Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 6601014, 18 pages
https://doi.org/10.1155/2022/6601014

Research Article

Stacking-Based Ensemble Learning Method for the Recognition of the Pedestrian Crossing Intention

Hongjia Zhang, Song Gao , and Pengwei Wang

School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, China

Correspondence should be addressed to Song Gao; gaosongsdut@163.com

Received 28 June 2022; Revised 13 September 2022; Accepted 12 October 2022; Published 2 November 2022

Academic Editor: Zhenning Li

Accurate recognition of pedestrian crossing intentions is essential for the safe operation of autonomous vehicles on urban roads. However, the current pedestrian crossing intention recognition model has the problems of relatively low recognition accuracy and short recognition advance time. Based on the above problems, this paper carried out a study on the recognition model of pedestrian crossing intention. Firstly, the pedestrian and vehicle crossing data were collected through laser radar and a high-definition monitor, and 1980 groups of valid samples were selected. Secondly, the pedestrian crossing intention characterization parameter set was determined through statistical analysis. Finally, this paper proposes a pedestrian crossing intention recognition model based on stacking ensemble learning. The ensemble learning framework integrates random forest (RF), support vector machine (SVM), long short-term memory network (LSTM), an attention mechanism, and bidirectional LSTM (AT-Bi-LSTM). Compared with traditional machine learning methods, the proposed method shows greater advantages in recognition accuracy. The model recognition accuracy reaches 95.36% when the model is recognized at 0.5 s before crossing the zebra crossing, and the model recognition accuracy is 89.27% when the model is recognized at 1s before crossing the zebra crossing. The research in this paper is of great significance for building a more intelligent pedestrian-vehicle collaboration and promoting the industrial application of the autonomous vehicle.

1. Introduction

A zebra crossing is an area for pedestrians to cross the road, and it is also a potential conflict area between vehicles and pedestrians [1, 2]. According to the accident statistics report issued by the road traffic management department, the number of pedestrian deaths rose from 14,923 in 2015 to 17,473 in 2019. The proportion of pedestrian deaths in the total number of traffic accident deaths increased from 25.72% to 27.84%. The number of injured pedestrians rose from 34,379 to 45,495. Further calculations found that from 2015 to 2019, in each pedestrian-related accident, 1.2 pedestrians were injured or died on average [3–7]. The above data shows that in recent years, the situation of vehicle-pedestrian accidents in China has been deteriorating year by year, and both the absolute number and the proportion of fatalities have been rising. At the same time, the above data prove that in the road traffic system, pedestrians belong to a vulnerable group. Once a pedestrian-vehicle accident occurs, even a slight scratching accident may induce serious pedestrian injury or even death [8, 9].

With the rapid development of current technology, autonomous vehicles are getting closer to reality. Autonomous vehicles have significant potential in reducing collision-related casualties, improving traffic conditions, and reducing traffic jams and vehicle emissions. The U.S. Department of Transportation released the Autopilot System Safety Vision 2.0 in 2017, which aims to improve the safety and reliability of the autopilot system in order to achieve the purpose of reducing the accident rate [10]. In 2016, the China Association of Automotive Engineers released a route for autonomous vehicle technology. The route mentioned that every vehicle will have a fully automated driving system or assisted driving system between 2026 and 2030 to improve road traffic safety [11].

Driving safely on urban roads is an important challenge for autonomous vehicles. In particular, it should be pointed out that there are a large number of pedestrians on urban roads. As relatively complex individuals, their movement behavior is affected by factors such as their own emotions, traffic environment, and weather. Through vision, sound, gestures, and actions, the driver can understand the pedestrians' intentions and then accurately complete the interaction with the pedestrian. However, for autonomous vehicles, it is difficult to understand the intentions of pedestrians and then accurately complete the pedestrian-vehicle interaction [12, 13]. The zebra crossing is the main interaction area between pedestrians and vehicles. Therefore, the research on the pedestrian crossing intention recognition model is carried out in this paper.

The main contributions of this paper are as follows:

(1) The current pedestrian crossing intention recognition models are mainly established based on traditional machine learning algorithms or deep learning algorithms, and the recognition accuracy is relatively low. This paper proposes a machine learning algorithm combination framework that can improve model recognition accuracy, namely, the stacking ensemble learning framework, which integrates four classical algorithms.

(2) The current pedestrian crossing intention recognition model usually cannot take into account the recognition accuracy and recognition advance time. Different from the current model, this model greatly increases the recognition advance time on the premise of ensuring recognition accuracy.

2. Related Works

At present, scholars at home and abroad have carried out a lot of research on pedestrian crossing intention recognition and have achieved relatively fruitful research results.

Mingus et al. [14] considered the trajectory and posture of pedestrians and established a pedestrian crossing intention recognition model based on the Gaussian dynamic model. The model recognition accuracy is 80%. Quintero et al. [15, 16] collected the posture data of pedestrians crossing the zebra crossing and divided the pedestrian movement posture into 11 key points of the human body. A pedestrian crossing intention recognition model is established based on the hidden Markov model. When it is recognized 0.125 s in advance, the accuracy of the model is 80%. Fang and Lopez [17] collected a large amount of posture data of pedestrians crossing the zebra crossing. The direction parameters were calculated between different points through the positioned human body key point data, and a pedestrian crossing intention recognition model was established using the support vector machine (SVM) algorithm. The model has high recognition accuracy, reaching 93%. Brehar et al. [18] proposed a method to identify pedestrian crossing behavior using a monocular far infrared. The method can still effectively identify pedestrian street cross action in low visibility environments such as nighttime, fog, heavy rain, or smoke, with an accuracy of 93.28%. Căilean et al. [19] propose a novel architecture for improving pedestrian safety at crosswalks. The architecture can effectively detect pedestrians and predict their street cross actions.

Völz et al. [20] established a pedestrian crossing intention recognition model based on a data-driven method. The main input parameters of the model are the distance between pedestrians and the zebra crossing, the distance between vehicles and the zebra crossing parameters, etc. The pedestrian recognition accuracy is 84.74%. Camara et al. [21] collected a large amount of pedestrian crossing data and established a pedestrian crossing intention recognition model by analyzing the relative position between pedestrians and vehicles. The recognition accuracy of the model can reach up to 96%. Zhao et al. [22] used lidar to collect a large amount of pedestrian crossing data and established a pedestrian crossing intention recognition model based on an artificial neural network (ANN) by analyzing the motion parameters of pedestrians and vehicles before crossing the zebra crossing. When recognized 0.5 s in advance, the model recognition accuracy is 92.6%. Zhang et al. [23] proposed a bidirectional long short-term memory network with an attention mechanism (AT-Bi-LSTM) to establish a pedestrian crossing intention recognition model. The recognition accuracy is 90.68% when the model is 0.6 s in advance.

Ghori et al. [24] proposed a new pedestrian crossing intention recognition framework, which combines convolutional neural networks (CNN) and LSTM networks. When recognized 1 s in advance, the recognition accuracy of the model is relatively low, at only 72%. Schulz and Stiefelhagen [25] and Brouwer et al. [26] established a pedestrian crossing intention recognition model by estimating the head movement posture of pedestrians crossing the zebra crossing. Hashimoto et al. [27] collected the intersection information and established a pedestrian crossing intention recognition model based on the dynamic Bayesian network (DBN). Schneemann and Heinemann [28] combined the image data and motion parameters of pedestrians crossing the zebra crossing and established a pedestrian crossing intention recognition model based on SVM.

Through the literature review, it can be seen that the current research on pedestrian crossing intentions has been relatively mature. The recognition accuracy of the intention model is already good, and the highest value has exceeded 90%. However, the recognition advance time of the model is relatively short. Overall, existing models do not seem to be able to maintain high recognition accuracy while maintaining a long recognition advance time.

In general, pedestrian crossing intention recognition can be regarded as a time-series modeling and forecasting problem. Therefore, this paper first collects the continuous data stream 2.1 s before pedestrians cross the zebra crossing. The data collection uses laser radar and a high-definition (HD) monitor. Secondly, the characteristic parameters related to the crossing intention are extracted. The characteristic parameters mainly include pedestrian speed, the distance between pedestrian and zebra crossing, age, gender, vehicle speed, the distance between vehicle and zebra crossing, and time to collision (TTC). Finally, a pedestrian crossing intention recognition model is established based on stacking ensemble learning. The SVM, random forest (RF), LSTM, and AT-Bi-LSTM algorithms were integrated. Figure 1 shows the research framework of this paper.

This paper is divided into five parts, namely, introduction, related works, proposed solution, experimental results, and conclusions. In the first and second parts, it mainly analyzed the conflict between pedestrians and vehicles and introduced the significance of the research on pedestrian crossing intention recognition. In the third part, the crossing intention recognition algorithm was introduced. This paper is based on the stacking ensemble learning algorithm, which integrates SVM, random forest (RF), LSTM, and AT-Bi-LSTM algorithms. Data acquisition equipment and acquisition methods were introduced. The main data acquisition equipment is the laser radar and an HD monitor. In the fourth part, the characteristic parameters of pedestrian crossing intention were analyzed, and the characteristic parameter set of pedestrian crossing intention was obtained. The fourth part also analyzed the results of the pedestrian crossing intention recognition model based on stacking ensemble learning and compares it with the traditional intention recognition algorithm. The fifth part elaborated on the conclusions of this paper.

3. Proposed Solution

3.1. Methodology

3.1.1. Ensemble Learning. Ensemble learning improves the performance of machine learning by combining multiple models. Compared with a single model, this method allows for better prediction performance. At present, it is widely used in some well-known international machine learning competitions (Netflix, KDD2009, and Kaggle) and has achieved good rankings. The ensemble learning method can be used to solve classification and regression tasks [29].

For ensemble learning, there are two main problems faced in the process of model integration, namely, (1) how to change the distribution or weight of the data. (2) How to combine multiple weak classifiers into a strong classifier. For the above two problems, there are three main solutions: (1) bagging method for reducing variance. (2) Boosting method for reducing bias. (3) Stacking method for improving prediction results [30–32]. Stacking ensemble learning has a better effect on improving recognition accuracy. Therefore, this paper chose stacking ensemble learning.

Stacking is a typical representative of ensemble learning methods. Individual weak classifiers are called base classifiers, and the classifiers used for combinations are called meta-classifiers. The base classifier is usually a heterogeneous classifier.

3.1.2. Base Classifier and Meta-Classifier

(1) SVM-Base Classifier. SVM [33] is a commonly used supervised learning algorithm for machine learning. It is a typical linear binary classifier. SVM is also regarded as the process of solving the optimal classification hyperplane. For the SVM, the key is the determination of the kernel function, the penalty function *C*, and the kernel function parameter g. The kernel function selected is the radial basis kernel function. The values of the penalty function C and the kernel function parameter g are determined by the grid search method. In this paper, when the pedestrian intention is identified at 0 s before crossing the zebra crossing, the values of *C* and g are 36 and 2.73, respectively. When the pedestrian intention is identified at 0.5 s before crossing the zebra crossing, the values of *C* and g are 48 and 2.32, respectively. When the pedestrian intention is identified at 1 s before crossing the zebra crossing, the values of *C* and g are 45 and 2.08, respectively. Since SVM is a common and mature algorithm, it will not be described in more detail in this paper.

