IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 10 May 2023, accepted 18 June 2023, date of publication 18 July 2023, date of current version 7 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3296314

RESEARCH ARTICLE

Prediction of Future Collision Risk for Truck Drivers Using the Time-Series Autonomic Nerve Function

NAO ITO, SHUNSUKE MINUSA, TAKESHI TANAKA, YUN LI, AND HIROYUKI KURIYAMA
Research and Development Group, Hitachi Ltd., Tokyo 185-8601, Japan

Corresponding author: Nao Ito (nao.ito.xr@hitachi.com)

This work was supported by the Research and Development Group, Hitachi Ltd.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Healthcare for truck drivers is an important issue. To prevent fatigue-related collisions among drivers, objective assessments of their physiological states are essential. A simple and quantitative evaluation method for fatigue involves the use of autonomic nerve function (ANF) indices obtained from heart rate variability analysis. However, predicting the occurrence of crashes using only physiological data is challenging. In most previous studies, the targets of driving situations have been often limited, or the prediction targets have been set as the driver's internal state rather than the accident. In this paper, we propose a novel collision risk prediction model using ANF and several simple external information types, which can be extracted from standard in-vehicle sensors without limiting the driving scene. Our experiments using actual truck drivers' data reveal that the proposed model can achieve collision risk prediction for the following 30 min with an accuracy of 74.9% recall and 0.79 AUC. Furthermore, we discover that simple external information obtained based on the vehicle speed significantly contributes to the prediction accuracy. As our prediction method only requires commonly equipped sensors as the sources of external information, this method is expected to be easily implemented not only for truck driving but also for general vehicle driving, where crashes are often likely.

INDEX TERMS Autonomic nerve function, collision, deep learning, fatigue, heart rate variability, long short-term memory, stress, time-series prediction, wearable device.

I. INTRODUCTION

Healthcare for professional drivers is an important issue. Especially for truck drivers, daily health checks are particularly important because of their hard work and the risk of accidents. Predicting traffic crashes caused by the bad health conditions of drivers is an important but challenging issue. So far, related research has considered two main perspectives for accident risk estimation. Among these, one involves predicting the risk of all types of traffic accidents from multiple types of extensive data. Previous studies on the prediction of accident occurrence [1] have primarily utilized environmental information (i.e., weather conditions and the traffic flow); however, driver-related data have been rarely utilized, except static information such as age [2], [3], [4]. For instance, Effati et al. [3] used driver data and several environmental data types including vehicle data, weather conditions, and road conditions to predict the crash severity of vehicles. Although they implemented driver-related data including the age, gender, and education level, they did not use non-static information data, such as physiological features. In this manner, most conventional risk-predicting studies have been focused on public and static datasets.

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

The other perspective involves restricting the target of estimation to use drivers' physiological data. Li et al. [5] proposed a method for predicting driving risk when drivers performed high-risk driving operations such as changing lanes. Based on physiological data (i.e., eye movement and heart rate) and environmental information such as vehicle speed, they evaluated whether lane changing was risky or safe. Fruet et al. [6] focused on stress as a driving risk and predicted the stress levels from physiological signals including galvanic skin responses, electrocardiograms, electromyograms, and breathing signals. These studies succeeded in utilizing physiological data during driving by specifying the situation and limiting the estimation target to the driver's state rather than crashes.

The common understanding for both perspectives is that a complex causal relationship exists between driver physiology and the occurrence of accidents. To resolve this complexity, we aim to predict the accident risk in generic situations by using driver biometrics. To achieve this aim, we focus on fatigue as a generic element within the driver related to accidents.

We assess fatigue based on heartbeat data obtained from wearable devices, which can quantitatively express a person's internal state, including fatigue. As the sensors are easy to wear, a versatile assessment is expected. Previously, we developed a prediction method for the collision risk from heartbeat data to realize safe truck driver operations [7], [8], [9], [10]. In these previous studies, we discovered that several indices from heartbeat data are associated with the risk of rear-end collision by truck drivers [8], [9].

In this paper, we propose a risk prediction method that uses the driver's physiological information and simple external information obtained primarily from in-vehicle sensors. To address the difficulty in handling physiological data during driving, we performed three different preprocessing steps on the physiological data. We also evaluated the proposed method using real data obtained from truck drivers.

II. PROPOSED METHODS

We utilize autonomic nerve function (ANF) indices for the driver's physiological data. The ANF can describe fatigue, as derived from the heartbeat signals. It is obtained by heart rate variability (HRV) analysis and changes owing to stress, fatigue, and sleepiness [11]. The HRV-based ANF has been widely used because heartbeat signals can be easily recorded using wearable devices.

This paper proposes a collision risk prediction method using an ANF based on a machine-learning time-series model. We focus on professional drivers, such as truck drivers, for whom fatigue-related accidents are a serious problem. For the input data of the model, we extract simple external data including the driver's static data and environmental data. The model comprises three key preprocessing steps for physiological data during driving.

A. PREDICTION MODEL USING TIME-SERIES ANF DATA

Fig. 1 shows the flow of collision risk prediction using ANF data during driving. In this study, a model was developed to predict collision risk for 30 min after the time of measurement; this ensured that drivers had a sufficient margin to act after receiving an alert.

