

Received 18 November 2022, accepted 4 December 2022, date of publication 12 December 2022, date of current version 22 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3228573

RESEARCH ARTICLE

Temporal Early Exiting With Confidence Calibration for Driver Identification Based on Driving Sensing Data

JAEBONG LIM[®]¹, YUNJU BAEK[®]¹, AND BUMHEE CHAE[®]²

¹School of Computer Science and Engineering, Pusan National University, Busan 46241, South Korea
²Suremobility Inc., Seoul 06179, South Korea

Corresponding author: Yunju Baek (yunju@pusan.ac.kr)

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: RS-2020-KD000201).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the OPEL Solution Inc.as part of the Data Voucher Project of the South Korean Government.

ABSTRACT Driver identification systems that use deep-neural-network-based sequential models have been studied for personalized intelligent vehicles. After a vehicle starts moving for a trip, the system identifies the driver at each time step using accumulated driving sensing data. We propose a novel driver identification system with temporal early exiting to identify a driver as early as possible while maintaining accuracy. Existing systems require entire-trip data or fixed-length partial trip data, regardless of driver identification difficulty. The proposed system automatically identifies the driver with less driving data for easy-to-identify trips and more driving data for hard-to-identify trips. To adaptively exit the identification by considering the difficulty of a trip, we propose a temporal early-exiting method by thresholding the confidence score. Sequential models output an identified driver and confidence score at each time step. However, the confidence score of deep neural networks is unreliable owing to the overconfidence problem. To overcome this problem, we propose three temporal confidence calibration methods that adjust the calibration strength according to the driving time and trip difficulty. Thus, the system can determine the best time to exit the identification, considering the trade-off between latency and accuracy. Our experiments on a naturalistic driving dataset show that the proposed system achieved 90.06% accuracy with early exiting at an average of 6.7 min, yielding the same accuracy with 74.2% latency reduction compared with driver identification with 26 min of fixed-length data for each trip.

INDEX TERMS Confidence calibration, driver identification, driving sensing data, in-vehicle sensors, label smoothing, temporal early exiting, temperature scaling.

I. INTRODUCTION

Intelligent vehicles (IVs), which have evolved based on computer and sensor technologies, have become essential in the field of intelligent transportation systems [1]. IVs provide safety and comfort to drivers by using driving sensing data from pedal pressure, vehicle motion, engine airflow, and environmental sensors. To provide a safer and more comfortable driving experience, IVs should consider the needs of different drivers through driving style features. Previous studies on driver identification have discussed how driving style features in advanced driver assistance systems and autonomous vehicles can improve driver experience [2]. Through driver identification, IVs can accommodate driving styles of various drivers. For example, Hyundai Inc. provides automatic drive-mode shifts according to driving styles,

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

and Tesla Inc. has released self-driving software with personalized driving styles. In a shared mobility context, driver identification helps car-sharing providers verify authorized drivers by tracking their driving styles and provide personalized rides for their customers [3].

Driver identification refers to determining the driver behind the driving data based on how they control the vehicle and react to different driving conditions [4]. Each driver has a different driving style, even under the same driving conditions. Machine learning approaches using deep neural networks (DNNs) are commonly used to learn the driving style features of each driver. A DNN identifies the driver using sliding-window driving data at each time step. Many previous studies have adopted the convolutional neural network (CNN) architecture to predict the confidence scores of drivers using a sliding window [5], [6], [7]. These studies independently identified the driver using a few seconds or minutes of data, regardless of the driving progression of a trip.

Recently, studies investigating the tendency of an entire or partial driving trip by aggregating the outcomes with a sliding window have attracted attention in driver identification. Because there are variations in the driving style of a driver, understanding the driving style of a trip rather than a short sliding window is required. For example, a driver's driving style may occasionally change depending on abnormal situations or emotions [8]. Related studies used deep sequential models such as recurrent neural networks (RNN) to aggregate the outcomes for an entire or partial trip [9], [10], [11], [12]. At each time step, the RNN outputs the driver confidence scores using the accumulated data from the beginning of the trip to the assessment time.

However, driver identification systems that rely on minimal driving are required. The vehicle must identify the driver as early as possible to switch to the driver's preferences [13]. After the vehicle starts moving, the system identifies the driver at each time step using the accumulated data. The final decision and identification exit can be made at any point during a trip. If the exit is made early, with relatively little trip data, the identification result has good latency but low accuracy. If the exit is made late, with considerable trip data, the result has good accuracy but a poor latency. Therefore, early exit and maintaining accuracy are crucial issues in driver identification systems.

It is challenging to decide when to exit the driver identification process because the best time may vary depending on the course and situation of a trip. For example, a driver may manipulate the brake frequently because of a traffic jam, but the same driver may brake little during a different trip. Therefore, it is difficult to identify the driver during some trips, whereas during others, it is simple [14]. Nevertheless, the disadvantage of existing systems is that they require entire-trip data or fixed-length partial trip data, regardless of driver identification difficulty. Consequently, a method to exit the identification adaptively is required to reduce latency while maintaining accuracy.

A simple and effective way to exit the identification adaptively is to use a confidence score threshold. Deep sequential models identify the driver and output a confidence score for each time step. A confidence score estimates the predicted probability that represents the true likelihood [15]. The confidence score is essential not only for model interpretability but also for early exiting. Despite the success of temporal early exiting with a confidence score in video recognition [16], [17], [18], no such studies have been conducted on driver identification. The challenge is that the confidence score is unreliable owing to the overconfidence problem, particularly in sequential sensing data [19]. Confidence calibration has been proposed as an effective solution to this problem [20]. Nevertheless, most previous works aimed at nonsequential data and did not consider early exiting.