(2) RF-Base Classifier. RF [34] is a classifier composed of a large number of decision trees, which is regarded as an ensemble learning method. Multiple decision tree classifiers are trained by sampling with replacement (bootstrap). Each decision tree classifier is independent of the others and has no correlation. Many classifiers are integrated into an RF classifier, and multiple decision tree classifiers obtain the final classification result through voting. To achieve a good recognition result, the adjustment of hyperparameters is essential. The hyperparameters refer to the number of decision trees and the maximum number of features. In this paper, we also use the grid search method to determine the two important parameter values. When the pedestrian intention is identified at 0 s before crossing the zebra crossing, the number of decision trees and the maximum number of features are 80 and 5, respectively. When the pedestrian intention is identified at 0.5 s before crossing the zebra crossing, the number of decision trees and the maximum number of features are 115 and 5, respectively. When the pedestrian intention is identified at 1s before crossing the zebra crossing, the number of decision trees and the maximum number of features are 125 and 5, respectively.

(3) LSTM-Base Classifier. At the end of the last century, Hochreiter and Schmidhuber proposed LSTM on the basis of RNN [35], which to some extent overcomes the problem of gradient disappearance and explosion in the back propagation process. The LSTM network introduces the concept of "gates," which are the input gate, forget gate, and output gate. These three gates are also called the memory unit of the network. The main purpose is to selectively delete and retain the associated information in the data to achieve the purpose of continuous update of the cell state and increase the model recognition accuracy. The grid search method was used to determine the hyperparameter values. When pedestrian intention is identified at 0 s before crossing the zebra crossing, the learning rate, hidden unit, and dropout values are 0.01, 128, and 0.4, respectively. When the pedestrian's intention is recognized at 0.5 s before crossing the zebra crossing, the values of the learning rate, hidden unit, and dropout are 0.05, 100, and 0.4, respectively. When the pedestrian intention is recognized at 1 s before crossing

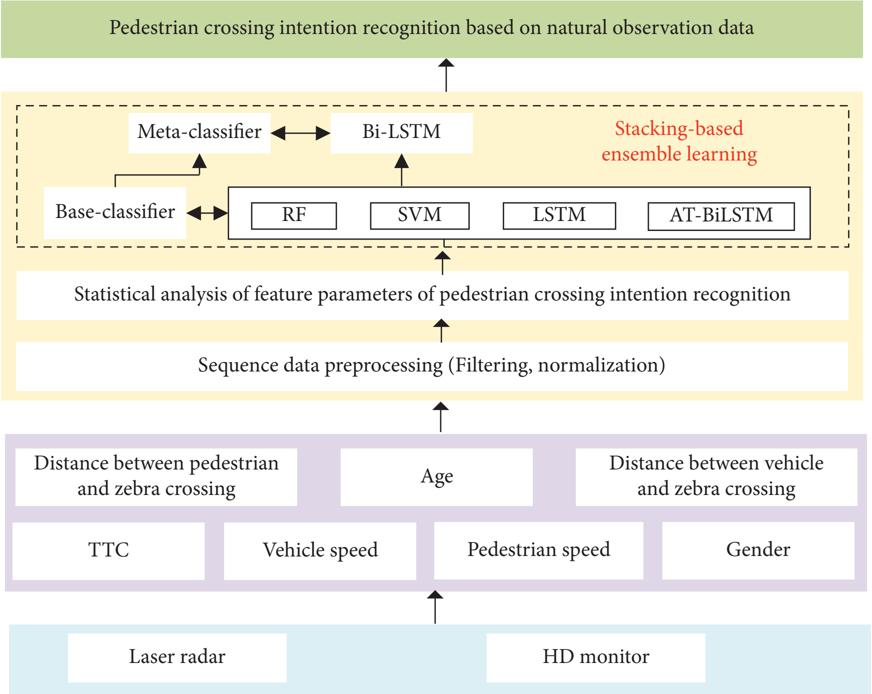

Pedestrian crossing intention recognition based on natural observation data
Meta-classifier
Bi-LSTM
Stacking-based ensemble learning
Base-classifier
RF
SVM
LSTM
AT-BiLSTM
Statistical analysis of feature parameters of pedestrian crossing intention recognition
Sequence data preprocessing (Filtering, normalization)
Distance between pedestrian and zebra crossing
Age
Distance between vehicle and zebra crossing
TTC
Vehicle speed
Pedestrian speed
Gender
Laser radar
HD monitor

Figure 1: Research framework of pedestrian crossing intention: laser radar and HD monitor acquire data before pedestrians cross the zebra crossing. After the data are filtered, the characterization parameters of pedestrians' crossing intention are obtained. After preprocessing, the characterization parameters are input into the stacking learning algorithm, and then the pedestrian crossing intention recognition model is established.

the zebra crossing, the learning rate, hidden unit, and dropout values are 0.001, 100, and 0.5, respectively. Adam was used as the optimizer. In addition, the LSTM network also solves the problem of interdependence before and after the input data so that the cell unit has a longer memory capacity. The specific working steps of the LSTM network are as follows:

Forget gate: the main function is to delete useless information in the cell unit, and the content of the information is determined by the sigmoid function.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right), \tag{1}$$

where σ is the forget gate sigmoid function, W_f is the weight matrix, b_f is the bias term, and the output range of f_t is [0, 1], and its value is inversely proportional to the degree of forgetting.

Input gate: it updates the information in the cell unit of the structure. The sigmoid layer and the tanh layer determine the updated information in the cell information.

$$i_t = \sigma\left(W_t \cdot [h_{t-1}, x_i] + b_i\right), \tag{2}$$

$$\widetilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_i] + b_c\right), \tag{3}$$

where σ is the input gate sigmoid function, tanh is the input gate function, W_t and W_c are weight matrices, b_i and b_c are bias terms, i_t is the input gate cell state update value, and $\widetilde{C}_t$ is the tanh function state update value.

Through formulas (2)–(4), the final updated state value of the cell unit is obtained, and the specific expression is 4.5.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t, \tag{4}$$

where C_{t-1} is the unit cell state value at the previous moment.

The main function of the output gate is to transfer the associated information to the cell unit at the next moment.

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right), \tag{5}$$

where o_t is the output value of the output gate, W_o is the weight matrix, and b_o is the bias term.

The final output h_t of the unit cell at the current moment can be expressed as follows:

$$h_t = o_t \cdot \tanh\left(C_t\right). \tag{6}$$

(4) At-Bi-LSTM-Meta Classifier. Pedestrian crossing intention recognition can be regarded as a sequence recognition problem. The movement state of pedestrians before crossing the zebra crossing can reflect the pedestrians' crossing decision. The data between a certain moment before crossing the zebra crossing and the next moment has a greater correlation. To better capture the characteristic information of pedestrian crossing intentions and fully obtain the correlation of sequence data in a period of time before crossing the zebra crossing, this paper adopts Bi-LSTM [36].

The input of the Bi-LSTM model at time t is x_t. During information processing, the state of Bi-LSTM from the forward to backward direction is updated as follows:

$$h_{fwt} = H\left(W_{fw}x_t + W_{fw1}h_{fwt-1} + b_{fw}\right), \tag{7}$$

where H is the backward output function, W_{fw} is the weight matrix from the input layer to the forward layer , W_{fw1} is the weight matrix between forward layers, and b_{fw} is the bias term.

The Bi-LSTM model is then updated from the backward to forward direction as follows:

$$h_{bwt} = H\left(W_{bw}x_t + W_{bw1}h_{bwt-1} + b_{bw}\right), \tag{8}$$

where $H^{`}$ is the forward output function, W_{bw} is the weight matrix from the input layer to the back layer, W_{bw1} is the weight matrix between back layers, and b_{bw} is the bias term.

Equation (9) describes the final output of the Bi-LSTM model following the forward and backward superimposition as follows:

$$h_t = \widetilde{H}\left(W_{fw2}h_{fwt} + W_{bw2}h_{bwt} + b_o\right), \tag{9}$$

where $\widetilde{H}$ is the output function of the output layer, W_{fw2} is the weight matrix from the forward layer to the output layer, and W_{bw2} is the weight matrix from the backward layer to the output layer.

The parameters of the pedestrian crossing intention are not equally important. To capture the most important information and shorten the flow distance of information, the Bi-LSTM-based attention mechanism was introduced [37]. The grid search method was used to determine the hyperparameter values. When the pedestrian intention is identified at 0 s before crossing the zebra crossing, the learning rate, hidden unit, and dropout values are 0.005, 120, and 0.4, respectively. When the pedestrian intention is recognized at 0.5 s before crossing the zebra crossing, the values of the learning rate, hidden unit, and dropout are 0.001, 120, and 0.4, respectively. When the pedestrian intention is recognized at 1 s before crossing the zebra crossing, the learning rate, hidden unit, and dropout values are 0.001, 100, and 0.2, respectively. Adam was used as the optimizer. Figure 2 presents the four components of the AT-Bi-LSTM framework, namely, (1) the input layer, which inputs the feature parameter sequence of the crossing intention, (2) the LSTM layer, (3) the attention layer, and (4) the output layer.

The correlation function of the attention layer is expressed as follows:

$$\begin{aligned} Q &= \tanh(P),\\ \beta &= softmax\left(\gamma^T Q\right),\\ \varepsilon &= P\beta^T,\\ h^* &= \tanh(\varepsilon), \end{aligned} \tag{10}$$

where P is a vector composed of $h_1, h_2, h_3 \ldots h_t$, T is the data length, γ is a trained parameter vector, and h^* is the final value used for classification.

3.1.3. *Stacking-Based Ensemble Learning Algorithm Description.*

The training set based on stacking ensemble learning includes a primary training set and a secondary training set. In the training phase, the secondary training set is generated using the base classifier. If the training set of the primary classifier is used directly to generate the secondary training set, the risk of over-fitting will be relatively high. Therefore, cross-validation is generally used to generate training samples for the meta-classifier. The method used in this paper is 5-fold cross-validation. Firstly, the base classifier (SVM, RF, LSTM, and AT-Bi-LSTM) is obtained through the primary training set training, and the primary training set is divided into 5 subsets. Secondly, the training set is reconstructed through 5-fold cross-validation to obtain the secondary training set, which is used to train the meta-classifier. Finally, the meta-classifier (Bi-LSTM) is obtained through the training of the secondary training set.

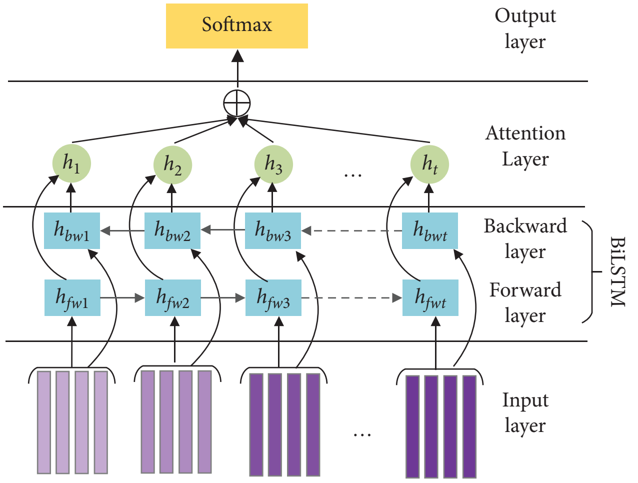


Figure 2: AT-Bi-LSTM structure: input layer is used to input data; the data flows into the forward and backward layers of the Bi-LSTM to obtain important clues in the data. The attention layer is used to remove useless information from data and extract key features. The softmax layer is responsible for outputting pedestrian intentions.