For the prediction, a deep-learning model was used, and the input layer was set for time-series data. Previous reports have shown that the ANF is useful for the prediction of a driver's health condition such as drowsiness [12] or stress [13]. In this study, based on the hypothesis that a driver's health condition affects the probability of an accident, we mainly utilized the ANF to predict collision risk. Conventionally, the ANF is recorded under stable conditions, such as resting or with the eyes closed. In this study, however, the ANF was obtained from drivers under constantly changing conditions. To account for this, three types of preprocessing were introduced.

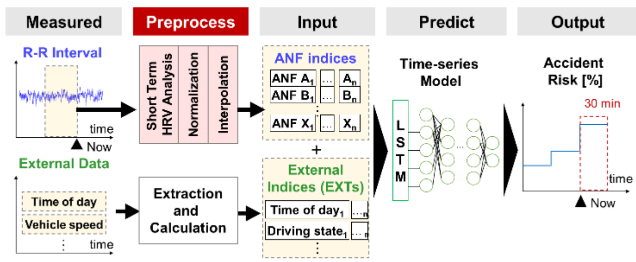

FIGURE 1. Overview of collision risk prediction. The method is based on a deep-learning time-series model. To extract features from complex and irregularly varying heart rate data, three types of preprocessing were applied.

First, HRV analysis was performed with a reasonable analysis window width for application to the driving state. In the HRV analysis, the ANF indicators were calculated from R-R interval (RRI) by frequency, time, and nonlinear domain analysis. In a time-series prediction, it is preferable to have a high temporal resolution, i.e., a short analysis window. However, if the time window is too short, the reliability of the HRV analysis may be reduced, and a trade-off could exist between the reliability and resolution. In a previous study [9], we found that the appropriate analysis window width for calculating the ANF is 2 min under several different conditions during driving. Here, following the previous study, we divided the measured RRI into 2 min segments without overlapping intervals. The HRV analysis was then performed on the segmented RRI data, and the ANF indices were calculated and combined to obtain the time-series data.

Second, we considered normalization focusing on individual differences in the ANF indices; notably, each type of ANF has a different numerical scale. For example, the distribution range of one ANF index is $10^1 - 10^4$, whereas that of another is $10^0 - 10^1$. To properly learn the model parameters, the distribution range of each indicator must be normalized before being input into the prediction model [14]. Min–max normalization, which uses the maximum and minimum values of the measurements, and Robust Scaler [15], which uses the quartile points, are commonly used. Robust Scaler, in particular, is expected to be stable as it considers outliers. However, the ANF indicator has large individual differences in distribution, and Robust Scaler may exclude most of an individual's data as outliers. Therefore, we devised a normalization process that considers the characteristics of the ANF data. We decided to normalize the data using the following parameters, which are more robust to outliers than Robust Scaler:

$$ANF_i' = (ANF_i - P_{50}) / (P_{97.5} - P_{2.5}), \tag{1}$$

where ANF_i represents the i-th type of ANF index. $P_{2.5}$ and $P_{97.5}$ represent the 2.5 and 97.5 percentiles of the data group containing X, respectively. To include both healthy and fatigued states in the data and to exclude measurement abnormalities such as noise, the significance level was set at 5%. The values below the 2.5 percentile and above the 97.5 percentile of each ANF index measurement for all subjects were treated as outliers.

Third, interpolation of the time-series ANF data was performed. For time-series ANF in driving, in some cases, the data were partially missing (e.g., in the case of a vehicle driving in a tunnel, poor communication conditions, a vehicle stopping for several minutes, and noise generation such as luggage being carried). Pseudo-missing data were randomly added to the dataset with no missing data. Then, after optimizing, the interpolation was applied to the dataset with true missing data in the actual measurements.

B. EXTRACTION OF EXTERNAL INFORMATION

External factors (EXTs) were defined to complement the information required for risk prediction. EXTs refer to all factors that a driver may encounter during an accident, other than their own physiological factors. Table 1 shows the candidate EXT types considered in this study. EXTs can be divided into two types. First, environmental data can be recorded by sensors mounted inside and outside the vehicle; such data may include the presence of oncoming or following vehicles, the presence of pedestrians, road type, traffic congestion, weather, time of day, delivery delays, and vehicle speed. Second, EXTs also include static data about the driver, such as age, personality, and driving skills. Thus, various EXTs provide useful information for prediction, but not all of them are necessarily available owing to the limited number of modalities provided by transport companies. Therefore, we limited the scope to a few modalities that are commonly or easily obtained.

We extracted the factors as EXTs closely related to the occurrence of driver fatigue and crash accidents. First, from environmental data that can be easily obtained from sensors installed inside and outside the vehicle, the vehicle speed and TIME OF DAY were selected; the vehicle speed was converted into two indices, DRIVING TIME and DRIVING STATE, to more clearly express the working conditions. DRIVING TIME is a numeric variable representing the time elapsed since the start of driving in a day. DRIVING STATE is a categorical variable (general road, highway, company premises, and stopped) that is estimated from the vehicle speed. Next, from the drivers' static data, AGE was chosen.

