also revealed in the survival curves. Compared to the survival curves of CRFormer, the survival curves of CRFormer-F are similar to the actual curves, thereby indicating that the utilization of claim features is also effective in both the time and mileage reliability analyses. Based on the above results, it is confirmed that the outstanding time-series prediction performance of the proposed models secures the reliability analysis. Finally, the survival probabilities of automobile components estimated by reliability analysis can be compared without constraints of considering the scales of failure frequencies and sales within warranty periods, and hence, can provide more practical information to decision-makers.

VI. CONCLUSION

This study proposed CRFormer, based on Transformer encoder, applied to the automotive component reliability prediction. Compared to the previous methods, CRFormer acquired enriched representations even with short-term inputs by designing the nested sequence and its embedding module, which fuses the time and mileage-based sequence, and achieved better prediction performance in both time and mileage reliability predictions. It indicates that the proposed models can assist early decision-making with more accurate information. In addition, through diverse evaluation metrics (i.e., overall claim prediction, pattern-wise claim prediction evaluation metrics, and pattern similarity metrics), the prediction results can be investigated in diverse aspects. Although using categorical features did not contribute to the improvement of the claim prediction performance, it significantly affected the improvement of the f1 score which is related to the accuracy of the increasing pattern. It suggests the proposed models can accurately predict future claims regardless of sequence patterns unlike the previous models which can accurately predict only irregular or decreasing patterns. Considering the importance of detecting automotive components which will be in failure, the proposed model can provide more practical information to decision-makers. Furthermore, visualizations of the prediction results and reliability analysis based on the survival curve can assist in intuitively understanding the prediction results with claim patterns and the survival probability of the automotive components. Finally, these predictions and analyses make it feasible not only to secure the lifespan of automotive components and decide on appropriate warranty periods but also to provide early feedback for manufacturers and stock managers, thus, it can result in reducing warranty and unexpected costs.

ACKNOWLEDGMENT

(Hyun Joon Park and Taehyeong Kim contributed equally to this work.)