Figure 3 presents the framework of stacking-based ensemble learning. Table 1 is the pseudocode of the stacking algorithm, and the main steps of model training are described as follows:

Step 1: divide the pedestrians' intention sample dataset S into the training set S_{train} and S_{test} according to the ratio of 3 : 1. According to the 5-fold cross-validation method, we randomly and equally divide S_{train} into 5 subsets, namely, S_1, S_2, S_3, S_4, and S_5, and select one of the subsets S_i $(i = 1, 2, \ldots, 5)$ as the verification subset in turn. Use the remaining $S_{+i} = S_{\text{train}} - S_i$ as the training subset.

Step 2: we use S_{+i} as the training set of base classifiers RF, SVM, LSTM, and AT-Bi-LSTM, use S_i as the verification subset, and output the test result x_i. Simultaneously, we predict the test set S_{test} and output the prediction result y_i.

Step 3: we iterate step 2 five times to obtain $\{x_1, x_2, x_3, x_4, \text{and } x_5\}$, and we merge the results according to the columns to get the column vector X_1 of the same length as the original training set S_{train}. We combine the test

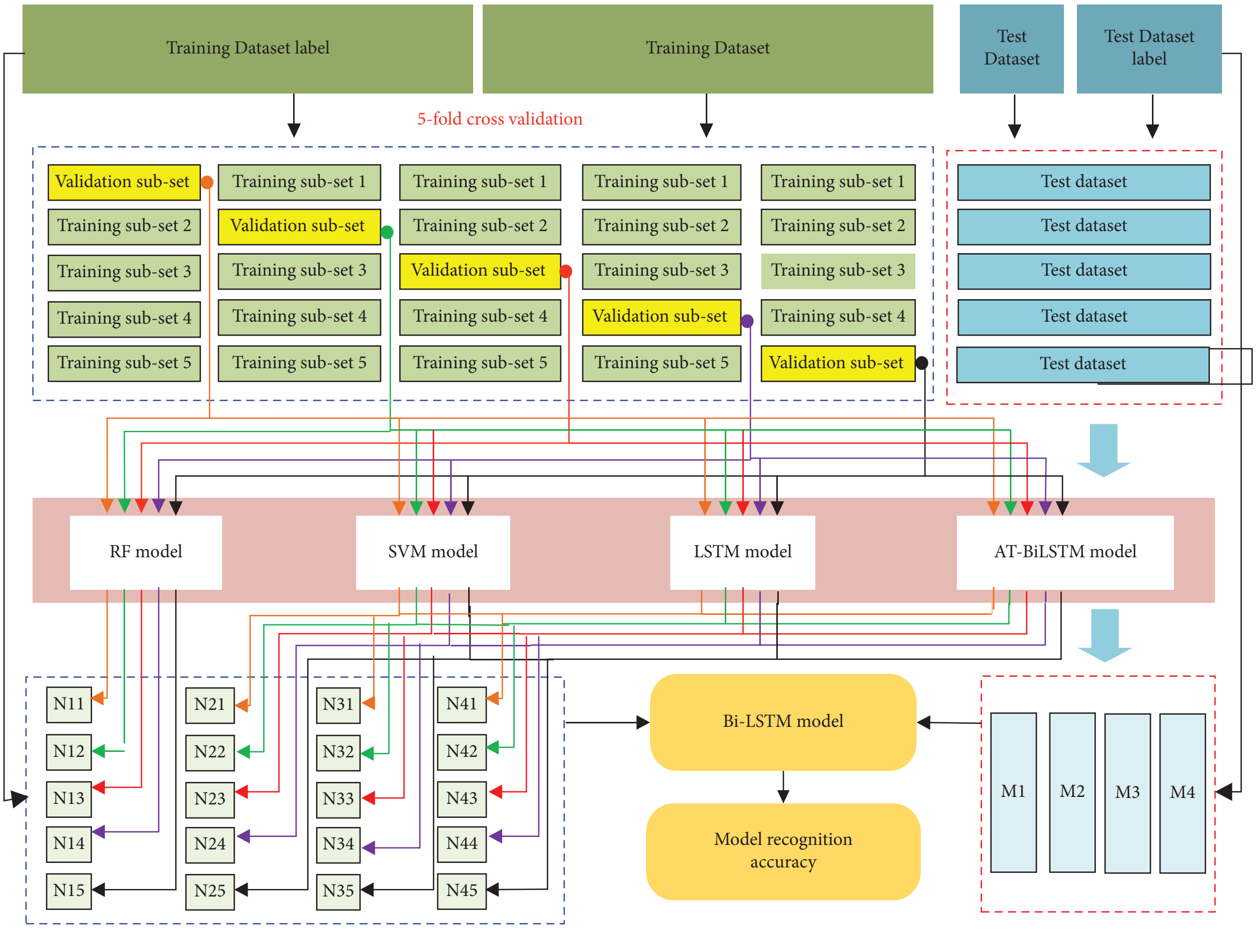


Figure 3: Stacking ensemble learning architecture: the data are divided into the training set and the test set. The training set is divided into four training subsets and one validation subset, and a new subset is obtained by the basic classifier RF, SVM, LSTM, and AT Bi LSTM. The new subset is trained by the meta-classifier to obtain the pedestrian crossing intention recognition model. Similarly, the new test set is obtained by four base classifiers. The test set is input into the intention recognition model to obtain the final recognition accuracy.

Table 1: Pseudocode of the stacking algorithm.

```
Input: training set $S_{\text{train}} = \{(x_1 y_1), (x_2 y_2), \ldots, (x_m y_m)\}$;
    Base classifier: $L_1, L_2, \ldots L_T$;
    Meta classifier: $L$ (Bi-LSTM).
Process:
  for $t = 1,2, \ldots, T$ do
    $h_t = L_t\ (S_{\text{train}})$% train the base classifier separately using the training set
  end for
  $N = \varnothing$; % create new datasets
  for $i = 1, 2\text{m}$ do
    for $t = 1,2, \ldots, T$ do
      $z_{it} = h_t\ (x_i)$% use the classifier $h_t$ to test the validation set
    end for
    $N = N \cup \{(z_{i1}, z_{i2}, \ldots z_{iT}), y_i\}$
  end for
    h' = $L\ (N)$; % training a meta-classifier based on the Bi-LSTM algorithm with the newly combined dataset
Output: $H(x) = h'\ (h_1\ (x), h_2\ (x) \ldots, h_T\ (x))$
```

samples and take the average to obtain a column vector Y_1 of the same length as the original test S_{test}.

Step 4: by sequentially performing step 3 on the base classifiers SVM, LSTM, and AT-Bi-LSTM, we obtain X_2, X_3, and X_4 from the original training set and Y_2, Y_3, and Y_4 from the original test set.

Step 5: we combine X_1, X_2, X_3, and X_4 and the label L of the original training set S_{train} to obtain a new sample dataset $N = \{X_1, X_2, X_3, X_4, \text{and } L\}$, and we use it as the training dataset of the meta-classifier Bi-LSTM. We obtain the accuracy of the meta-classifier via the test dataset $M = \{Y_1, Y_2, Y_3, Y_4, \text{and } P\}$.

3.2. Experimental

3.2.1. Study Site. Figures 4 and 5 are diagrams of the study site and equipment placement location, respectively. The zebra crossing section has no signal light control and monitoring equipment. The width of the zebra crossing is 12 m, a two-way four-lane. The road gradient is small and negligible, and the road is separated by a double yellow line. There is no green belt or buffer waiting area. The selected road is a common road in the city. The traffic flow in this section is mainly composed of small passenger vehicles.

3.2.2. Experimental Equipment. The laser radar model LUX4L-4 selected in this experiment is produced by the German IBEO company, as shown in Figure 6. The radar used in the experiment belongs to the four-line radar, and the scanning frequency is set to 12.5 Hz. The detection range of the lidar is 0.3–200 m, the vertical viewing angle is 3.2°FOV, and the horizontal viewing angle can reach 110°. The radar used in the experiment can perform real-time scanning of all objects within the detection field, including moving objects and stationary objects. At the same time, the data collected by the radar are read through the associated software ILV-Premium, as shown in Figure 6. Through this software, the type, speed, and position of the target detected by the radar can be displayed in real time. The specific display interface of software is shown in Figure 6.

The selected HD monitor is small in size, and the video resolution is 1920 × 1080. Figure 6 shows the physical image. Both the LUX radar and the driving recorder are powered by small batteries. The data collection location is 15 m away from the zebra crossing. In addition, the use of radar alone will miss a large amount of data, making the selection work more complicated. At the same time, the gender and age of pedestrians cannot be judged. In order to overcome this problem, radar and HD monitors are used together. After the two devices are synchronized in time, the HD monitor is used to determine whether the pedestrian wants to cross the zebra crossing. The data of the pedestrian before or when crossing the zebra crossing are collected by the laser radar. The radar point cloud image recorded by ILV-Premium is the main, and the video recorded by the HD monitor is auxiliary to realize the precise selection of data.

3.2.3. Data Collection and Analysis. To overcome the influence of time heterogeneity, all observation experiments were conducted on sunny days. Pedestrian crossing intention recognition is a continuous-time series classification problem. The pedestrians' crossing intention is determined according to the speed change within a period of time before the pedestrians cross the zebra crossing or the time series change of the surrounding environment (vehicle speed or the distance between the vehicle and the zebra crossing, etc.). Generally speaking, when pedestrians are crossing the zebra crossing, they determine their intention to cross the zebra crossing by observing the surrounding environment (such as the distance between the vehicle and themselves), which is reflected in the speed of the pedestrian crossing the zebra crossing. If the pedestrian does not slow down, it may be a direct crossing behavior. Figure 7 shows a schematic diagram of the pedestrian crossing. In this paper, pedestrian crossing intentions are divided into three categories, namely, "walking-walking intention (WWI)," "walking-stopping intention (WSI)," and "stopping-walking intention (SWI)." WWI refers to a pedestrian crossing the zebra crossing without stopping after reaching the curb. WSI means that after considering the road traffic environment, pedestrians did not choose to cross directly after reaching the curb but waited. SWI means that pedestrians start to cross the zebra crossing after waiting at the curb.

In this paper, the main process of selecting the characteristic parameters of pedestrian intention before crossing the zebra crossing is as follows.

Check whether the pedestrian has the intention of crossing the zebra crossing through the HD monitor. If the video shows that the pedestrian is WWI, then we need to go back for a certain period of time and collect the pedestrian-related data and vehicle-related data during this period of time through the laser radar. If it is determined through the video that the pedestrians' intention to cross the zebra crossing is WSI or SWI, we use the same method to reverse the laser radar and record it.

The intention characterization parameters selected in this paper are mainly pedestrian speed, the distance between the pedestrian and the zebra crossing, vehicle speed, the distance between the vehicle and the zebra crossing, and TTC. In addition, the paper also introduces the influence of pedestrian age, gender, and group on pedestrians' intention to cross the zebra crossing. The specific definition is as follows:

Pedestrian speed is the mean speed value of pedestrians during a period of time before crossing the zebra crossing, obtained by laser radar. In the process of collecting pedestrian speed by radar, the true speed value is obtained after Kalman filtering, and the speed value of each frame is counted to finally get the mean speed of the pedestrian before crossing the zebra crossing.

The distance between the pedestrian and the zebra crossing (DPZC) refers to the square and root result of the two parameters of the vertical distance between the pedestrian and the curb and the vertical distance between the pedestrian and the zebra crossing.