In this study, we propose a novel driver identification system with a temporal early-exiting method to identify a driver as early as possible while maintaining accuracy. The proposed system provides not only the identification result for each time step but also the time to exit the identification with the final decision during driving. By using temporal early exiting with a combination of CNN and RNN, the system can determine the tendency of sequential driving data for real driving scenarios. To exit the identification adaptively by considering the difficulty of a trip, we propose three types of temporal confidence calibration methods for temporal early exiting. The proposed calibration methods automatically train and calibrate a model to identify the driver with less driving data for easy-to-identify trips and more driving data for hard-to-identify trips. By adjusting the calibration strength according to the driving time and trip difficulty, we improved the trade-off between latency and accuracy.

We aimed to identify drivers as early as possible while maintaining accuracy. The main contributions of this study can be summarized as follows:

1) We study temporal early exiting for driver identification to reduce latency while maintaining accuracy using naturalistic driving data.

2) We propose three types of temporal confidence calibration for early exiting to improve the trade-off between accuracy and latency.

3) We propose a novel hybrid confidence calibration to emphasize the consideration of trip difficulty.

The remainder of this paper is organized as follows. In Section II, we provide a review of related works on driver identification systems, early-exiting methods, and confidence calibration. In Section III, we explain our proposed driver identification system with temporal early exiting, and in Section IV, we propose three types of temporal confidence calibration methods for early exiting. In Section V, the experimental results are presented and discussed using collected and public datasets. Finally, we discuss our conclusions and plans for future work in Section VI.

II. RELATED WORK

A. DRIVER IDENTIFICATION

Driver identification using driving sensing data has been actively studied. The driver can be identified using a machine learning algorithm using sliding-window data as input. Wakita et al. [5] proposed a driver identification system based on a Gaussian mixture model with 6 s window data. Public driving sensing datasets extracted from vehicle on-board diagnostics (OBD) ports have become popular, and much research has been conducted on them using CNNs and RNNs to process the window data. Chen et al. [6] used an input with a 3 s window and 2 s overlap. Xun et al. [7] proposed a CNN using a 30 s window with a 29 s overlap. These studies independently identified drivers for each short window, regardless of their driving progress during a trip. They reported good accuracy because existing public datasets have many restrictions, such as few drivers and fixed routes. However, recent studies using large natural datasets without restrictions have shown inferior accuracy for driver identification using short-window data [9], [21], [22].

Research attempting to understand the driving style of an entire or partial trip by aggregating the results of sliding windows is attracting attention in the field of driver identification. Voting is a straightforward approach that aggregates the outcomes for each window in a trip. Previous studies identified the driver for each window and then selected a driver for a trip by voting on the outcomes of each window [21], [22]. The combination of CNN and RNN is commonly used for developing driver identification systems because of their superior accuracy with sequential data [9], [10], [11], [12]. These works used a CNN to extract features from window data and RNN to identify the driver at each time step using accumulated features from the beginning of the trip. Therefore, the RNN prediction at each time step utilizes not only the current observation but also previous observations, which provide context for the progression of the sequence.

Existing driver identification systems have disadvantages in that they require entire-trip or fixed-length partial trip data, regardless of driver identification difficulty. Dong et al. [9] proposed a method that processed entire-trip data with a CNN-RNN to improve accuracy. Similarly, Zhang et al. [10] presented a driver identification system based on CNN-RNN using entire-trip data. They extracted features via a sliding window using a CNN and collected them using an RNN. Ren et al. [11] proposed a deep learning framework that combines the Siamese architecture and CNN-RNN to process entire-trip data. To date, there is no existing work on the optimal time to exit driver identification during a trip. Therefore, temporal early-exiting studies on driver identification are required.

B. TEMPORAL EARLY EXITING AND CONFIDENCE CALIBRATION

Early exiting, a method of reducing the latency of a DNN, can be conducted layer-wise and temporally. Layer-wise

early exiting reduces the computational load by exiting the inference early via the exit branches in the neural network architecture. If an intermediate exit branch is sufficiently confident, the model short-circuits and halts the inference at that layer [23]. Temporal early exiting has been actively studied for deep sequential models such as RNNs. An RNN with temporal early exiting adaptively exits recognition after observing only a fraction of the sequence. Temporal early exiting, which helps determine when to exit such that accuracy can be maintained while reducing latency, is widely used in video recognition. Moreover, Ghodrati et al. [24] proposed a method for determining when to exit by using a gated unit.

Research on determining when to exit using confidence scores has been actively conducted in temporal early exiting. Wang et al. [16] determined when to exit by using a confidence score threshold in online video recognition. This approach can score partial events by monitoring the degree of event completion, as it monotonically increases toward termination. Tang et al. [17] proposed an early-exiting system when the confidence score for streaming speech command recognition is sufficiently high. Because it is difficult to trust these confidence scores, Ma et al. [18] proposed a ranking loss to gradually increase the confidence score.

Even though most modern DNNs natively produce an estimated confidence score over class labels for a given instance, the scores do not always closely reflect the true probabilities of each class. The confidence score is unreliable owing to the overconfidence problem, particularly in sequential sensing data [19]. Confidence calibration has been proposed to solve the overconfidence problem effectively. Temperature scaling [15] and label smoothing [27] are widely used confidence calibration methods. Temperature scaling prevents overconfidence by leveling the confidence per label of the learned DNN. Label smoothing prevents overconfidence by averaging the correct answer values during the DNN training process. However, a limitation of these calibration methods is that they do not consider sequential models.

To overcome this limitation, recently, studies have been on temporal confidence calibration [28], [29]. Leathart and Polaczuk [28] applied different temperatures according to the data sequence length for natural language. Shen et al. [29] proposed class-specific errors to adjust the smoothing strength by considering the sequence context. These studies aimed to alleviate overconfidence; however, to date, there is no temporal confidence calibration study for early exiting.