These four EXT indicators were chosen not only because they are basic data commonly handled by many transportation companies but also because they are associated with the ANF indices or the occurrence of driving crashes. Previous studies have reported changes in the ANF behavior with AGE in personal attribute data [16]. TIME OF DAY and DRIVING TIME can express the working conditions of drivers. A fatigue analysis study on logistics drivers in Japan [17] confirmed a significant relationship between variables representing working conditions, such as working hours and fatigue appearing in the ANF. Furthermore, it has been reported that the frequency of fatal crashes varies depending on the TIME OF DAY and that the type of behavior that mainly causes fatal crashes varies greatly depending on the DRIVING STATE (e.g., public roads and highways) [18].

Based on these findings, AGE, TIME OF DAY, DRIVING TIME, and DRIVING STATE were inputted as EXT data into the model, in addition to the ANF indices.

TABLE 1. Environmental-factor (EXT) candidates.

Type of data	Information
Data from sensors set inside or outside the vehicles	Oncoming or following vehicles Pedestrians Traffic congestion Weather **TIME OF DAY** Delivery delays **DRIVING TIME** **DRIVING STATE** etc.
Data from driver's attributes	**AGE** Personality Driving skill etc.

The four underlined EXTs were included as input to contribute to the prediction.

C. STRUCTURE OF THE RISK PREDICTION MODEL

The structure of the prediction model was fine-tuned using a machine-learning deep neural network. First, we adjusted the size of the prediction model, i.e., the number of layers in the model and the number of nodes per layer. Based on the preliminary hyperparameters search in terms of the prediction accuracy, five layers were chosen as the number of layers ranged from two to ten. The number of nodes for each layer was also compared for accuracy from 32, 64, 128 and 256 candidates. The number of nodes in the first layer was 128, halving to 64, 32 as one went through the layers, and doubling to 64 in the final layer. Second, the type of layers in the model was determined. Table 2 details the structure of the model. A feature extraction policy was set up so that the time-series characteristics of each indicator were extracted independently, and the interrelationships between indicators were captured. With this policy, we introduced long short-term memory (LSTM) [19] in the first layer of the model and a dense layer (fully connected layer) in layers 2–5. The model was set to solve a binary classification task to identify high and low collision risk.

The preprocessed ANF data (see section A) and the EXT data selected in section B were inputted into the model to learn collision risk prediction.

TABLE 2. Parameters of each layer in the proposed time-series prediction model.

Index	Layer type	Number of input, output	Activate function
1	LSTM	length of input data, 128	ReLU
2	Dense	128, 64	ReLU
3	Dense	64, 32	ReLU
4	Dense	32, 64	ReLU
5	Dense	64, 1	Sigmoid

III. EXPERIMENT

A. DEFINITION OF CRASH RISK

We defined the accident risks to be predicted by the model. The crashes considered in this study was limited to rear-end collisions, as this type of collision risk can be estimated from the driver's level of fatigue [8], [9]. Because the data on actual crashes are rare and insufficient for analysis, this study focuses on "near-miss situations," which may lead to rear-end collisions [9]. A classifier developed in a previous study [8] to detect near misses from vehicle behavior data was used, and the detected near-miss probability was considered as the collision risk. Among the near misses detected by the classifier, when a collision warning was issued owing to insufficient distance between vehicles, the scene was targeted for detection. The classifier outputted a continuous value of zero–one, which represents the probability of a near-miss. An output value of 0.83 was used as the decision threshold for near misses, where the classifier had a detection performance of 50% or more in terms of the recall [8]. In this study, the time range to be predicted was set at 30 min to give the driver sufficient time to act if a risk was detected. If the predicted value exceeded the threshold value of 0.83 at least once during the 30 min period, the interval was labeled "near-miss;" if it never exceeded the threshold value, the interval was labeled "not near-miss."

B. DATA MEASUREMENT AND EXTRACTION

To confirm the validity of the proposed method, we evaluated the performance on a real-world dataset obtained in a previous study [9]. The data corresponding to 26 drivers who were in daily truck transport operations were recorded over a period of approximately 3 months. Drivers with noisy or unusual HRV data, such as arrhythmias, and drivers with very limited experience were excluded from the analysis. The measurement data of the remaining 20 drivers were used as evaluation data for the model. This study's data were obtained according to the standards of internal review board on Research & Development group, Hitachi, Ltd., and the study was conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent before enrollment in this study.

The RRI of drivers during driving was measured using a wearable ECG-type heartbeat sensor (myBeat WHS-1, UNION TOOL CO., Tokyo, Japan) with a belt-type electrode on the chest. After measurement, the data were selected with correct measurements, normal arrival and departure times, and the majority of working hours spread throughout the day. The ANF indices were calculated from the RRI data based on HRV analysis. They represent sympathetic and parasympathetic nervous system activities and are used in clinical medicine as a means of quantifying and evaluating the degree of fatigue [20]. Following the analysis method guidelines [21], the RRI data were segmented using a moving window with a width of 120 s and a moving width of 120 s. The ANF indices shown in Table 3 were then calculated. The ANF can be quantified by frequency-domain analysis, time-domain analysis, and nonlinear analysis of RRI data [21], [22]. For example, the total power (TP) and sympathetic/parasympathetic activity balance (LF/HF), which is calculated based on the power spectrum obtained by frequency analysis, have been reported to correlate with subject fatigue in clinical science and automobile driving studies [20], [23]. Because the relationship between each ANF and the risk of a crash while driving is unknown, we decided to comprehensively utilize the ANF indices in Table 3. Henceforth, the ANF groups listed in Table 3 are expressed as ANFs. The calculated ANFs were combined in a time series every 30 min.