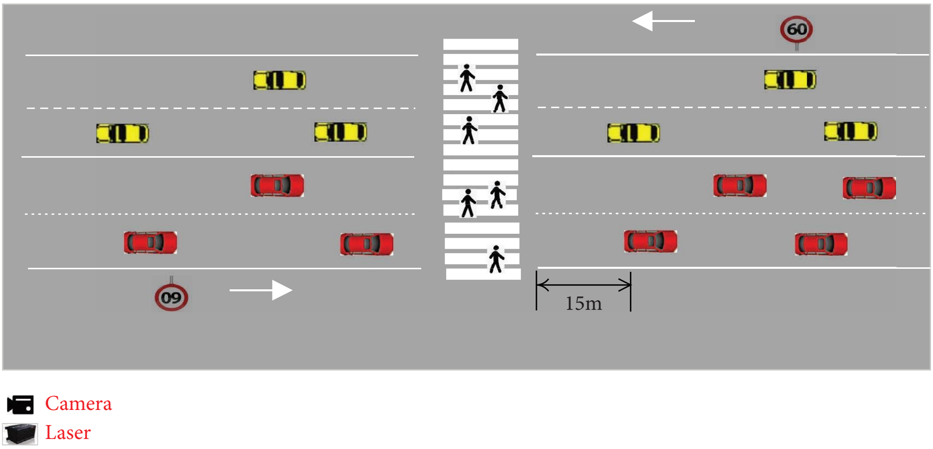


Figure 4: Schematic diagram of the study site: the equipment is placed at the curb, about 15 meters away from the zebra crossing.

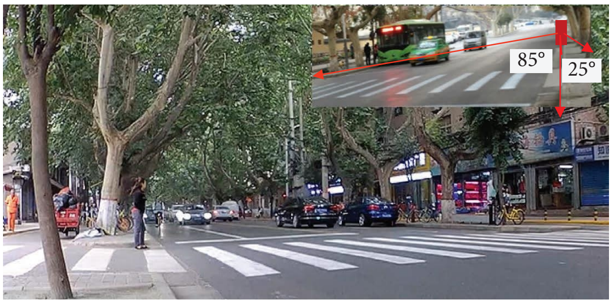


Figure 5: Photograph of the study site: lidar detection angle is 110°. It can completely cover the whole road.

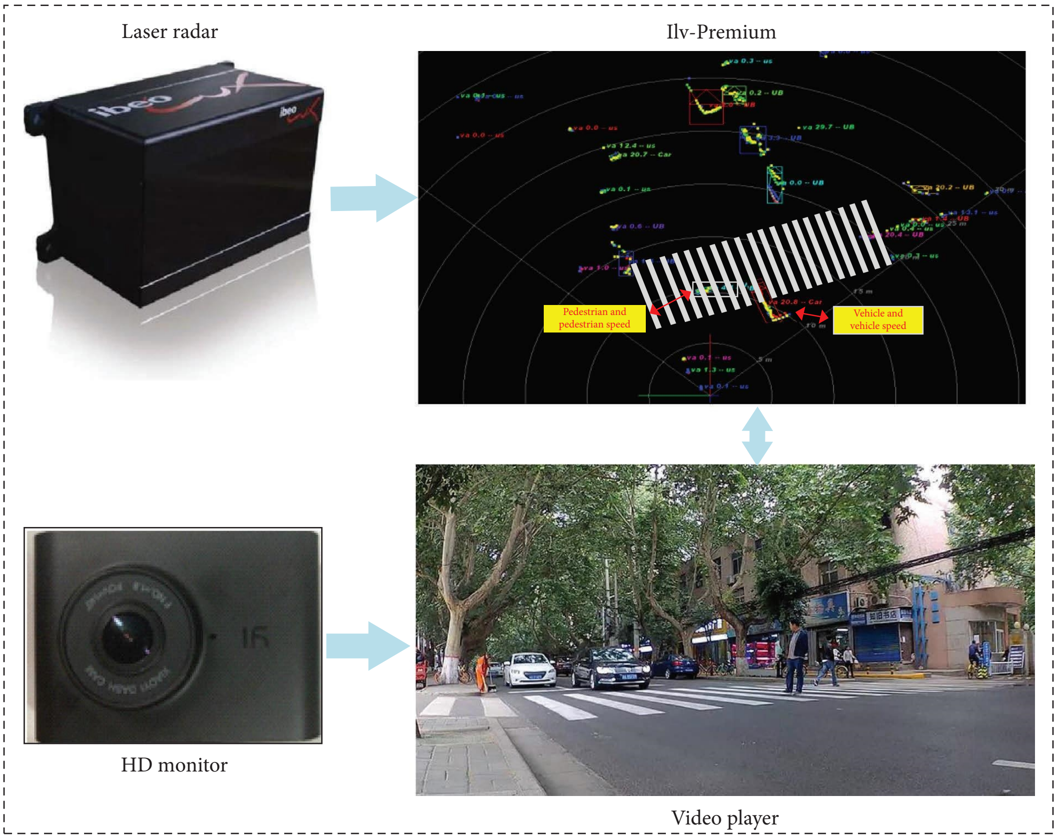


Figure 6: Laser radar and HD monitor: the upper part of the picture is a radar map, and the lower part is a camera map. Time synchronization between two devices.

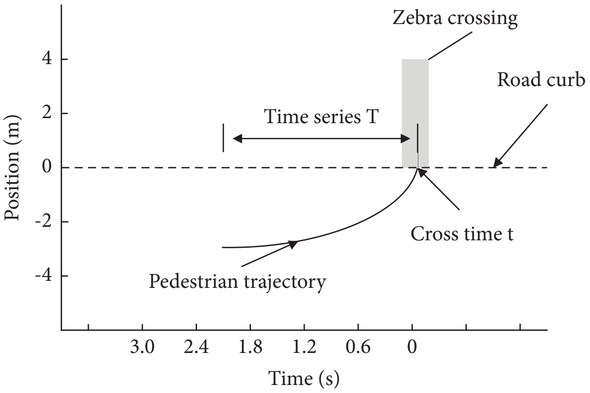


Figure 7: Time-series schematic diagram before pedestrians cross the zebra crossing: the dotted line is the curb, and the gray box is the zebra crossing. Time series *T* refers to the time from the beginning of the pedestrian trajectory to the time when pedestrians arrive at the zebra crossing.

The distance between the vehicle and the zebra crossing (DVZC) refers to the vertical distance between the vehicle and the zebra crossing.

TTC refers to the distance between the vehicle and the zebra crossing divided by the current speed of the vehicle.

3.2.4. Data Preprocessing. The data obtained from the radar will bring a lot of noise and interference signals. In order to make the collected data closer to the real value, this paper uses a Kalman filter to filter the data directly collected by the radar. It should be pointed out that the distance value between the vehicle and the zebra crossing and the vehicle speed value is larger than the pedestrian speed value and the value between the pedestrian and the zebra crossing, in order to more accurately capture the key information in the data, reduce the training time of the model, and improve model recognition accuracy. This paper uses the min-max function to normalize the characteristic parameters.

4. Experimental Results

4.1. Characteristic Parameter Analysis Results

4.1.1. Time to Collision. Figure 8(a) shows the TTC line chart under different crossing intentions within 2.1 s before crossing the zebra crossing. It can be seen that when the intention is WWI, the selected TTC value when pedestrians cross the zebra crossing is the largest, which is at the top of the three curves. When the intention is SWI, the TTC value selected by pedestrians crossing the zebra crossing is second, in the middle of the three curves. When the intention is WSI, the TTC value selected by pedestrians crossing the zebra crossing is the smallest, which is at the bottom of the three curves. As time goes by, the TTC value under different intentions shows a steady downward trend.

Figure 8(b) is a box diagram of the TTC under different crossing intentions. When the intentions are WWI, SWI, and WSI, the mean values of TTC are 5.79 s, 5.22 s, and 2.51 s, respectively. One-way analysis of variance (ANOVA) found that there were significant differences in TTC values under different intentions (F (2, 1977) = 1719.60, $p < 0.001$), and the post-hoc test found that there were significant differences in TTC values after pairings with different intentions ($p < 0.001$).

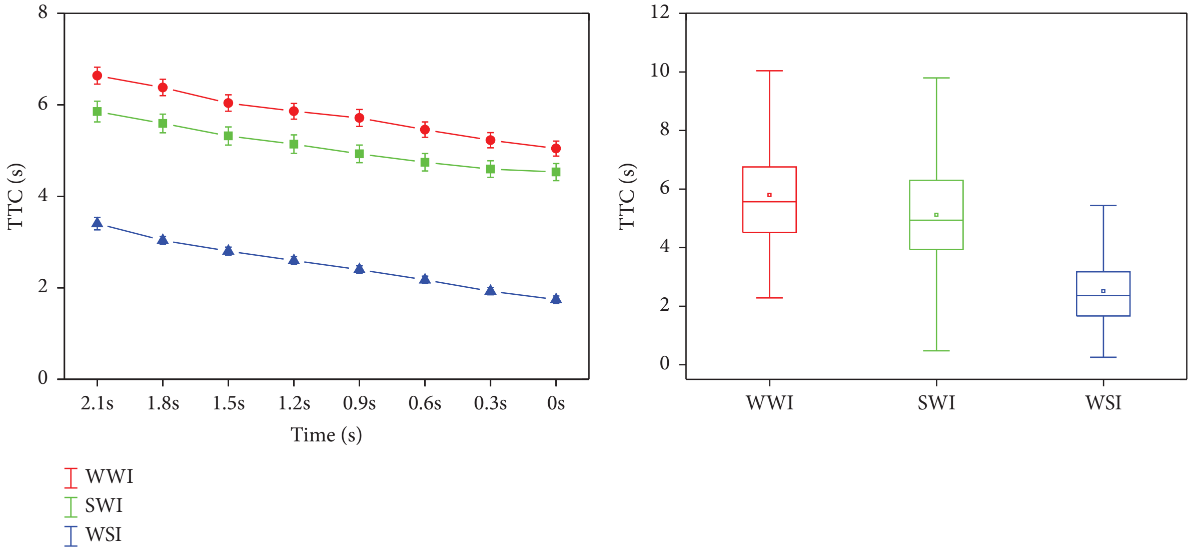
4.1.2. Vehicle Speed. Figure 9(a) shows the vehicle speed line chart under different crossing intentions within 2.1 s before crossing the zebra crossing. It can be seen that when the intention is SWI, the vehicle speed value when pedestrians' cross the zebra crossing is the largest, which is at the top of the three curves. When the intention is WWI, the vehicle speed value is the second, in the middle of the three curves. When the intention is WSI, the vehicle speed value is the smallest, which is at the bottom of the three curves. Generally speaking, with the change of time, the value of vehicle speed does not change much, and the value is relatively stable.

Figure 9(b) is a box diagram of vehicle speed under different crossing intentions. When the crossing intentions are WWI, SWI, and WSI, the mean values of vehicle speed are 30.61 km/h, 29.94 km/h, and 31.21 km/h. One-way ANOVA found that there were significant differences in vehicle speed values under different intentions (F (2, 1977) = 83.69 and $p < 0.001$), and the post-hoc test found that there was no significant difference in the vehicle speed values between WWI and SWI ($p = 0.15 > 0.05$). There is a significant difference in the vehicle speed value between WWI and WSI ($p < 0.001$). There is a significant difference in the vehicle speed value between SWI and WSI ($p < 0.001$).

4.1.3. Distance between Pedestrian and Zebra Crossing. Figure 10(a) shows the DPZC changes under different crossing intentions within 2.1 s before crossing the zebra crossing. It can be seen that when the intention is WWI, the DPZC value when pedestrians cross the zebra crossing is the largest, which is at the top of the three curves. When the intention is WSI, the DPZC value is the second, in the middle of the three curves. When the intention is SWI, the DPZC value is the smallest, which is at the bottom of the three curves. Generally speaking, as time goes by, the DPZC value with the intention of WWI and WSI shows a steady downward trend. The DPZC value with the intention of SWI did not change significantly.