III. PROPOSED DRIVER IDENTIFICATION SYSTEM

In this paper, we propose a novel driver identification system using minimum driving data while maintaining accuracy through temporal early exiting. The proposed system consists of two parts: a driver identification system with temporal early exiting and three types of temporal confidence calibration for temporal early exiting. The proposed driver identification system is characterized by temporal early



FIGURE 1. Overview of the proposed driver identification system.

exiting during a driving trip. The proposed system identifies the driver when the confidence score exceeds a threshold, and three types of confidence calibration methods for temporal early exiting are introduced to improve the trade-off between latency and accuracy.

Fig. 1 shows an overview of the proposed system, which consists of three components: driving data preparation, driver identification, and temporal early exiting. In the data preparation stage, the system obtains driving data from vehicular sensors during the trip. Moreover, it preprocesses the data to normalize them and composes slide-window data. In the driver identification stage, the deep learning model identifies the driver at each time step using the preprocessed data. Finally, in the temporal early-exiting stage, the final result of the trip is determined when the confidence score output exceeds the threshold value.

We used a standard preprocessing method and a wellknown CNN-RNN structure for the proposed system to provide an example of a driver identification system with temporal early exiting for quick application to any other driver identification system. Similar to existing studies, the driver is identified from among specified drivers. The proposed system classifies a set of drivers as follows. The sequential classification task assigns a sequential label vector $Y = \{y_1, y_2, \dots, y_t, \dots, y_T\}$ to input $X = \{x_1, x_2, \dots, y_t, \dots, y_T\}$:

$$f: X_{1 \sim t} \to y_t = \{y_t^1, y_t^2, \dots, y_t^k, \dots, y_t^K\}.$$
 (1)

In the driver identification system, X represents time-series driving sensing data, and Y is the sequence of a driver identified at each time step; y_t is one-hot encoded ground truth at time step t for K drivers. Thus, y_t^k is the binary truth for the k-th driver.

A. DRIVING DATA PREPARATION

During driving data preparation, the proposed system acquires driving data from vehicular sensors and preprocesses the acquired driving data. The driving data are multidimensional sequential data consisting of various sensing data, such as velocity and pedal pressure. The system applies the normalization and sliding-window techniques in the preprocessing step. Driving data are prepared sequentially to pass to the CNN-RNN for driver identification at each time step.

To mitigate the effects of different sensor scales, the system normalizes the driving data before using them for driver identification. Through normalization, the sensing data are treated equally by the CNN-RNN. The normalization equation is shown in (2):

$$\hat{X}_{t}^{i} = \frac{X_{t}^{i} - \min X_{t}^{i}}{\max X_{t}^{i} - \min X_{t}^{i}},$$
(2)

where min (X_t^i) and max (X_t^i) are the minimum and maximum values of feature column X_t^i , respectively. \hat{X}_t^i is the normalized variable. Normalization transforms each sensor data point of the driving data to a range between 0 and 1.

Before being input in the CNN-RNN, the normalized data are processed using a sliding window. Algorithms with sliding windows are widely used for processing sequential data. Driving data are continuous; therefore, the sliding-window technique is adopted to divide the entire-trip driving data into multiple discrete data windows by period. The sliding-window technique extracts the window data with a certain overlapping ratio.

B. DRIVER IDENTIFICATION

In the driver identification stage, the proposed system sequentially processes the prepared driving data and identifies the driver at each time step. Sliding-window data, which are a fraction of the sequential driving data, are passed into the CNN, which in turn extracts the date features from each sliding window. The features accumulated from the beginning of the trip to the present are input into the RNN. At each time step, the RNN outputs the driver confidence scores.

The driver confidence scores at each time step result from the SoftMax activation. The last layer of the CNN-RNN outputs the driver confidence score as a real value ranging between 0 and 1 with the SoftMax activation presented in (3):

$$p_t^k = \frac{e^{z_t^k}}{\sum_{k=1}^K e^{z_t^k}}$$
(3)

where p_t^k represents the confidence score of driver k at time step t. K is the number of drivers to classify, and z_t^k represents the last layer outputs of the CNN-RNN before the SoftMax activation. The CNN-RNN model is trained with cross-entropy loss to find the best model parameters θ as follows:

$$\underset{\theta}{\operatorname{argmin}} \sum_{k=1}^{K} -y_t^k \log\left(p_t^k\right). \tag{4}$$

The detailed architecture of the CNN-RNN is as follows. The CNN consists of six layers: five convolutional layers and an average pooling layer. Each convolutional layer has a 3×3 filter and ReLU activation. The second and fourth convolutional layers have a 3×1 stride. The numbers of channels in the convolutional layers are 64, 128, 128, 256, and 256. An average pooling layer converts the 2D feature output of the last convolutional layer into a 1D feature to be passed to the RNN. The size of the 1D feature is 256. The CNN has 1.1M parameters.

The RNN consists of four layers: two long short-term memory (LSTM) layers and two fully connected layers. Each LSTM layer has 256 channels, sigmoid activation for recurrent units, and tanh activation for output. The number of channels in the first fully connected layer is 128. The number of channels in the last fully connected layer equals the number of drivers that require identification. Therefore, the last fully connected layer as a real value ranging between 0 and 1 with SoftMax activation. The RNN has 1.2M parameters. Consequently, the total size of the CNN-RNN architecture is 2.3M parameters.

C. TEMPORAL EARLY EXITING

Driver identification based on SoftMax activation assigns a confidence score to each driver. We set a threshold *TH* for early exiting using the confidence score for each driver as follows:

$$\max_{1 \le k \le k} p_t^k \ge TH,\tag{5}$$

where p_t^k represents the confidence score of driver k at time step t. As shown in (5), the proposed system exits the processing and identifies the most likely driver once the network surpasses this confidence threshold. This model adaptively exits per sequential driving data. The threshold *TH* controls the timing of early exiting with a certain tradeoff between accuracy and latency. We aim to improve this trade-off between accuracy and latency for all *TH* ranges.