EXTs were measured in the same interval in which RRIs were measured. EXTs were obtained at each 2 min range and combined every 30 min (i.e., 15 time points). The mean of AGE was 48.0 ± 8.37, and the mean of DRIVING TIME was 4.31 ± 1.40 h. Note that DRIVING TIME is not the time of continuous driving but the total driving time during a single working day. Fig. 2 shows the distribution of the most frequent DRIVING STATE over 30 min. DRIVING STATE is estimated from the vehicle speeds every 2 min. The most frequent DRIVING STATE was general road, and the least frequent DRIVING STATE was stopped.

C. EVALUATION

The proposed method was evaluated from two different perspectives. First, we evaluated the effect of the presence or absence of actual missing data on the prediction performance. Two types of data were prepared: one with no missing data at all during the measurement period and one that always contained missing data. Then, after a preliminary experimental

TABLE 3. Calculated autonomic nerve function (ANF) indices.

Index Types	Index
Frequency-domain Index	LF; low-frequency (0.04–0.15 Hz) band in the power spectral density [ms^2]
	HF; high-frequency (0.15–0.4 Hz) band in the power spectral density [ms^2]
	TP; total power (0.04–0.4 Hz band) in the power spectral density [ms^2]
	LF/HF [a.u.]
	devLF; deviation score of LF [a.u.]
	devHF; deviation score of HF [a.u.]
	devTP; deviation score of TP [a.u.]
	CCVLF; coefficients of component variance (CCV) in LF [%]
	CCVHF; CCV in HF [%]
	CCVTP; CCV in TP [%]
	lnCCVLF; natural logarithm of CCVLF [a.u.]
	LFnu; density of the low-frequency band of TP [%]
Time-domain Index	Beats [bpm]
	AVGHR; average heart Rate [a.u.]
	MAXHR; maximum heart Rate [a.u.]
	MINHR; minimum heart Rate [a.u.]
	MeanNN; average RRI [ms]
	CVRR; coefficient of variation RRI [a.u.]
	RMSSD; Root mean square successive difference [ms]
	SDNN; standard deviation of normal-to-normal RRI [ms]
	SDSD; standard deviation of the differences between successive NN intervals [ms]
	NN50; number of NN intervals exceeding 50 ms [counts]
	pNN50; percentage of NN50 [%]
	absNN50; absolute value of NN50 [counts]
	abspNN50; percentage of absNN50 [%]
Nonlinear Index	Poincaré SD1; short-term RRI variability in Poincaré plots [ms]
	Poincaré SD2; long-term RRI variability in Poincaré plots [ms]
	Poincaré SD1/SD2 [a.u.]
	Poincaré S; product of SD1 and SD2 [ms]
	DFA alpha1; short-term scaling index in detrended fluctuation analysis (DFA) [a.u.]
	DFA alpha2; long-term scaling index in DFA [a.u.]
	Tone; index of the balance between sympathetic and vagal nerves in Tone-Entropy (T-E) analysis [a.u.]
	Entropy; index representing the sum of autonomic nervous activity in T-E analysis [a.u.]
	ApEn; approximate entropy of HRV [a.u.]

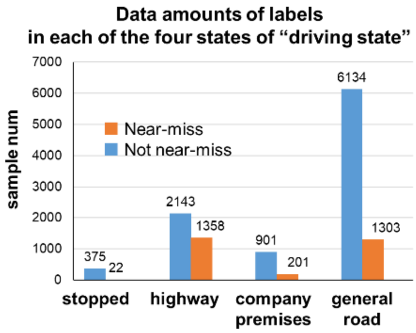

FIGURE 2. Near-miss and not near-miss data in each DRIVING STATE class. General road has the highest near-miss rate, and highway has the highest near-miss rate.

TABLE 4. Number of data with and without missing data in 30 min dataset.

Dataset type	Near-miss situations	Not Near-miss situations	Total
With missing	11372	38145	49517
Without missing	2884	9553	12437
With missing + interpolation	11343	38037	49380

comparison, mean-value interpolation was applied to the missing data. Table 4 shows the sizes of the 30 min input datasets with and without missing data. For the dataset with no missing data, the number of usable samples decreased from 49517 to 12437, and interpolation restored the number of samples to 49380. Second, we evaluated the effects of the changes in the types of input indicators: we trained and evaluated the ANFs only, EXTs only, and ANF plus EXT data. In addition, evaluation by ablation of the EXT indices was also performed. For the same trained model, each of the four EXTs was pseudo-missing, and the accuracies were compared. From this ablation experiment, it is possible to determine which of the EXTs contribute to the model's predictions. In these evaluations, if the data were pseudo-missed, the number was replaced by zero. In terms of the DRIVING STATE, which was represented as one-hot encoding, all elements of the one-hot vector were also replaced by zero when pseudo-missed.