Figure 10(b) is a box diagram of DPZC for pedestrians under different crossing intentions. When the crossing intentions are WWI, SWI, and WSI, the mean values of DPZC are 1.05 m, 0.44 m, and 0.18 m. One-way ANOVA found that there were significant differences in DPZC values under different intentions (F (2, 1977) = 2018.46, $p < 0.001$), and the post-hoc test found that there were significant differences in DPZC values after pairings with different intentions ($p < 0.001$).

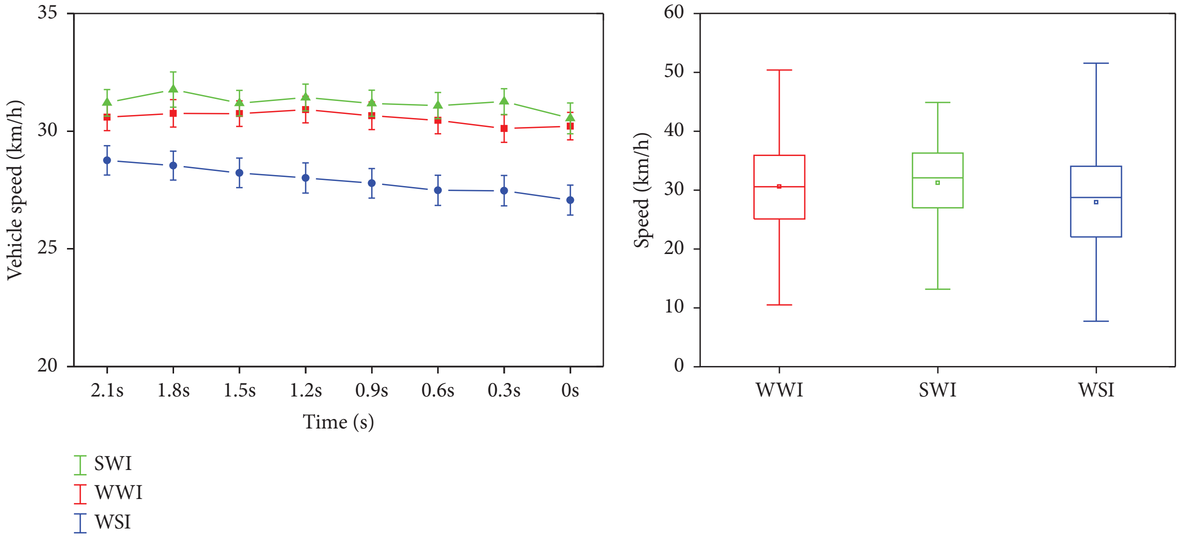
4.1.4. Pedestrian Speed. Figure 11(a) shows the pedestrian speed changes under different crossing intentions within 2.1 s before crossing the zebra crossing. It can be seen that when the intention is WWI, the pedestrian speed value when



(a)

(b)

Figure 8: TTC under different crossing intentions. (a) Line chart of TTC change with time under different intentions. (b) Box diagram of TTC under different crossing intentions.



(a)

(b)

Figure 9: Vehicle speed under different crossing intentions. (a) Line chart of vehicle speed change with time under different intentions. (b) Box diagram of vehicle speed under different crossing intentions.

pedestrians cross the zebra crossing is the largest, which is at the top of the three curves. When the crossing intention is WSI, the pedestrian speed value is the second, in the middle of the three curves. When the intention is SWI, the pedestrian speed value is the smallest, which is at the bottom of the three curves. Generally speaking, as time goes by, there is no significant change in the value of pedestrian speed with WWI. The value of pedestrian speed whose intention is WSI drops rapidly. The pedestrian speed value with the intention of SWI shows a slow upward trend.

Figure 11(b) is a box diagram of pedestrian speed for pedestrians under different crossing intentions. When the crossing intentions are WWI, SWI, and WSI, the mean values of pedestrian speed are 4.27 km/h, 0.39 km/h, and 2.22 km/h. One-way ANOVA found that there were significant differences in pedestrian speed values under different intentions (F (2, 1977) = 2274.09 and $p < 0.001$), and the post-hoc test found that there were significant differences in pedestrian speed values after pairings with different intentions ($p < 0.001$).

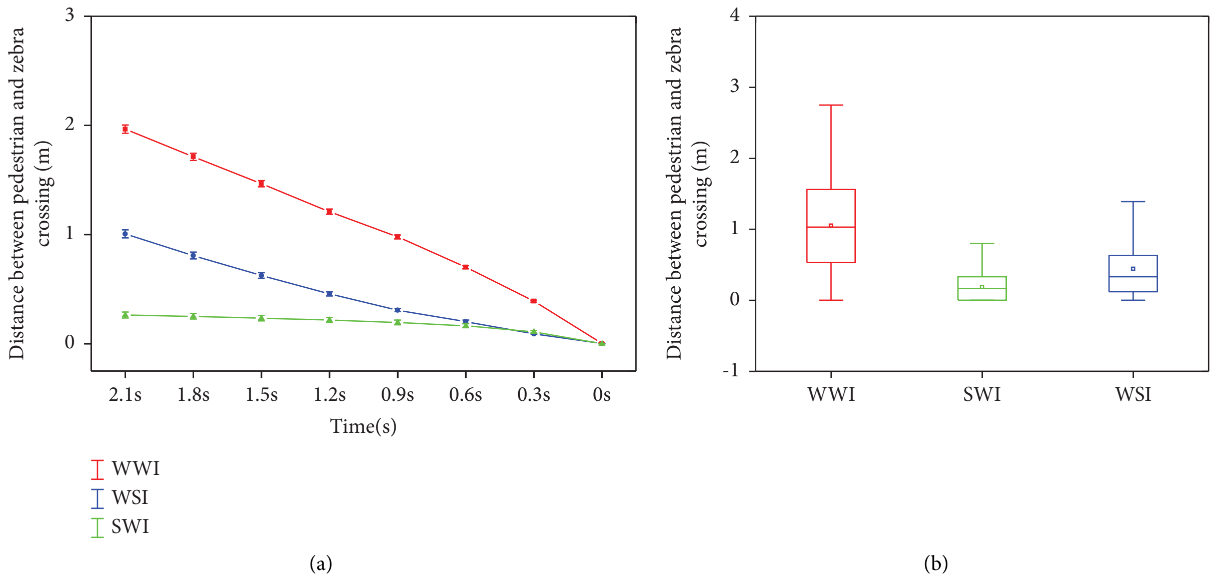


FIGURE 10: DPZC under different crossing intentions. (a) Line chart of DPZC change with time under different intentions. (b) Box diagram of DPZC under different crossing intentions.

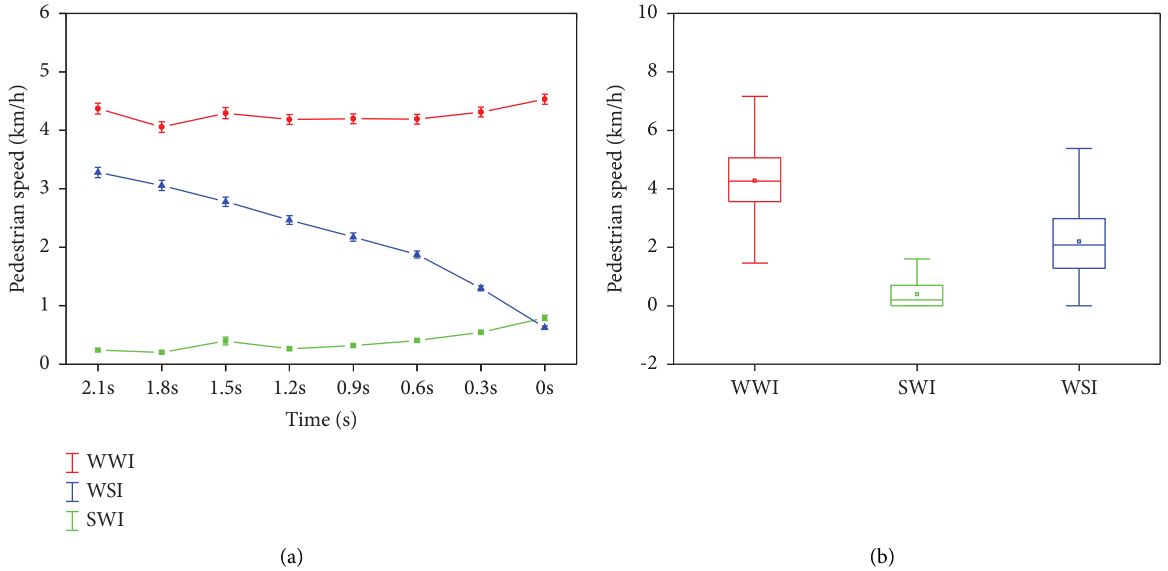


FIGURE 11: Pedestrian speed under different crossing intentions. (a) Line chart of pedestrian speed change with time under different intentions. (b) Box diagram of pedestrian speed under different crossing intentions.

4.1.5. Distance between Vehicle and Zebra Crossing. Figure 12(a) shows the DVZC changes under different crossing intentions within 2.1 s before crossing the zebra crossing. It can be seen that when the intention is WWI, the DVZC value when pedestrians cross the zebra crossing is the largest, which is at the top of the three curves. When the intention is SWI, the DVZC value is the second, in the middle of the three curves. When the intention is WSI, the DVZC value selected by pedestrians crossing the zebra crossing is the smallest, which is at the bottom of the three curves. Generally speaking, as time goes by, the DVZC value under different intentions shows a steady downward trend.

Figure 12(b) is a box diagram of DVZC for pedestrians under different crossing intentions. When the crossing intentions are WWI, SWI, and WSI, the mean values of DVZC are 49.28 m, 45.13 m, and 19.44 m. One-way ANOVA found that there were significant differences in DVZC values under different intentions (F (2, 1977) = 2247.65, $p < 0.001$), and

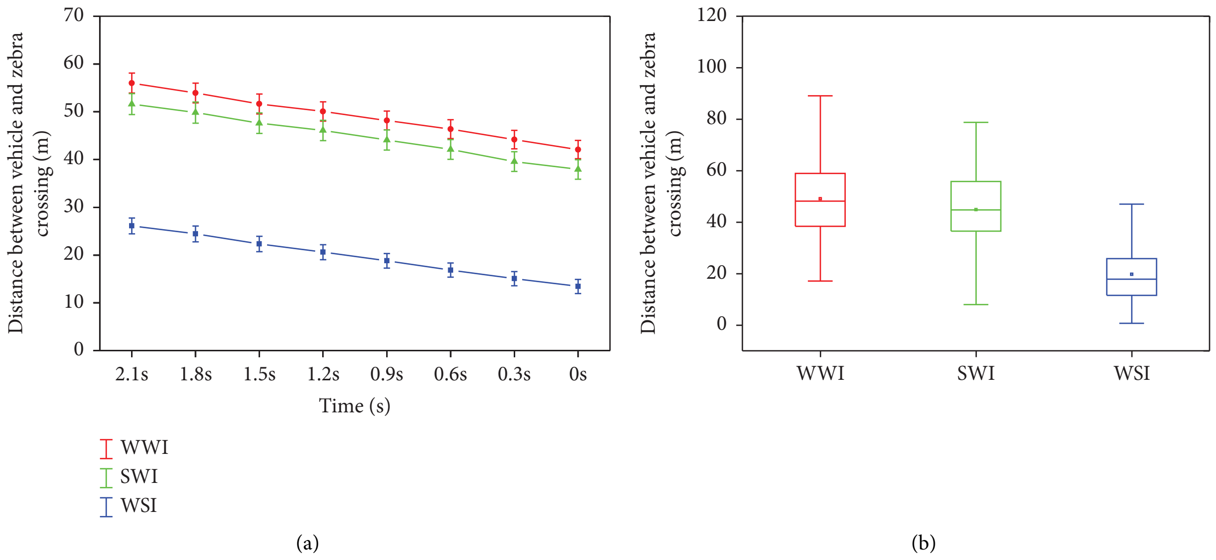


Figure 12: DVZC under different crossing intentions. (a) Line chart of DVZC change with time under different intentions. (b) Box diagram of DVZC under different crossing intentions.