IV. PROPOSED CONFIDENCE CALIBRATION

In this paper, we propose three temporal confidence calibration methods, which adjust the calibration strength over time. The proposed methods aim to reduce the latency to exit identification during a driving trip and to maintain identification accuracy. Therefore, the objectives of the proposed method are as follows:

 $\begin{array}{l} \operatorname{argmin}_{\text{calibraion}} \operatorname{AVG_Latency}_{\text{exiting}}(\mathcal{D}_{\text{val}}) \\ \operatorname{subject to} \operatorname{ACC}_{\text{exiting}}(\mathcal{D}_{\text{val}}) = \operatorname{ACC}_{\text{last}}(\mathcal{D}_{\text{val}}). \end{array}$ (6)

Here, \mathcal{D}_{val} is the validation set, which means that the proposed calibration method finds the best method for the validation set. AVG_Latency_{exiting}(\mathcal{D}_{val}) represents the average early-exiting time for each trip on the validation set. ACC_{exiting}(\mathcal{D}_{val}) and ACC_{last}(\mathcal{D}_{val}) are the accuracies with partial-trip data before early-exiting and entire-trip data, respectively. ACC_{exiting}(\mathcal{D}_{val}) and ACC_{last}(\mathcal{D}_{val}) are equal if the system with early-exiting maintains its accuracy compared to the system without early-exiting. Therefore, the proposed calibration methods aim to find the best way to minimize AVG_Latency_{exiting}(\mathcal{D}_{val}) when ACC_{exiting}(\mathcal{D}_{val}) and ACC_{last}(\mathcal{D}_{val}) are equal.

The problem of overconfidence occurs in the early stage when using little driving data. It is especially a problem on trips in which driver identification is difficult. To overcome this problem, calibration methods should be improved to adjust the calibration strength according to driving time. The calibration strength should be increased at the beginning of driving and decreased during the latter part of driving. Therefore, the proposed temporal calibration methods adjust the calibration strength according to driving time and trip difficulty.

The proposed temporal temperature scaling and label smoothing are advanced methods that modify existing standard calibration methods to control the calibration strength over driving time. The proposed hybrid calibration method is characterized by its calibration strength adjustment according to driver identification difficulty, i.e., by decreasing the strength for easy-to-identify trips and increasing it for hard-to-identify trips.

A. TEMPORAL TEMPERATURE SCALING

Temperature scaling has been widely used as a simple and effective calibration method to mitigate the overconfidence problem of modern DNNs. Temperature scaling is a post-hoc calibration strength adjustment method via a temperature parameter for DNN output. The temperature parameters cannot be adjusted during the training period. The DNN learns to decrease the temperature as much as possible so that it can be very confident in the training examples [28]. Therefore, a temperature parameter should be obtained after training the DNN. Temperature scaling multiplies the logits by the selected temperature, which is a scalar, before applying the SoftMax:

$$\hat{p}_{t}^{k} = \frac{e^{z_{t}^{k}/T}}{\sum_{k=1}^{K} e^{z_{t}^{k}/T}},$$
(7)

where \hat{p}_t^k and T are the calibrated confidence score of p_t^k and the calibration strength for SoftMax activation, respectively.

Although temperature scaling is an effective calibration method, applying the same temperature on the output at each time step is ineffective for sequential model calibration because the sequence data context varies over time [29]. Accordingly, we modified the standard temperature scaling to control the calibration strength over driving time. We control



FIGURE 2. Temporal temperature scaling examples.

T for temperature scaling over driving time t as follows:

$$T = \left(1 - \frac{1}{L}t\right)^{\beta} (\alpha - 1) + 1.$$
(8)

Here, the best temperature T, which is a function of driving time t not a scalar, is selected via grid search among predefined α and β sets. α and β should be greater than 0. Grid search aims to enhance the performance of temporal early-exiting according to (6). α controls the initial temperature, and β controls the strength reduction. Fig. 2 and 3 show examples of temperature T according to α and β .

B. TEMPORAL LABEL SMOOTHING

Recently, label smoothing has become a popular choice for DNN calibration. One-hot encoding, in which all data are labeled with 0 or 1, is prone to make the model overconfident for prediction. Label smoothing is applied during training and focuses on increasing the entropy of the predicted values, thereby reducing overconfident events. It smooths the one-hot distribution with a hyperparameter ϵ to obtain a soft distribution for every y_t :

$$\hat{y}_t^k = y_t^k \left(1 - \epsilon\right) + \epsilon/K,\tag{9}$$

where y_t^k and \hat{y}_t^k are the one-hot encoded ground truth and label smoothed ground truth, respectively.

Although label smoothing has proven to be helpful in mitigating the overconfidence problem by smoothing confidence in a soft distribution, using the same smoothing strength on the label of all sequences is ineffective for calibrating sequence data [29]. We modified the standard label smoothing to control the calibration strength over driving time. Therefore, the proposed temporal label smoothing controls ϵ over driving time *t* as follows:

$$\epsilon = \alpha \left(1 - \frac{1}{L}t \right)^{\beta},\tag{10}$$

where α controls the initial smoothing strength, and β leads to an exponential reduction in confidence. We select the best ϵ value with the validation set via grid search according to (6) among predefined α and β sets. β should be greater than 0, and α should be between 0 and 1.



FIGURE 3. Overview of the proposed hybrid confidence calibration.

C. HYBRID CONFIDENCE CALIBRATION

The proposed hybrid calibration determines a suitable calibration strength for each trip because applying a high calibration strength for easy-to-identify trips increases latency without a gain in accuracy. Therefore, the proposed hybrid calibration increases the calibration strength for hard-toidentify trips and decreases it for easy-to-identify trips. The main idea behind the proposed hybrid calibration is to find a smoothed label for each trip via post-hoc temperature scaling and then apply label smoothing with the acquired smoothed label. Via temperature scaling, we can obtain the temperature for each trip in the validation set but cannot obtain the temperature for a new trip. However, label smoothing can apply calibration implicitly for a new trip but cannot obtain the calibration strength for each trip. Therefore, we used temperature scaling to find an appropriate strength for each trip and applied label smoothing with this strength while training the model.