REFERENCES

- [1] Z. Zhang, S. He, Z. He, and A. Dai, "Two-dimensional warranty period optimization considering the trade-off between warranty cost and boosted demand," *Comput. Ind. Eng.*, vol. 130, pp. 575–585, Apr. 2019.
- [2] B. Rai and N. Singh, "A modeling framework for assessing the impact of new time/mileage warranty limits on the number and cost of automotive warranty claims," *Rel. Eng. Syst. Saf.*, vol. 88, no. 2, pp. 157–169, May 2005.
- [3] D. N. P. Murthy, "Product warranty and reliability," *Ann. Oper. Res.*, vol. 143, no. 1, pp. 133–146, Jan. 2006.
- [4] S. Ho and M. Xie, "The use of ARIMA models for reliability forecasting and analysis," *Comput. Ind. Eng.*, vol. 35, nos. 1–2, pp. 213–216, 1998.
- [5] B. Yang and X. Li, "A study on software reliability prediction based on support vector machines," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage.*, Dec. 2007, pp. 1176–1180.
- [6] B. Kareem and A. S. Lawal, "Spare parts failure prediction of an automobile under criticality condition," *Eng. Failure Anal.*, vol. 56, pp. 69–79, Oct. 2015.
- [7] A. R. Saputro, H. Murfi, and S. Nurrohman, "Analysis of deep neural networks for automobile insurance claim prediction," in *Proc. Int. Conf. Data Mining Big Data*. Singapore: Springer, 2019, pp. 114–123.
- [8] R. Khoshkangini, P. S. Mashhadi, P. Berck, S. G. Shahbandi, S. Pashami, S. Nowaczyk, and T. Niklasson, "Early prediction of quality issues in automotive modern industry," *Information*, vol. 11, no. 7, p. 354, Jul. 2020.
- [9] T. Aziz and M. R. Karim, "Assessment and prediction of reliability of an automobile component using warranty claims data," *Rel., Theory Appl.*, vol. 14, no. 1, pp. 63–78, 2019.
- [10] M.-W. Lu, "Automotive reliability prediction based on early field failure warranty data," *Qual. Rel. Eng. Int.*, vol. 14, no. 2, pp. 103–108, Mar. 1998.
- [11] H. Moskowitz and Y. H. Chun, "A Poisson regression model for two-attribute warranty policies," *Nav. Res. Logistics*, vol. 41, no. 3, pp. 355–376, Apr. 1994.
- [12] N. D. Singpurwalla and S. Wilson, "The warranty problem: Its statistical and game-theoretic aspects," *SIAM Rev.*, vol. 35, no. 1, pp. 17–42, Mar. 1993.
- [13] A. Amin, L. Grunske, and A. Colman, "An approach to software reliability prediction based on time series modeling," *J. Syst. Softw.*, vol. 86, no. 7, pp. 1923–1932, Jul. 2013.
- [14] J.-G. Lee, T. Kim, K. W. Sung, and S. W. Han, "Automobile parts reliability prediction based on claim data: The comparison of predictive effects with deep learning," *Eng. Failure Anal.*, vol. 129, Nov. 2021, Art. no. 105657.
- [15] S. Meng, H. Wang, Y. Shi, and G. Gao, "Improving automobile insurance claims frequency prediction with telematics car driving data," *ASTIN Bull., J. IAA*, vol. 52, no. 2, pp. 1–29, 2021.
- [16] G. S. Wasserman, "An application of dynamic linear models for predicting warranty claims," *Comput. Ind. Eng.*, vol. 22, no. 1, pp. 37–47, Jan. 1992.
- [17] A. Kleyner and P. Sandborn, "A warranty forecasting model based on piecewise statistical distributions and stochastic simulation," *Rel. Eng. Syst. Saf.*, vol. 88, no. 3, pp. 207–214, Jun. 2005.
- [18] K. D. Majeske, "A non-homogeneous Poisson process predictive model for automobile warranty claims," *Rel. Eng. Syst. Saf.*, vol. 92, no. 2, pp. 243–251, Feb. 2007.
- [19] A. Akbarov and S. Wu, "Warranty claim forecasting based on weighted maximum likelihood estimation," *Qual. Rel. Eng. Int.*, vol. 28, no. 6, pp. 663–669, Oct. 2012.
- [20] Z. Xu and J. H. Saleh, "Machine learning for reliability engineering and safety applications: Review of current status and future opportunities," *Rel. Eng. Syst. Saf.*, vol. 211, Jul. 2021, Art. no. 107530.
- [21] S. Wu and A. Akbarov, "Support vector regression for warranty claim forecasting," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 196–204, Aug. 2011.
- [22] M. A. Fauzan and H. Murfi, "The accuracy of XGBoost for insurance claim prediction," *Int. J. Adv. Soft Comput. Appl.*, vol. 10, no. 2, pp. 1–13, 2018.
- [23] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 379, no. 2194, Apr. 2021, Art. no. 20200209.
- [24] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*.
- [25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