Table 2: Number of intention samples.

Label	Train sample	Test sample
WWI	494	164
SWI	482	160
WSI	510	170

the post-hoc test found that there were significant differences in DVZC values after pairings with different intentions ($p < 0.001$).

4.1.6. Age and Gender. Numerous studies have shown that the age and gender of pedestrians have great differences in the choice of pedestrians to cross the zebra crossing. Generally speaking, men's choice of crossing the zebra crossing is relatively aggressive, and women's choice is relatively cautious [38, 39]. The ages of pedestrians are usually divided into young, middle-aged, and old. When crossing the zebra crossing, elderly pedestrians choose relatively cautiously, while middle-aged pedestrians choose more aggressively. Generally, 18–30, 30–59, and >59 are young, middle-aged, and old, respectively [40–42].

4.2. Model Results. Through the analysis in the previous chapter, the input parameter set of the model is finally determined, which includes TTC, DPZC, DVZC, vehicle speed, pedestrian speed, age, and gender. In this paper, a total of 1980 sets of valid data are selected, of which 75% are used as the training set, and the remaining 25% are used as the test set. The training set uses a five-fold cross-validation method. Table 2 shows the number of training samples and the number of test samples under different intentions. In this paper, the pedestrian crossing intention recognition models at 0 s, 0.5 s, and 1 s before crossing the zebra crossing are established, respectively. The performance of the model was evaluated by precision, recall, F1 score, confusion matrix, and receiver operating characteristic (ROC) curve.

4.2.1. Model Recognition Results at 0 s before Crossing the Zebra Crossing. Table 3 shows the model evaluation results when the model is 0 s before crossing the zebra crossing. Compared with several traditional machine learning algorithms, it is found that the pedestrian crossing intention model based on stacking ensemble learning has the highest recognition accuracy, reaching 98.79%. The precision, recall, and F1 score of this model for identifying WWI are 98.78%, 98.78%, and 98.78%, respectively. In the same way, the precision, recall, and F1 scores of the model for identifying SWI are 99.38%, 98.76%, and 99.07%, respectively. The precision, recall, and F1 scores of the model for identifying WSI are, respectively, 99.24%, 98.82%, and 98.53%. The comprehensive evaluation found that the pedestrian crossing intention model based on stacking-based ensemble learning introduced in this paper has the best recognition performance. The running time of the stacking model is 0.0083 s, and the running times of the AT-Bi-LSTM, LSTM, RF, and SVM models are 0.0032 s, 0.0054 s, 0.0065 s, and 0.0046 s, respectively. It can be seen that the running times of the above models are all in milliseconds, which can meet the actual needs.

Figure 13 shows the ROC curve of each model. It can be seen from the figure that when the false positive rate is 5%, the pedestrian crossing intention recognition model based on stacking ensemble learning has the highest true positive rate, followed by AT-Bi-LSTM, LSTM, RF, and SVM. Secondly, the area under the ROC curve based on the stacking ensemble learning method is the largest, which is higher than the other four algorithms. In addition, the ROC curves of the five algorithms are relatively far from the straight-line $y = x$, which shows that the recognition performance of the

Table 3: Model evaluation result at 0 s before crossing the zebra crossing.

Algorithm	Accuracy (%)	WSI			WWI			SWI		
		Pr (%)	Re (%)	F1 (%)	Pr (%)	Re (%)	F1 (%)	Pr (%)	Re (%)	F1 (%)
SVM	90.08	88.24	89.29	88.76	89.63	89.63	89.63	92.50	91.36	91.93
RF	92.12	90.59	91.67	91.12	92.07	92.07	92.07	93.75	92.59	93.17
LSTM	93.54	91.76	93.41	92.58	93.9	93.33	93.62	95.00	93.83	94.41
AT-Bi-LSTM	96.15	95.29	95.86	95.58	96.34	95.76	96.05	96.88	96.88	96.88
Stacking	98.79	98.24	98.82	98.53	98.78	98.78	98.78	99.38	98.76	99.07

Note. Pr represents precision, Re represents recall, and F1 represents F1 scores.

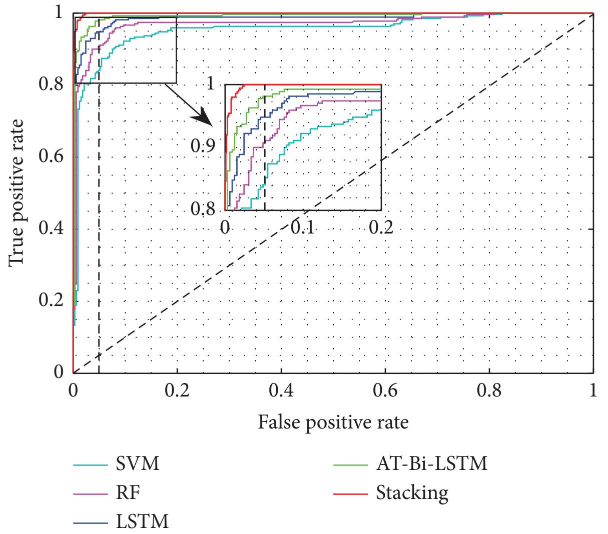


Figure 13: ROC curves of different models at 0 s before crossing the zebra crossing.

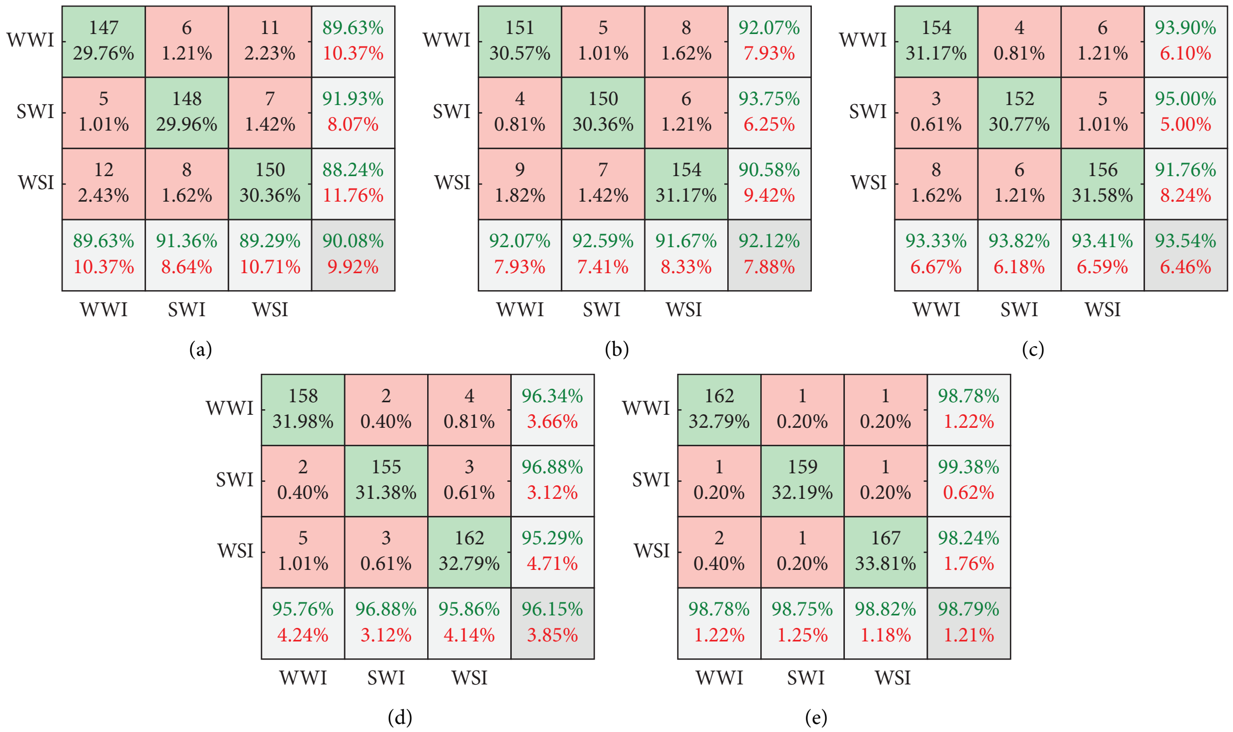
five models is better. A comprehensive comparison found that the performance of the pedestrian crossing intention recognition model based on stacking ensemble learning introduced in this paper is the best.

Figure 14 shows the confusion matrix of the five algorithms. It can be seen from the confusion matrix that the SVM-based intention recognition model has the most misrecognition times. The number of times that WWI is recognized as SWI and WSI is 6 and 11, respectively, and the times that SWI is recognized as WWI and WSI are, respectively, 5 and 7, and the number of times that WSI is recognized as WWI and SWI is 12 and 8, respectively. In contrast, the pedestrian crossing intention recognition model based on stacking integrated learning has the least number of misrecognitions and the best model performance. Among them, the times that WWI is recognized as SWI and WSI are 1 and 1, respectively, and the times that SWI is recognized as WWI and WSI are 1 and 1, respectively, and the times of WSI being recognized as WWI and SWI are 2 and 1, respectively.

4.2.2. Model Recognition Results at 0.5 s before Crossing the Zebra Crossing. Table 4 shows the model evaluation results when the model is 0.5 s before crossing the zebra crossing. Compared with several traditional algorithms, it is found that the intention recognition model based on stacking ensemble learning has the highest accuracy of 95.36%, the model recognition accuracy based on AT-Bi-LSTM is 92.12%, the model recognition accuracy based on LSTM is 89.30%, and the model recognition accuracy based on RF is 87.07%. The SVM-based model has the lowest recognition accuracy, which is 85.26%. It can be seen from Table 4 that the precision, recall, and F1 score of the pedestrian crossing intention model based on stacking ensemble learning are significantly higher than the other four algorithms. It can be seen that the stacking ensemble learning method introduced in this paper has the best recognition performance at 0.5 s before crossing the zebra crossing. Compared with Table 3, it can be seen that when the model is recognized at 0.5 s before crossing the zebra crossing, the accuracy has decreased to a certain extent. The main reason is that some key features contained in the sequence data have been deleted. However, in general, the accuracy of the model can still meet actual needs. The running time of the stacking model is 0.0076 s, and the running times of the AT-Bi-LSTM, LSTM, RF, and SVM models are 0.0027 s, 0.0060 s, 0.0061 s, and 0.0052 s, respectively.