Fig. 3 presents an overview of the proposed hybrid confidence calibration. The proposed method comprises four components: temperature acquisition, smoothed label conversion, training, and rescaling. In the temperature acquisition phase, we applied temperature scaling trip by trip. All trip data match one optimal temperature to obtain the minimum latency while maintaining accuracy for early exiting. Using temperature scaling, we can obtain the temperature for each trip in the validation set. In the smoothed label conversion phase, smoothed labels suitable for each trip data point were obtained using the temperature acquired in the previous phase. The result of applying temperature to the confidence of each trip data point was used as a smoothed label. In the next step, the model was trained using the obtained label as the standard label-smoothing method. Finally, the temporal temperature was applied again. Although it has been found that temperature scaling and label smoothing can be used as calibration techniques and have similar effects for a DNN, the relationship between temperature scaling and label smoothing is an open research topic in the field of machine learning theory. Therefore, in this study, the scale was adjusted by applying temperature scaling again in the last phase when combining the two calibration methods.

Algorithm 1 shows the procedure for temperature acquisition and smoothed label conversion. The temperature

Algorithm 1 Temperature Acquisition and Smoothed Label		
Conversion		
Input: Train set (X, Y)		
Output: Train set with the smoothed labels (X', Y')		
01: Split train set to <i>P</i> folds		
02: For <i>p</i> in 1: <i>P</i> do		
03: Select <i>p</i> fold for validation fold		
Select remain folds for train folds		
Train a model <i>M</i> using train folds		
6: For <i>d</i> in validation fold do		
7: For α , β in a predefined α , β set do		
Get temperature T by α , β		
If <i>T</i> is the best on model <i>M</i> according		
to (6)		
10: Best temperature $\hat{T} \leftarrow T$		
11: End if		
12: End for		
13: Get confidence of d		
14: Apply the best temperature \hat{T} to confidence		
: Use the scaled confidence as smoothed label		
16: End for		
17: End for		

acquisition uses only the training set. We split the training set into P folds and obtained the temperature for each trip in a cross-validation manner. Each fold became a validation fold, and the remaining folds became a training fold. We performed a grid search among predefined α and β sets to find temperature T for each trip on a validation fold. The best \hat{T} value, which resulted in the minimum latency while maintaining accuracy for early exiting, was selected. Then, the smoothed label was obtained by determining the confidence for each trip data and applying temperature scaling to the confidence with the best temperature \hat{T} .

V. EXPERIMENTS

In this section, we evaluate the effectiveness of temporal early exiting with the three types of temporal confidence calibration for the proposed driver identification system. To evaluate the proposed driver identification approach, we experimented with a collected dataset and an open public dataset. We compared the existing driver identification system using fixed-length driving data and the proposed system with temporal early exiting. Moreover, we analyzed the effects of three types of temporal confidence calibration methods. In addition, we explored hyperparameters for temporal confidence calibration. Finally, we evaluated the proposed system using an open public dataset.

A. DATA COLLECTION

Data were collected from different types of real vehicles to verify the effectiveness of the proposed system. Four sets of sensing data from the vehicle OBD port, namely vehicle speed, revolutions per minute, steering wheel angle, and brake pedal pressure, were recorded every 1 s during driving.



FIGURE 4. OBD data collector distributed for data collection.



FIGURE 5. Example trips of the collected dataset.

Data were collected using an OBD data collector designed by OPEL Solution Inc. (Fig. 4), which was distributed to the public with consent to collect data using crowdsourcing. The data were collected and anonymized by OPEL Solution Inc. as part of the Data Voucher Project of the South Korean government. All trip data were collected in South Korea using various routes without any restrictions, thus creating a naturalistic dataset.

Details of the collected dataset are as follows. After excluding the data of drivers with insufficient trips, the number of subject drivers was 55. Each remaining driver drove ten trips or more. Example trips in Fig. 5 show the driving route of the vehicles during data collection. Most trips differ in terms of the route. The dataset contained 1941 trip data points for a total of 2181 hours. Each trip was driven for a minimum of 30 min, an average of 1 h, and a maximum of 5.5 h.

B. EXPERIMENTAL SETTINGS

We evaluated the performance of the proposed system over the collected natural dataset. To evaluate the efficiency of the proposed system, we compared its identification results with those of various confidence calibration methods. The experimental settings included the input data settings, dataset separation, and metrics.

For all experiments, the window size was 20 s with 10 s overlapping. The system identified the driver every 10 s using accumulated driving sensing data from the beginning to the present time until early exiting occurred. The sampling rate of the four types of sensing data was 1 Hz. Therefore, the



FIGURE 6. Early exiting results according to calibration.

prepared input data for the DNN with a 20 s window through the four sensors represented 80 real number data points.

We divided the dataset using an 8:1:1 ratio for training, validation, and testing, respectively, such that all trips in the dataset were divided into training, validation, and test sets. Although the number of trips for each driver was imbalanced, the validation and test sets had at least one trip by each driver. The CNN-RNN model for driver identification was trained only with the training set. The model was trained with the Adam optimizer, where the number of epochs was 100. The CNN was pretrained with the sliding-window data. Confidence calibration used the training and validation sets. The test set was only used for performance evaluation after the model was trained.