We evaluated the accuracy of the proposed method to discuss its suitability for predicting the occurrence of collisions. The prediction model was optimized by Adam and trained with a learning rate of 0.001. Binary cross entropy was used as the loss function. As evaluation metrics, we used recall and the area under the curve (AUC), which represents the area under the receiver operating characteristic curve. The AUC was used to evaluate the discriminative ability to detect the presence or absence of near misses. Recall was used to evaluate the ability to detect near misses. Three-fold cross-validation was used for the evaluation. The original dataset was randomly assigned to three groups by subject. Data pre-processing, model training, and evaluation were conducted using Python 3.6, including Tensorflow 1.14, Keras 2.2, and scikit-learn 0.22.

IV. RESULTS

Table 5 compares the prediction accuracy with and without missing time-series data and with and without interpolation processing for missing data. The recall was 30.3%, and the AUC was 0.15 lower in the dataset with missing data than that in the dataset without missing data, but the recall recovered by 5.8% after interpolation. Table 6 compares the prediction performance for each combination of indices which are the input of the model. The accuracy was the best at 74.9% for recall and 0.70 for the AUC when both ANFs and all EXTs were used as input data. From ablation of EXTs, recall was most decreased when AGE was missing, and the AUC was most decreased when DRIVING STATE was missing. Across all evaluation results, the best performance was achieved when ANFs and all EXTs were used and trained on input data with a time range of 30 min without missing data.

TABLE 5. **Comparison of time-series predictions with and without missing measurements (30 min dataset).**

Input test data	Interpolation	recall	AUC
ANFs + EXTs (without missing)	-	<u>**74.9%**</u>	<u>**0.79**</u>
ANFs + EXTs (with missing)	No	44.6%	0.64
	Yes	50.4%	0.63

TABLE 6. **Comparison of prediction accuracy for each combination of indicators.**

Index of input data					Recall	AUC
ANFs	EXTs: AGE	EXTs: DRIVING STATE	EXTs: DRIVING TIME	EXTs: TIME OF DAY		
✓(*1)	-	-	-	-	29.0%	0.55
-	✓(*2)	✓(*2)	✓(*2)	✓(*2)	52.7%	0.69
✓	✓	-	-	-	51.4%	0.47
✓	-	✓	-		66.7%	0.25
✓	-	-	✓		53.1%	0.57
✓	-	-	-	✓	59.1%	0.56
✓	✓	✓	✓	✓	<u>74.9%</u>	<u>0.79</u>

*1 Trained with only ANF data.
*2 Trained with only EXT data.

V. DISCUSSION

In this study, we developed a risk prediction model using physiological and simple external data, considering the importance of dealing with multimodal data sources. Compared with existing models, the proposed model has no restrictions in driving situations, such as lane changing, but always predicts the risk of rear-end collisions. Our experimental results showed the high predictive performance of future accident risk during driving (recall, 74.9%; AUC, 0.79) and the effectiveness of the combination of the two types of input indices.

A. PERFORMANCE OF THE PROPOSED METHOD

First, interpolation increased the amount of usable data, as shown in Table 4, but significantly reduced performance, as shown in Table 5. Although interpolation improved the recall by 5.8%, it was not sufficient for practical use. The interpolation process was introduced to increase the amount of available data by dealing with missing measurements of physiological data during driving. These results suggest that our proposed method can only be applied during stable measurement situations, such as expressway driving, where no missing data are generated during the 30 min period.

Second, the contributions of the EXT indices were evaluated in the ablation experiments (Table 6). When the DRIVING STATE was pseudo-missing, the AUC seemed to decrease most effectively, whereas the recall did not decrease considerably. This indicates that when DRIVING STATE was missing, the ability to determine the presence or absence of a near-miss was greatly reduced, and the model tended to classify the scene as a ''near-miss,'' independent of the input. This is probably owing to the existence of label bias related to the DRIVING STATE. Fig. 2 shows the number of labels in each of the four states of the DRIVING STATE. The general road with the largest number of labels (7473) has a near-miss rate of 0.17. Conversely, 1358 of the 3501 highway data include near misses, which corresponds to a value of 0.39, more than twice that for the general road data. The likelihood of crashes is considered to be closely related to both the environment and internal conditions such as the driver's nervousness. External information on the driving speed was shown to be a potentially effective factor in controlling models that predict crashes from physiological data.

B. COMPARISON OF PREVIOUS STUDIES AND PROPOSED METHOD

Our results show that DRIVING STATE made a significant contribution to prediction. In a previous study [3], drivers' data (age, gender, and education level) and environmental data (vehicle data, weather conditions, and road infrastructure) were used to predict crashes; the prediction accuracy was 81.8% for the recall and 0.81 for the AUC. Their performance was better than the best results of our study: 74.9% for recall and 0.79 for AUC. They also suggested that high-speed driving was one of the most important causes of motor vehicle fatalities. As DRIVING STATE contributed the most to the prediction accuracy among the EXT data, our results also indirectly support their results.