Figure 15 shows the ROC curve of each model. It can be seen from the figure that when the false positive rate is 5%, the pedestrian crossing intention recognition model based on stacking ensemble learning has the highest true positive rate, followed by AT-Bi-LSTM, LSTM, RF, and SVM. Secondly, the area under the ROC curve based on the stacking ensemble learning method is the largest, which is higher than the other four algorithms. Compared with Figure 16, it can be seen that the area under the ROC curve corresponding to each algorithm has been reduced, and the performance of the model has begun to decline.

Figure 16 shows the confusion matrix of the five algorithms. It can be seen from the confusion matrix that the SVM-based intention recognition model still has the most misrecognition times. The number of times that WWI is recognized as SWI and WSI is 10 and 15, respectively, and the times that SWI is recognized as WWI and WSI are, respectively, 7 and 11, and the number of times that WSI is recognized as WWI and SWI is 19 and 11, respectively. In contrast, the pedestrian crossing intention recognition model based on stacking ensemble learning has the least number of misrecognitions and the best model performance. Among them, the times that WWI is recognized as SWI and WSI are 3 and 5, respectively, and the times that SWI is recognized as WWI and WSI are 2 and 5, respectively; the times of WSI being recognized as WWI and SWI are 7 and 3, respectively. Compared with Table 3, the number of misrecognition times has increased.



WWI	147 29.76%	6 1.21%	11 2.23%	89.63% 10.37%
SWI	5 1.01%	148 29.96%	7 1.42%	91.93% 8.07%
WSI	12 2.43%	8 1.62%	150 30.36%	88.24% 11.76%
	89.63% 10.37%	91.36% 8.64%	89.29% 10.71%	90.08% 9.92%
	WWI	SWI	WSI	

(a)

WWI	151 30.57%	5 1.01%	8 1.62%	92.07% 7.93%
SWI	4 0.81%	150 30.36%	6 1.21%	93.75% 6.25%
WSI	9 1.82%	7 1.42%	154 31.17%	90.58% 9.42%
	92.07% 7.93%	92.59% 7.41%	91.67% 8.33%	92.12% 7.88%
	WWI	SWI	WSI	

(b)

WWI	154 31.17%	4 0.81%	6 1.21%	93.90% 6.10%
SWI	3 0.61%	152 30.77%	5 1.01%	95.00% 5.00%
WSI	8 1.62%	6 1.21%	156 31.58%	91.76% 8.24%
	93.33% 6.67%	93.82% 6.18%	93.41% 6.59%	93.54% 6.46%
	WWI	SWI	WSI	

(c)

WWI	158 31.98%	2 0.40%	4 0.81%	96.34% 3.66%
SWI	2 0.40%	155 31.38%	3 0.61%	96.88% 3.12%
WSI	5 1.01%	3 0.61%	162 32.79%	95.29% 4.71%
	95.76% 4.24%	96.88% 3.12%	95.86% 4.14%	96.15% 3.85%
	WWI	SWI	WSI	

(d)

WWI	162 32.79%	1 0.20%	1 0.20%	98.78% 1.22%
SWI	1 0.20%	159 32.19%	1 0.20%	99.38% 0.62%
WSI	2 0.40%	1 0.20%	167 33.81%	98.24% 1.76%
	98.78% 1.22%	98.75% 1.25%	98.82% 1.18%	98.79% 1.21%
	WWI	SWI	WSI	

(e)

Figure 14: Confusion matrix at 0 s before crossing the zebra crossing: the cyan color indicates the number of correct recognitions and their proportion in all samples, and the light red color indicates the number of misrecognitions and their proportion in all samples. The rightmost column in the figure is precision, and the bottom column is recall. (a) Confusion matrix of SVM; (b) confusion matrix of LSTM; (c) confusion matrix of AT-Bi-LSTM; (d) confusion matrix of stacking.

Table 4: Model evaluation result at 0.5 s before crossing the zebra crossing.

Algorithm	Accuracy (%)	WSI			WWI			SWI		
		Pr (%)	Re (%)	F1 (%)	Pr (%)	Re (%)	F1 (%)	Pr (%)	Re (%)	F1 (%)
SVM	85.26	84.76	84.24	84.50	88.75	87.12	87.93	82.35	84.34	83.33
RF	87.07	86.59	86.59	86.59	90.63	88.41	89.51	84.12	86.14	85.12
LSTM	89.30	90.24	88.62	89.43	91.88	90.74	91.30	87.06	88.62	87.83
AT-Bi-LSTM	92.12	92.07	91.52	91.79	93.75	93.17	93.46	90.59	91.67	91.12
Stacking	95.36	95.12	94.55	94.83	96.88	96.27	96.57	94.12	95.24	94.67

Note. Pr represents precision, Re represents recall, and F1 represents F1 scores.

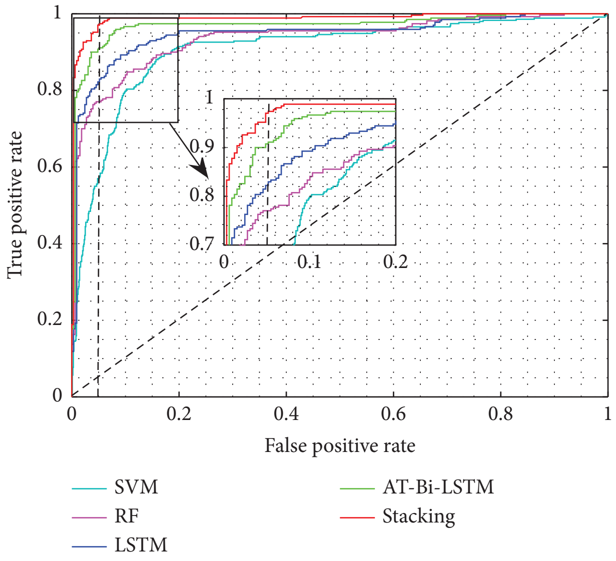


Figure 15: ROC curves of different models at 0.5 s before crossing the zebra crossing.

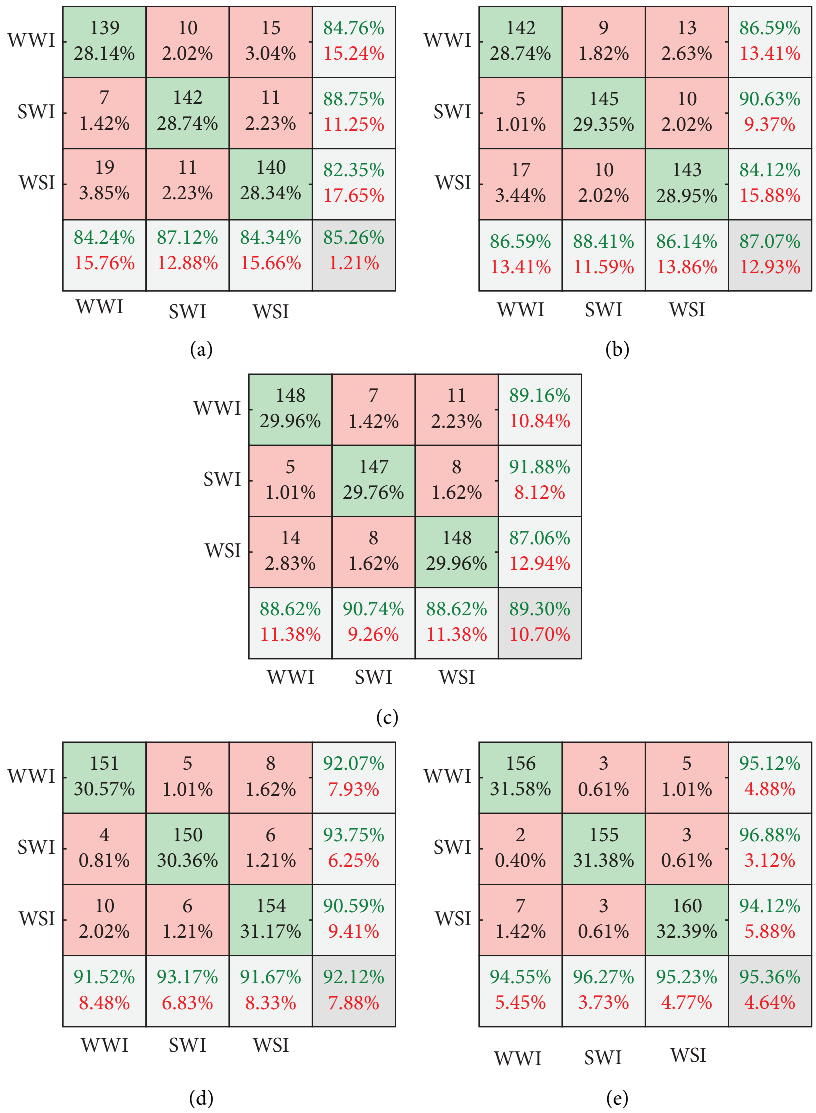


Figure 16: Confusion matrix at 0.5 s before crossing the zebra crossing: the cyan color indicates the number of correct recognitions and their proportion in all samples, and the light red color indicates the number of misrecognitions and their proportion in all samples. The rightmost column in the figure is precision, and the bottom column is recall. (a) Confusion matrix of SVM; (b) confusion matrix of LSTM; (c) confusion matrix of AT-Bi-LSTM; (d) confusion matrix of stacking.

Table 5: Model evaluation result at 1 s before crossing the zebra crossing.

Algorithm	Accuracy (%)	WSI			WWI			SWI		
		Pr (%)	Re (%)	F1 (%)	Pr (%)	Re (%)	F1 (%)	Pr (%)	Re (%)	F1 (%)
SVM	76.33	76.22	75.30	75.76	82.50	78.57	80.49	70.59	75.00	76.22
RF	78.35	78.05	77.58	77.81	83.75	80.72	82.21	73.53	76.69	78.05
LSTM	81.18	81.71	79.76	80.72	86.25	83.64	84.92	75.88	80.12	81.71
AT-Bi-LSTM	85.23	85.98	83.43	84.68	90.00	87.80	88.89	80.00	84.47	85.98
Stacking	89.27	89.63	88.02	88.82	93.13	90.85	91.98	85.88	88.48	89.63

Note. Pr represents precision, Re represents recall, and F1 represents F1 scores.