- [26] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
- [27] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.
- [28] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.
- [29] M. Liu, A. Zeng, Z. Xu, Q. Lai, and Q. Xu, "Time series is a special sequence: Forecasting with sample convolution and interaction," 2021, *arXiv:2106.09305*.
- [30] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," *Int. J. Forecasting*, vol. 36, no. 3, pp. 1181–1191, Jul. 2020.
- [31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [32] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
- [33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [34] 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 *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.
- [35] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer*, 2020, pp. 213–229.
- [36] H. Akbari, L. Yuan, R. Qian, W.-H. Chuang, S.-F. Chang, Y. Cui, and B. Gong, "VATT: Transformers for multimodal self-supervised learning from raw video, audio and text," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 24206–24221.
- [37] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," 2019, *arXiv:1912.09363*.
- [38] N. Wu, B. Green, X. Ben, and S. O'Banion, "Deep transformer models for time series forecasting: The influenza prevalence case," 2020, *arXiv:2001.08317*.
- [39] H. J. Park, M. S. Lee, D. I. Park, and S. W. Han, "Time-aware and feature similarity self-attention in vessel fuel consumption prediction," *Appl. Sci.*, vol. 11, no. 23, p. 11514, Dec. 2021.
- [40] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, and X. Yan, "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
- [41] N. Kitaev, L. Kaiser, and A. Levskaya, "Reformer: The efficient transformer," 2020, *arXiv:2001.04451*.
- [42] 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 *Proc. AAAI*, 2021, pp. 11106–11115.
- [43] J. Xu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 1–12.
- [44] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.
- [45] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2015, *arXiv:1508.04025*.
- [46] M. Müller, "Dynamic time warping," in *Information Retrieval for Music and Motion*. Berlin, Germany: Springer, 2007, pp. 69–84.
- [47] J. Lin, S. Williamson, K. Borne, and D. DeBarr, "Pattern recognition in time series," in *Advances in Machine Learning and Data Mining for Astronomy*, vol. 1. New York, NY, USA: Taylor & Francis, 2012.
- [48] L. Attardi, M. Guida, and G. Pulcini, "A mixed-Weibull regression model for the analysis of automotive warranty data," *Rel. Eng. Syst. Saf.*, vol. 87, no. 2, pp. 265–273, 2005.
- [49] V. Krivtsov, "A Bayesian estimation procedure of reliability function for lifetime distributions," *Int. J. Performability Eng.*, vol. 13, no. 2, p. 129, 2017.
- [50] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Amer. Statist. Assoc.*, vol. 53, no. 282, pp. 457–481, 1958.



HYUN JOON PARK received the B.S. degree in industrial engineering from Hongik University, Seoul, South Korea, in 2020. He is currently pursuing the M.S. degree in industrial and management engineering with Korea University, Seoul. His research interests include attention mechanism, time series modeling, speech enhancement, and deep learning.



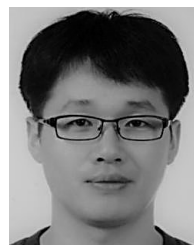
TAEHYEONG KIM received the B.S. degree in information statistics from Korea University, Sejong, Seoul, South Korea, in 2019, where he is currently pursuing the Ph.D. degree in industrial and management engineering. His research interests include anomaly detection, semi-supervised learning, and deep learning.



YOUNG SEOK KIM received the B.S. degree in industrial engineering from Dongguk University, Seoul, South Korea, in 2021. He is currently pursuing the M.S. degree in industrial and management engineering with Korea University, Seoul. His research interests include NLP, recommendation systems, and deep learning.



JINHONG MIN received the B.S. degree in mathematics and computer science and engineering from the Konkuk University of Korea, Seoul, South Korea, in 2022. He is currently pursuing the M.S. degree with the School of Industrial and Management Engineering, Korea University, Seoul. His current research interests include computer vision, image generation, and multimodal learning.



KI WOO SUNG received the Ph.D. degree in industrial engineering from Sungkyunkwan University, South Korea, in 2016. He is currently a Data Analyst and a Senior Research Engineer with Hyundai Motor Company Research and Development Center. His research interests include reliability analysis and data analysis related to automobiles.



SUNG WON HAN received the M.S. degrees in operations research, in statistics, and in mathematics from the Georgia Institute of Technology, in 2006, 2007, and 2010, respectively, and the Ph.D. degree from the School of Industrial and Systems Engineering (statistics), Georgia Institute of Technology. He was a Senior Research Scientist with the Division of Biostatistics, School of Medicine, New York University, and a Postdoctoral Researcher with the Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics and Epidemiology, School of Medicine, University of Pennsylvania. He is currently an Associate Professor with the School of Industrial and Management Engineering, Korea University. His research interests include probabilistic graphical model, network analysis, machine learning, and deep learning.

• • •