The result of index ablation also showed the importance of using the ANF indices for risk prediction. Most crash prediction studies using machine learning have utilized real data [1]. Most of the information on the environment is publicly available and easy to obtain, but personal information is not. It has been suggested [1] that this data bias should be eliminated, and emphasis should be placed on hard-to-obtain data such as driver information. In our experiments, we extracted the ANF, i.e., physiological data that can represent the driver's internal state at a temporal resolution of 2 min. Although predicting crash risk using ANF data alone is difficult, we achieved an AUC performance of 0.79 by adding simple environmental data and drivers' static data. The effectiveness of using both environmental and personal information, the importance of which has been previously highlighted, was confirmed in this study.

C. LIMITATIONS AND FUTURE RESEARCH

One of the limitations of this study is the type of data. We proposed four EXTs, DRIVING TIME, DRIVING STATE, TIME OF DAY, and AGE, related to small-scale modalities that many companies can provide. However, not all EXT options could be verified. Future studies would include expanding the type of data. For example, the location data of individual drivers can be monitored by adding one condition: the driver is carrying a smartphone. Once the location over time is known, the four EXT indices could be estimated and obtained through applications and networks, even without in-vehicle sensors. Furthermore, we would be able to utilize public environmental information such as weather and traffic congestion, which have been used in many previous studies.

VI. CONCLUSION

To prevent fatigue-related accidents, we proposed a model for predicting future accident risk using driver ANF data and simple external information (EXT). Using the model with 34 ANFs and 4 EXT indicators, we predicted the accident risk within 30 min into the future with a recall of 74.9% and an AUC of 0.79 for a real-world truck drivers' dataset. The ablation evaluation of the input indicators showed that ANF and EXT each contributed to the prediction of the model. Our results suggest that the use of driver physiological and environmental data is effective for accident prediction, which finally contributes to accident prevention. If high-quality physiological data can be obtained appropriately, it may be possible to predict the occurrence of crashes related to a driver. Introduction of both ANF and EXT indicators to accident prediction studies will be useful for future driver safety management in the workplace.

APPENDIX

In this study, we set input time-series ANF data of 30 min for the evaluation. Here, we describe how we determined the input time range. As a preliminary study, the suitable input length was confirmed. Fig. 3 shows the process of creating time-series datasets in several different time ranges. If the RRI data for 2 min up to a point in time t are expressed as RRI_{t-2}^{t}, each of the ANF indices calculated in the HRV analysis and the risk calculated by the model are expressed by the following equations:

$$ANF_{i,t} = HRV(RRI_{t-2}^{t}, i), \tag{2}$$

$$X(t,n) = \begin{pmatrix} ANF_{1,t-2(n-1)} & \cdots & ANF_{1,t} \\ \vdots & \ddots & \vdots \\ ANF_{i,t-2(n-1)} & \cdots & ANF_{i,t} \\ EXT_{1,t-2(n-1)} & \cdots & EXT_{1,t} \\ \vdots & \ddots & \vdots \\ EXT_{j,t-2(n-1)} & \cdots & EXT_{i,t} \end{pmatrix}, \tag{3}$$

$$risk = model\,\{X(t,n)\}. \tag{4}$$

First, the ANF and EXT were calculated from RRI. i is the number of ANF index types, and j is the number of EXT index types. In this study, i takes values from one to 34, and j takes values from one to four, representing 34 ANF indices and four EXT indices, respectively. t is the end time of the measurement in minutes. Here, $ANF_{i,t}$ is calculated from the 2 min RRI_{t-2}^{t} by the function HRV. HRV applies the corresponding calculation formula depending on the type of ANF index i. $EXT_{i,t}$ is the external information in the interval for which RRI_{t-2}^{t} was measured. Second, the time-series data $X(t,n)$ are constructed. n is the number of data in the time series for each analysis window width. As the analysis window width was determined to be 2 min, for example, n=15 is used to represent a 30 min RRI, and the most past-time point is represented by $t-2(n-1) = t-28$ [min]. Finally, the time-series data $X(t,n)$ are input to the *model* proposed in this study, which outputs the predicted value *risk*.

As shown in Fig. 3, the time lengths of the input data ranged from 6 to 60 min in 6 min increments (i.e., 10 window conditions). The number of data pieces ranged from 3 to

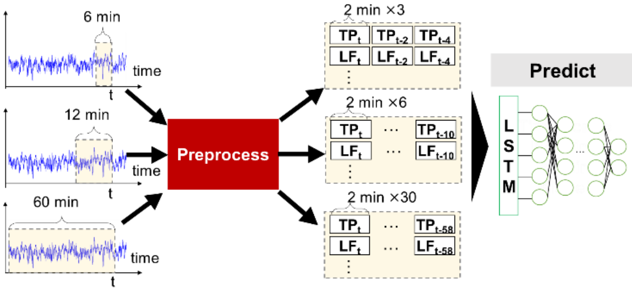

FIGURE 3. Generation of time-series datasets in different time ranges. The ANF indices are calculated from 6–60 min RRI data at a 2-min time width; thus, the dataset length is 3–30.

30 in length, respectively. Table 7 shows the total amount of data for each input time width. Only datasets not containing missing data in each time width are counted here. The time range for the future prediction was fixed at 30 min.