Most existing studies evaluated the accuracy of window data but not trips. However, in a realistic scenario, it is important to identify the driver accurately and quickly during the trip. Therefore, we used trip accuracy as the evaluation metric, similar to some existing studies based on per-trip identification. The trip accuracy calculates the proportion of driving trips that belong to the target driver among all trips in the dataset, which can be denoted as

$$Trip Accuracy = \frac{the number of true positive trip}{the total number of trip} (11)$$

The calibration parameters were determined via grid search. During temporal temperature scaling, the search space of the grid search was such that α and β had integer values ranging from 1 to 10. For temporal label smoothing, the search space of the grid search was such that β has an integer value from 1 to 10, and α was a real number from 0 to 1 with a step of 0.1. For hybrid calibration, the search space of the grid search was the same as the temporal temperature scaling, and the number of folds was 8 except for the hyperparameter experiments.

C. EXPERIMENTAL RESULTS

1) TEMPORAL EARLY EXITING RESULTS

Fig. 6 summarizes the results for the system with early exiting according to the calibration methods. As demonstrated, the proposed system after the three types of temporal calibration is highly improved compared to the system without



FIGURE 7. Results without early exiting: accuracy when exiting at a fixed time.



FIGURE 8. Early exiting results according to calibration (enlarged view).

calibration. The reason may be that the overconfidence problem in the early stage of a driving trip makes early exiting difficult. Temporal confidence calibrations are effective in mitigating the overconfidence problem in the early stages. The results for driver identification without early exiting, which means that the system always exited at a fixed time, are shown in Fig. 7. Driver identification with 26 min of fixed-length data for each trip reached 90.06% accuracy. The system with early exiting outperformed the system without early exiting, which had 84.53% accuracy with 10 min of driving data, regardless of the calibration method. In particular, the proposed system with early exiting applied using hybrid calibration achieved 90.06% accuracy at an average of 6.7 min, which is the same accuracy with 74.2% latency reduction compared to driver identification using 26 min of fixed-length data for each trip.

2) TEMPORAL CALIBRATION RESULTS

Fig. 8 shows an enlarged view of the early exiting results according to the calibration methods. Temperature scaling showed a slightly better trade-off between accuracy and latency than label smoothing. The hybrid confidence calibration outperformed the other temporal calibrations. The reason may be that the hybrid method considers the difficulty of a trip.

We compared the proposed temporal calibration and standard calibration results. Fig. 9 illustrates a comparison



FIGURE 9. Comparison between temporal temperature scaling and standard temperature scaling.



FIGURE 10. Comparison between temporal label smoothing and standard label smoothing.

between temporal temperature scaling and standard temporal scaling. As illustrated, the proposed system with temporal temperature scaling outperformed that with standard temperature scaling. Fig. 10 illustrates a comparison between temporal label smoothing and standard label smoothing. As illustrated, the proposed system with temporal label smoothing outperformed that with standard label smoothing. The reason may be that the overconfidence problem in the early stage of a driving trip makes early exiting difficult. Temporal confidence problem in the early stages.

We performed an ablation study for the hybrid calibration, as shown in Table 1. Results show latency: the average driving time when maintaining accuracy during temporal early exiting. The latency was 9.2 min when the temperature acquisition for each trip process was omitted. One best temperature was applied to all trip data to obtain smoothed labels for this experiment. Similarly, the latency was 9.8 min when the temperature acquisition and smoothed label conversion process was omitted. This experiment is equivalent to a simple combination method applying temperature scaling after label smoothing. Therefore, we showed that both

TABLE 1. Ablation study for hybrid confidence calibration.

	Average driving time with temporal early-exiting while maintaining accuracy (min)
Proposed hybrid calibration	6.7
Proposed hybrid calibration without temperature acquisition	9.2
Proposed hybrid calibration without temperature acquisition and smoothed label conversion (Simple combination that applies temporal temperature scaling after temporal label smoothing)	9.8



FIGURE 11. Examples of the change in confidence score while driving a trip before and after hybrid calibration (Driver ID: 1).

temperature acquisition and smoothed label conversion are essential.

3) EFFECTS OF TEMPORAL CALIBRATION

To analyze the obtained results in more detail and investigate the effects of temporal calibration, we visualized the change in the confidence score over time during a trip. Fig. 11 illustrates examples before and after calibration. Before calibration, overconfidence occurred at the beginning of driving. In the figure, the top row presents an example of early identification, while the bottom row presents an example of late identification. If a selected temperature is low in hybrid calibration, there is no overconfidence in the early driving stage of the trip. If a selected temperature is high in hybrid calibration, there is overconfidence in the early driving stage of the trip.

To evaluate the characteristics of temporal calibration, we analyzed the timing of early exiting using a box-plot presentation. Through the box plot, we analyzed the timing of early exiting with no calibration and three types of temporal calibration. These box plots include the shortest timing, lower quartile (25%), median (50%), upper quartile (75%), and the longest timing according to the threshold with accuracy.



FIGURE 12. Early exiting without calibration.



FIGURE 13. Early exiting with temporal temperature scaling.



FIGURE 14. Early exiting with temporal label smoothing.



FIGURE 15. Early exiting with hybrid confidence calibration.

As shown in Fig. 12, early exiting without calibration showed a low variation in exit timing. Figs. 13 and 14 show the early exiting results with temporal temperature scaling and temporal label smoothing, respectively. Early exiting with calibration showed a high variation in exit timing. This implies that the system exits at various times according to a trip. Therefore, the proposed calibration methods

automatically train and calibrate a model to identify the driver with less driving data for easy-to-identify trips and more driving data for hard-to-identify trips. Hybrid calibration had the highest variation in exit timing, as shown in Fig. 15. This is because the calibration strength is adjusted according not only to driving time but also trip difficulty. The temperature



FIGURE 16. Temporal temperature scaling results according to hyperparameters.



FIGURE 17. Temporal label smoothing results according to hyperparameters.

acquisition phase, which finds temperatures for each trip in hybrid calibration, emphasizes adaptiveness by considering trip difficulty.

4) TEMPORAL CALIBRATION HYPERPARAMETERS

As described in Section IV, the hyperparameters α and β control the calibration strength and thus play an important role in model performance. Therefore, we set the hyperparameters to different magnitudes to observe their effects on model performance.