TABLE 7. Number of data samples for each input time range (without missing dataset).

Input time width	Without near misses	With near misses
6 min	40419	13061
12 min	26952	9072
18 min	18558	6180
24 min	13080	4216
30 min	9553	2884
36 min	7069	1993
42 min	5224	1417
48 min	3894	1072
54 min	2944	819
60 min	2235	620

TABLE 8. Comparison of time-series forecasting accuracy for different time ranges of input data.

Input data	Input time width	recall	AUC
ANFs + EXTs (without missing)	6 min	60.9%	0.69
	12 min	56.0%	0.69
	18 min	58.4%	0.71
	24 min	56.0%	0.71
	30 min	**74.9%**	**0.79**
	36 min	52.7%	0.69
	42 min	39.2%	0.64
	48 min	59.0%	0.71
	54 min	72.7%	0.65
	60 min	59.4%	0.74

Table 8 shows the application results of the proposed method for each input time range. The highest accuracy was obtained for the input time width of 30 min for both recall and the AUC. These results can be interpreted from the following points. In general, prediction accuracy is affected by the training data size of the model. The word "data size" has two meanings here: the number of data samples and the time length of a single piece of data. Because we used a dataset without missing, the number of samples tends to be larger for shorter time widths. As a result, the dataset with a shorter time width is considered to be successful in learning different ANF patterns, leading to an improvement in the generalizability of the model. Conversely, longer input data sizes have richer time-domain features and are considered to be more effective for future prediction. Therefore, our results suggest that, in our condition, the 30 min time width provided the best compromise. Finally, we decided to set the input time range to 30 min.

ACKNOWLEDGMENT

The authors wish to acknowledge Dr. Hirohiko Kuratsune, Professor with the School of Medicine, Osaka Metropolitan University, for helpful comments.

REFERENCES

[1] P. Marcillo, Á. L. V. Caraguay, and M. Hernández-Álvarez, "A systematic literature review of learning-based traffic accident prediction models based on heterogeneous sources," *Appl. Sci.*, vol. 12, no. 9, p. 4529, Apr. 2022, doi: 10.3390/APP12094529.

[2] M. Effati, M. A. Rajabi, F. Hakimpour, and S. Shabani, "Prediction of crash severity on two-lane, two-way roads based on fuzzy classification and regression tree using geospatial analysis," *J. Comput. Civ. Eng.*, vol. 29, no. 6, Sep. 2014, Art. no. 04014099, doi: 10.1061/(ASCE)CP.1943-5487.0000432.

[3] M. Effati and A. Sadeghi-Niaraki, "A semantic-based classification and regression tree approach for modelling complex spatial rules in motor vehicle crashes domain," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 5, no. 4, pp. 181–194, Jul. 2015, doi: 10.1002/WIDM.1152.

[4] B. Sharma, V. K. Katiyar, and K. Kumar, "Traffic accident prediction model using support vector machines with Gaussian kernel," in *Proc. 4th Int. Conf. Soft Comput. Problem Solving* (Advances in Intelligent Systems and Computing), vol. 437, Apr. 2016, pp. 1–10, doi: 10.1007/978-981-10-0451-3_1.

[5] Y. Li, F. Wang, H. Ke, L.-L. Wang, and C.-C. Xu, "A driver's physiology sensor-based driving risk prediction method for lane-changing process using hidden Markov model," *Sensors*, vol. 19, no. 12, p. 2670, Jun. 2019, doi: 10.3390/s19122670.

[6] D. Fruet, C. Barà, R. Pernice, L. Faes, and G. Nollo, "Assessment of driving stress through SVM and KNN classifiers on multi-domain physiological data," in *Proc. IEEE 21st Medit. Electrotechnical Conf. (MELECON)*, Jun. 2022, pp. 920–925, doi: 10.1109/MELECON53508.2022.9842891.

[7] Hitachi Transport System. *Smart and Dafety Connected Vehicle System SSCV*. Accessed: Mar. 2, 2021. [Online]. Available: https://www.hitachi-transportsystem.com/jp/sscv/pdf/sscv.pdf

[8] K. Mizuno, D. Ojiro, T. Tanaka, S. Minusa, H. Kuriyama, E. Yamano, H. Kuratsune, and Y. Watanabe, "Relationship between truck driver fatigue and rear-end collision risk," *PLoS ONE*, vol. 15, no. 9, Sep. 2020, Art. no. e0238738, doi: 10.1371/JOURNAL.PONE.0238738.

[9] S. Minusa, K. Mizuno, D. Ojiro, T. Tanaka, H. Kuriyama, E. Yamano, H. Kuratsune, and Y. Watanabe, "Increase in rear-end collision risk by acute stress-induced fatigue in on-road truck driving," *PLoS ONE*, vol. 16, no. 10, Oct. 2021, Art. no. e0258892, doi: 10.1371/journal.pone.0258892.

[10] N. Ito, T. Tanaka, S. Minusa, and H. Kuriyama, "Accident risk time series prediction based on autonomic nervous function while driving," in *Proc. EMBC*, 2021, p. 5174.

[11] D. Kos, J. Raeymaekers, A. Van Remoortel, M. D'hooghe, G. Nagels, M. D'Haeseleer, E. Peeters, T. Dams, and T. Peeters, "Electronic visual analogue scales for pain, fatigue, anxiety and quality of life in people with multiple sclerosis using smartphone and tablet: A reliability and feasibility study," *Clin. Rehabil.*, vol. 31, no. 9, pp. 1215–1225, Sep. 2017, doi: 10.1177/0269215517692641.

[12] K. Fujiwara, E. Abe, K. Kamata, C. Nakayama, Y. Suzuki, T. Yamakawa, T. Hiraoka, M. Kano, Y. Sumi, F. Masuda, M. Matsuo, and H. Kadotani, "Heart rate variability-based driver drowsiness detection and its validation with EEG," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1769–1778, Jun. 2019, doi: 10.1109/TBME.2018.2879346.

[13] N. Munla, M. Khalil, A. Shahin, and A. Mourad, "Driver stress level detection using HRV analysis," in *Proc. ICABME*, 2015, pp. 61–64, doi: 10.1109/ICABME.2015.7323251.

[14] R. Sen, H.-F. Yu, and I. Dhillon, "Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting," in *Proc. NIPS*, 2019, pp. 1–10.

[15] C. Nakagawa, A. Suzuki, F. Kikuchi, T. Watanabe, N. Akiu, and T. Kojima, "Estimation of strong tension state using multiple physiological indices optimized for each individual," *Jpn. J. Ergonom.*, vol. 58, no. 2, pp. 84–95, 2022.

[16] N. Tamura, K. Shimazu, M. Hienuki, K. Oh-iwa, H. Kim, and K. Hamaguchi, "Effects of physiological ageing on autonomic reflexes," *Nippon Ronen Igakkai Zasshi. Jpn. J. Geriatrics*, vol. 19, no. 6, pp. 563–570, 1982, doi: 10.3143/GERIATRICS.19.563.

[17] Ministry of Health Labour and Welfare. (Apr. 2021). *Project on Medical Investigation of Fatigue Levels of Motor Vehicle Drivers*. Accessed: Apr. 25, 2022. [Online]. Available: https://www.mhlw.go.jp/content/11601000/000773658.pdf

[18] *Statistical Analysis of Traffic Accidents in January-June 2009—by Place of Occurrence*, Japan Trucking Association, 2017.

[19] S. Hochreiter and J. Schmidhuber. (1996). *LSTM Can Solve Hard Long Time Lag Problems*. Accessed: Mar. 1, 2021. [Online]. Available: https://proceedings.neurips.cc/paper/1996/hash/a4d2f0d23dcc84ce983ff9157f8b7f88-Abstract.html

[20] N. Okawa, D. Kuratsune, J. Koizumi, K. Mizuno, Y. Kataoka, and H. Kuratsune, "Application of autonomic nervous function evaluation to job stress screening," *Heliyon*, vol. 5, no. 2, Feb. 2019, Art. no. e01194, doi: 10.1016/j.heliyon.2019.e01194.

[21] M. Malik, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996, doi: 10.1161/01.cir.93.5.1043.

[22] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers Public Health*, vol. 5, p. 258, Sep. 2017, doi: 10.3389/fpubh.2017.00258.

[23] K. Yokoyama and I. Takahashi, "Feasibility study on estimating subjective fatigue from heart rate time series," *Trans. Inst. Electron. Inf. Commun. Eng. A*, vol. 96, no. 11, pp. 756–762, 2013.

NAO ITO was born in Japan. She received the bachelor's and master's degrees in engineering from Chiba University, Chiba, Japan, in 2017 and 2019, respectively.

Since 2019, she has been a Researcher with the Research and Development Group, Hitachi Ltd. Her current research interests include machine learning and physiological data analysis.

Ms. Ito is a member of the Japan Ergonomics Society.

SHUNSUKE MINUSA was born in Japan. He received the B.E. and master's degrees in information science and technology from Hokkaido University, Sapporo, Japan, in 2016 and 2018, respectively.

Since 2018, he has been a Researcher with the Research and Development Group, Hitachi Ltd. His current research interests include human sensing, biostatistics, and psychophysiological data analysis for healthcare applications.

Mr. Minusa is a member of the Behaviormetric Society of Japan and the Japan Ergonomics Society.

TAKESHI TANAKA was born in Japan. He received the bachelor's and master's degrees in electronics and information engineering from Hokkaido University, Sapporo, Japan, in 2002 and 2005, respectively.

Since 2005, he has been a Researcher with Hitachi Ltd.

Mr. Tanaka is a member of the Information Processing Society of Japan.

YUN LI was born in China. She received the bachelor's and master's degrees in intelligent and mechanical interaction systems from the University of Tsukuba, Tsukuba, Japan, in 2013 and 2015, respectively.

Since 2015, she has been a Researcher with the Research and Development Group, Hitachi Ltd.

HIROYUKI KURIYAMA was born in Japan. He received the master's degree from the Department of Biomedical Control, Hokkaido University, Sapporo, Japan, in 1990.

Since 1990, he has been a Researcher with the Research and Development Group, Hitachi Ltd.

Mr. Kuriyama is a member of the Japanese Society of Fatigue Science.