We explored hyperparameter settings for temporal temperature scaling and temporal label smoothing. Fig. 16 shows the temporal temperature scaling results for various α and β settings. Fig. 17 shows the temporal label smoothing results for various α and β settings. In both cases, we observed similar trade-offs between accuracy and latency, regardless of the hyperparameter settings.

Finally, we explored the hyperparameter settings for the hybrid calibration. Fig. 18 shows the results according to the grid search space. A grid equal to 10 means that α and β take integer values from 1 to 10 for temperature scaling. The grid equal to 5 showed a relatively low performance, and grids equal to 10 and 15 showed similar performances.



FIGURE 18. Hybrid calibration results according to grid search space.



FIGURE 19. Hybrid calibration results according to the number of folds during temperature acquisition.

Fig. 19 shows the results according to the number of P folds. Grids of 10 or more are sufficient for hybrid calibration. The large fold is advantageous for the temperature selection efficiency per trip.

D. EXPERIMENTAL RESULTS ON A PUBLIC DATASET

As mentioned above, public datasets have many restrictions, such as few drivers and fixed routes. Most previous studies on public datasets have reported good accuracy. Mekki et al. [3] compared the performance with four public datasets. Performance with the OSF dataset [30] was relatively poor because of the large number of drivers and various driver conditions. Therefore, we analyzed early exiting with the OSF dataset, although the dataset is based on a simulator and has only one route. After excluding the data of drivers with insufficient trips, we experimented with the trip data of 35 drivers.

The driver identification results without early exiting, which means that the system always exited at a fixed time, showed a latency of 6 min to reach 91.43% accuracy on the OSF dataset. Table 2 summarizes the results for the system with early exiting according to calibration method on the OSF dataset. The system with early exiting outperformed the system without early exiting regardless of the calibration method. The system without calibration showed an inferior

TABLE 2. OSF dataset results.

	Average driving time with temporal early-exiting while maintaining accuracy (min)
Proposed temporal temperature scaling	1.7
Proposed temporal label smoothing	1.5
Proposed hybrid calibration	1.1

accuracy of 82.86%. The proposed system after the three types of temporal calibration was highly improved compared to the system without calibration. Temporal confidence calibration methods are effective in mitigating the overconfidence problem in the early stages even on the public dataset with many restrictions. In particular, the proposed system with early exiting applied using hybrid calibration achieved 91.43% accuracy at an average of 1.1 min, which is the same accuracy with 81.67% latency reduction compared to driver identification using 6 min of fixed-length data for each trip.

VI. CONCLUSION

In this paper, a driver identification system is proposed that benefits from temporal early exiting. First, we conducted temporal early exiting for driver identification. Additionally, we proposed three types of temporal confidence calibration to address the overconfidence problem so that the system can work effectively during early exit. In particular, we proposed a novel hybrid confidence calibration method to improve the trade-off between accuracy and latency by considering the difficulty of identifying the driver during a trip. The proposed system achieved 90.06% accuracy with early exiting at an average of 6 min, the same accuracy with 74.2% latency reduction compared to driver identification with 26 min of fixed-length data on a naturalistic dataset. Furthermore, the proposed system achieved 91.43% accuracy at an average of 1.1 min, the same accuracy with 81.67% latency reduction on a public dataset.

IVs should consider the needs of different drivers to provide a safer and more comfortable driving experience. We proposed a novel driver identification system that allows drivers to provide personalized features as early as possible. In a shared mobility context, driver identification helps car-sharing providers verify authorized drivers by tracking their driving styles and providing personalized rides for their customers. In an autonomous mobility context, self-driving software with personalized driving styles provides a safer and more comfortable driving experience.

Although it has been found that temperature scaling and label smoothing have similar calibration effects for a DNN, the relationship between temperature scaling and label smoothing is an open research topic in machine learning theory. In this paper, hybrid calibration combining temperature scaling and label smoothing is proposed; however, there is a limitation as their relationship is not considered. Further improvement is possible instead of rescaling by applying temperature scaling again in the last step of hybrid calibration. It is also necessary to study the relationship with knowledge distillation.

To the best of our knowledge, this paper is the first to explore temporal early exiting in driver identification and the first to consider trip difficulty for temporal confidence calibration. We believe that the proposed methods can be easily applied to existing driver identification systems and will promote applications in driver identification.

REFERENCES

- H. Cai, Z. Hu, Z. Chen, and D. Zhu, "A driving fingerprint map method of driving characteristic representation for driver identification," *IEEE Access*, vol. 6, pp. 71012–71019, 2018.
- [2] M. N. Azadani and A. Boukerche, "Driving behavior analysis guidelines for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6027–6045, Jul. 2022.
- [3] A. E. Mekki, A. Bouhoute, and I. Berrada, "Improving driver identification for the next-generation of in-vehicle software systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7406–7415, Aug. 2019.
- [4] M. N. Azadani and A. Boukerche, "Driver identification using vehicular sensing data: A deep learning approach," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–6.
- [5] T. Wakita, K. Ozawa, C. Miyajima, K. Igarashi, K. Itou, K. Takeda, and F. Itakura, "Driver identification using driving behavior signals," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2005, pp. 396–401.
- [6] J. Chen, Z. C. Wu, and J. Zhang, "Driver identification based on hidden feature extraction by using adaptive nonnegativity-constrained autoencoder," *Appl. Soft Comput.*, vol. 74, pp. 1–9, Jan. 2019.
- [7] Y. Xun, J. Liu, N. Kato, Y. Fang, and Y. Zhang, "Automobile driver fingerprinting: A new machine learning based authentication scheme," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1417–1426, Feb. 2020.
- [8] N. M. Nor and A. Wahab, "Driver identification and driver's emotion verification using KDE and MLP neural networks," in *Proc. 3rd Int. Conf. Inf. Commun. Technol. Moslem World (ICT4M)*, Dec. 2010, pp. E96–E101.
- [9] W. Dong, J. Li, R. Yao, C. Li, T. Yuan, and L. Wang, "Characterizing driving styles with deep learning," 2016, arXiv:1607.03611.
- [10] J. Zhang, Z. Wu, F. Li, C. Xie, T. Ren, J. Chen, and L. Liu, "A deep learning framework for driving behavior identification on in-vehicle CAN-BUS sensor data," *Sensors*, vol. 19, no. 6, pp. 1–17, Mar. 2019.
- [11] H. Ren, M. Pan, Y. Li, X. Zhou, and J. Luo, "ST-SiameseNet: Spatiotemporal Siamese networks for human mobility signature identification," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 1306–1315.
- [12] M. N. Azadani and A. Boukerche, "Convolutional and recurrent neural networks for driver identification: An empirical study," in *Proc. IEEE/IFIP Netw. Oper. Manage. Symp.*, Apr. 2022, pp. 1–6.
- [13] G. Kar, S. Jain, M. Gruteser, J. Chen, F. Bai, and R. Govindan, "PredriveID: Pre-trip driver identification from in-vehicle data," in *Proc.* 2nd ACM/IEEE Symp. Edge Comput., Oct. 2017, pp. 1–12.
- [14] R. Khan, N. Saxena, O. Rana, and P. Gope, "ATVSA: Vehicle driver profiling for situational awareness," in *Proc. IEEE Eur. Symp. Secur. Privacy Workshops*, Jun. 2022, pp. 348–357.
- [15] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. ICML*, Aug. 2017, pp. 1321–1330.
- [16] S. Ma, L. Sigal, and S. Sclaroff, "Learning activity progression in LSTMs for activity detection and early detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1942–1950.
- [17] Y. Wang, K. Lv, R. Huang, S. Song, L. Yang, and G. Huang, "Glance and focus: A dynamic approach to reducing spatial redundancy in image classification," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, Dec. 2020, pp. 1–15.

- [18] Y. Wang, Y. Yue, Y. Lin, H. Jiang, Z. Lai, V. Kulikov, N. Orlov, H. Shi, and G. Huang, "AdaFocus V2: End-to-end training of spatial dynamic networks for video recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 20030–20040.
- [19] P. Boyer, D. Burns, and C. Whyne, "Out-of-distribution detection of human activity recognition with smartwatch inertial sensors," *Sensors*, vol. 21, no. 5, p. 1669, Mar. 2021.
- [20] R. Müller, S. Kornblith, and G. Hinton, "When does label smoothing help?" in *Proc. Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, Dec. 2019, pp. 1–10.
- [21] S. Jafarnejad, G. Castignani, and T. Engel, "Towards a real-time driver identification mechanism based on driving sensing data," in *Proc. IEEE* 20th Int. Conf. Intell. Transp. Syst. (ITSC), Oct. 2017, pp. 1–7.
- [22] S. H. Sanchez, R. F. Pozo, and L. A. H. Gomez, "Driver identification and verification from smartphone accelerometers using deep neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 97–109, Jan. 2022.
- [23] S. Teerapittayanon, B. McDanel, and H. Kung, "BranchyNet: Fast inference via early exiting from deep neural networks," in *Proc. Int. Conf. Pattern Recognit.*, Dec. 2016, pp. 2464–2469.
- [24] A. Ghodrati, B. E. Bejnordi, and A. Habibian, "FrameExit: Conditional early exiting for efficient video recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15608–15618.
- [25] R. Tang, K. Kumar, J. Xin, P. Vyas, W. Li, G. Yang, Y. Mao, C. Murray, and J. Lin, "Temporal early exiting for streaming speech commands recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.* (*ICASSP*), May 2022, pp. 7567–7571.
- [26] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classifiers*, vol. 10, no. 3 pp. 61–74, Mar. 1999.
- [27] G. Pereyra, G. Tucker, J. Chorowski, L. Kaiser, and G. E. Hinton, "Regularizing neural networks by penalizing confident output distributions," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Apr. 2017, pp. 1–15.
- [28] T. Leathart and M. Polaczuk, "Temporal probability calibration," 2020, arXiv:2002.02644.
- [29] S. Huang, Y. Luo, Z. Zhuang, J.-G. Yu, M. He, and Y. Wang, "Context-aware selective label smoothing for calibrating sequence recognition model," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 4591–4599.
- [30] S. Taamneh, P. Tsiamyrtzis, M. Dcosta, P. Buddharaju, A. Khatri, M. Manser, T. Ferris, R. Wunderlich, and I. Pavlidis, "A multimodal dataset for various forms of distracted driving," *Sci. Data*, vol. 4, no. 1, Aug. 2017, Art. no. 170110.



JAEBONG LIM was born in 1992. He received the B.S. and M.S. degrees from Pusan National University, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. His research interests include embedded systems, lowpower devices, embedded AI, TinyML, and driver behavior analysis.



YUNJU BAEK was born in 1967. He received the Ph.D. degree in computer science from KAIST, Republic of Korea, in 1997. He was an Invited Professor at KAIST, a CTO at Naver Corporation, and an Assistant Professor at Sookmyung Women's University. He is currently a Professor with the School of Computer Science and Engineering, Pusan National University. His research interests include embedded systems, RTLS systems, wireless sensor networks, embedded AI, TinyML, and driver behavior analysis.



BUMHEE CHAE was born in 1983. He received the B.S. degree in software engineering from the Kumoh National Institute of Technology, Republic of Korea, in 2009. He was a BigData Platform Team Leader at SureSoft Technologies Inc. He is currently a Devops Group Leader at Suremobility Inc. His research interests include software testing, mobility platform development, and autonomous driving systems.

...