4.2.3. Model Recognition Results at 1 s before Crossing the Zebra Crossing. Table 5 shows the model evaluation results when the model is 1 s before crossing the zebra crossing. Compared with several traditional algorithms, it is found that the intention recognition model based on stacking ensemble learning has the highest accuracy of 89.27%, the model recognition accuracy based on AT-Bi-LSTM is 85.23%, the model recognition accuracy based on LSTM is 81.18%, and the model recognition accuracy based on RF is 78.35%. The SVM-based model has the lowest recognition

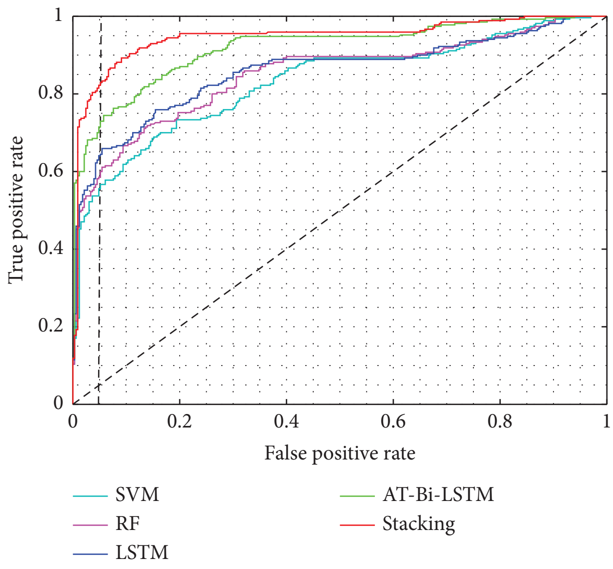
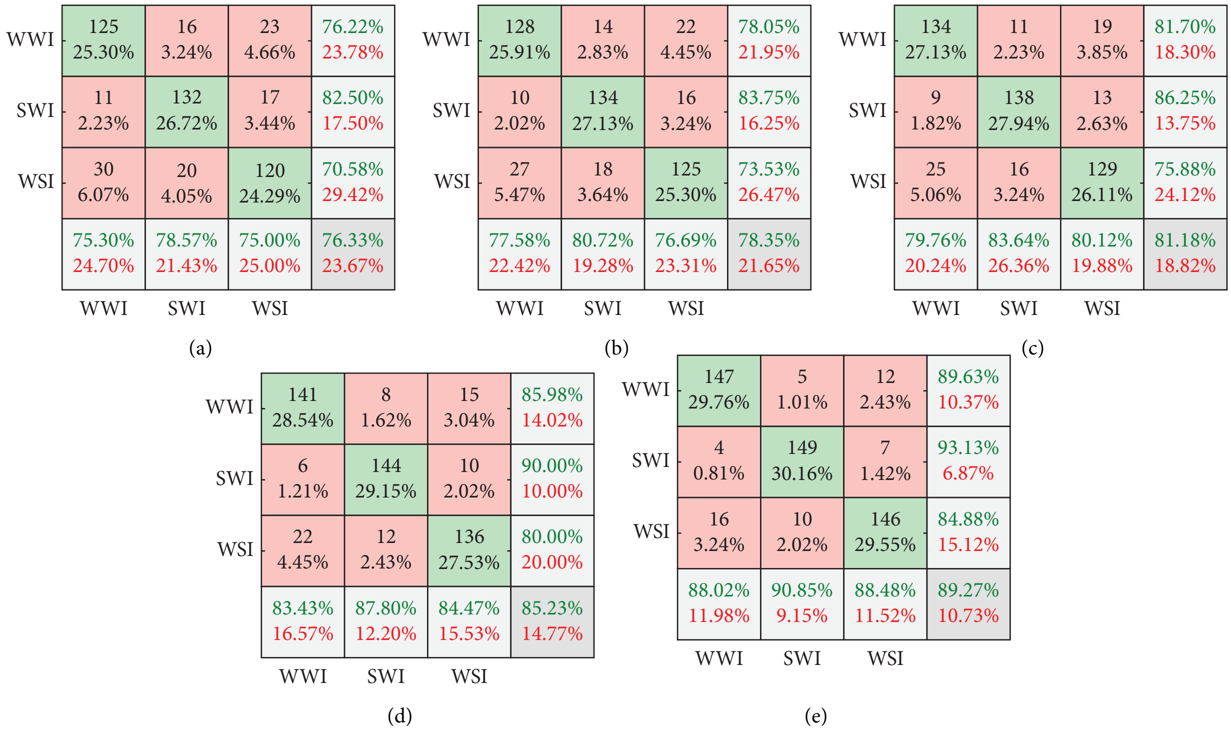


FIGURE 17: ROC curves of different models at 1s before crossing the zebra crossing.



	WWI	SWI	WSI	
WWI	125 25.30%	16 3.24%	23 4.66%	76.22% 23.78%
SWI	11 2.23%	132 26.72%	17 3.44%	82.50% 17.50%
WSI	30 6.07%	20 4.05%	120 24.29%	70.58% 29.42%
	75.30% 24.70%	78.57% 21.43%	75.00% 25.00%	76.33% 23.67%

(a)

	WWI	SWI	WSI	
WWI	128 25.91%	14 2.83%	22 4.45%	78.05% 21.95%
SWI	10 2.02%	134 27.13%	16 3.24%	83.75% 16.25%
WSI	27 5.47%	18 3.64%	125 25.30%	73.53% 26.47%
	77.58% 22.42%	80.72% 19.28%	76.69% 23.31%	78.35% 21.65%

(b)

	WWI	SWI	WSI	
WWI	134 27.13%	11 2.23%	19 3.85%	81.70% 18.30%
SWI	9 1.82%	138 27.94%	13 2.63%	86.25% 13.75%
WSI	25 5.06%	16 3.24%	129 26.11%	75.88% 24.12%
	79.76% 20.24%	83.64% 26.36%	80.12% 19.88%	81.18% 18.82%

(c)

	WWI	SWI	WSI	
WWI	141 28.54%	8 1.62%	15 3.04%	85.98% 14.02%
SWI	6 1.21%	144 29.15%	10 2.02%	90.00% 10.00%
WSI	22 4.45%	12 2.43%	136 27.53%	80.00% 20.00%
	83.43% 16.57%	87.80% 12.20%	84.47% 15.53%	85.23% 14.77%

(d)

	WWI	SWI	WSI	
WWI	147 29.76%	5 1.01%	12 2.43%	89.63% 10.37%
SWI	4 0.81%	149 30.16%	7 1.42%	93.13% 6.87%
WSI	16 3.24%	10 2.02%	146 29.55%	84.88% 15.12%
	88.02% 11.98%	90.85% 9.15%	88.48% 11.52%	89.27% 10.73%

(e)

FIGURE 18: Confusion matrix at 1s before crossing the zebra crossing: the cyan color indicates the number of correct recognitions and their proportion in all samples, and the light red color indicates the number of misrecognitions and their proportion in all samples. The rightmost column in the figure is precision, and the bottom column is recall. (a) Confusion matrix of SVM; (b) confusion matrix of LSTM; (c) confusion matrix of AT-Bi-LSTM; (d) confusion matrix of stacking.

accuracy, which is 76.33%. It can be seen from Table 5 that the precision, recall, and F1 score of the pedestrian crossing intention model based on stacking ensemble learning are significantly higher than the other four algorithms. It can be seen that the stacking ensemble learning method introduced in this paper has the best recognition performance at 1s before crossing the zebra crossing. Compared with Tables 3 and 4, it can be seen that when the model is recognized at 1s before crossing the zebra crossing, the accuracy has decreased. The main reason is that most of the key features contained in the sequence data have been deleted. However, the method introduced in this paper still has high recognition accuracy. The running time of the stacking model is 0.0094 s, and the running times of the AT-Bi-LSTM, LSTM, RF, and SVM models are 0.0035 s, 0.0059 s, 0.0071 s, and 0.0040 s, respectively.

Figure 17 shows the ROC curve of each model. It can be seen from the figure that when the false positive rate is 5%, the pedestrian crossing intention model based on stacking ensemble learning has the highest true positive rate, over 80%. The recognition accuracy of the remaining four algorithms has dropped significantly, and the corresponding value is less than 80%. Secondly, the area under the ROC curve based on stacking ensemble learning is the largest, which is higher than the other four algorithms. Compared with Figures 16 and 17, it can be seen that the area under the ROC curve corresponding to each algorithm has been reduced.

Figure 18 shows the confusion matrix of the five algorithms. It can be seen from the confusion matrix that the SVM-based intention recognition model has the most misrecognition times. In contrast, the pedestrian crossing intention recognition model based on stacking ensemble learning has the least number of misrecognitions and the best model performance. Compared with Figures 14 and 16, the number of misrecognition times has significantly increased.

5. Conclusions

This paper first collected the motion parameters of pedestrians and vehicles with laser radar and HD monitor and selected 1980 effective samples. Secondly, the statistical method is used to obtain the characteristic parameter set that can reflect the pedestrians' crossing intention. Finally, using the characteristic parameter set as the input of the stacking integrated learning method, a pedestrian crossing intention model with high recognition accuracy is trained and compared with traditional machine learning algorithms. The results show that the accuracy rate of the pedestrian crossing intention recognition model based on stacking ensemble learning is 98.79% when it is recognized at 0 s before crossing the zebra crossing. When it is recognized at 0.5 s before crossing the zebra crossing, the accuracy rate of the pedestrian crossing intention recognition model based on stacking ensemble learning is 95.36%. When it is recognized at 1 s before crossing the zebra crossing, the accuracy of the pedestrian crossing intention recognition model based on stacking ensemble learning is 89.27%. Compared with traditional machine learning algorithms, the method introduced in this paper has the best recognition performance. The method introduced in this paper has a high accuracy of intention recognition, which is of practical significance for future fully autonomous vehicles to effectively avoid human-vehicle conflicts and improve the efficiency of urban road driving.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Postdoctoral Startup Fund under grant number 522010 and in part by the National Natural Science Foundation Project, 52102465.

References

[1] C. Wang, H. Zhang, H. Wang, and R. Fu, "The effect of "yield to pedestrians" policy enforcement on pedestrian street crossing behavior: a 3-year case study in Xi'an, China," *Travel Behaviour and Society*, vol. 24, pp. 172–180, 2021.

[2] H. Zhang, Y. Guo, Y. Chen, Q. Sun, and C. Wang, "Analysis of pedestrian street-crossing decision-making based on vehicle deceleration-safety gap," *International Journal of Environmental Research and Public Health*, vol. 17, no. 24, p. 9247, 2020.

[3] Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China, *Annual Report of Road Traffic Accident Statistics of the People's Republic of China*, Beijing, 2020.

[4] Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China, *Annual Report of Road Traffic Accident Statistics of the People's Republic of China*, Beijing, 2019.

[5] Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China, *Annual Report of Road Traffic Accident Statistics of the People's Republic of China*, Beijing, 2018.

[6] Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China, *Annual Report of Road Traffic Accident Statistics of the People's Republic of China*, Beijing, 2017.

[7] Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China, *Annual Report of Road Traffic Accident Statistics of the People's Republic of China*, Beijing, 2016.

[8] J. Zhao, Y. Tang, and Y. Han, "Gap acceptance probability model for pedestrians at unsignalized mid-block crosswalks based on logistic regression," *Accident Analysis & Prevention*, vol. 129, pp. 76–83, 2019.

[9] J. Zhao, J. O. Malenje, J. Wu, and R. Ma, "Modeling the interaction between vehicle yielding and pedestrian crossing behavior at unsignalized midblock crosswalks," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 73, pp. 222–235, 2020.

[10] Us Department of Transportation, "Automated driving Systems2.0," *A Vision for Safety*, vol. 24, p. 57, 2017.

[11] Sae-China, "Driverless technology roadmap," *Safety Now*, vol. 67, p. 435, 2016.

[12] B. Yang and R. Ni, "Vision-based recognition of pedestrian crossing intention in an urban environment," in *Proceedings of the 2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems*, pp. 992–995, Suzhou, China, 29 July 2019 - 02 August 2019.

[13] S. Kalantarov, R. Riemer, and T. Oron-Gilad, "Pedestrians road crossing decisions and body parts movements," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 53, pp. 155–171, 2018.

[14] R. Mínguez, I. Alonso, D. Fernández-Llorca, and M. Sotelo, "Pedestrian path, pose, and intention prediction through Gaussian process dynamical models and pedestrian activity recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 5, pp. 1803–1814, 2018.

[15] R. Quintero, I. Parra, J. Lorenzo, D. Fernández-Llorca, and M. Sotelo, "A. Pedestrian intention recognition by means of a Hidden Markov Model and body language," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems*, pp. 1–7, Yokohama, Japan, 16-19 October 2017.

[16] R. Quintero, I. Parra, D. Llorca, and M. Sotelo, "Pedestrian path prediction based on body language and action classification," in *Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems*, pp. 679–684, Qingdao, China, 2010.

[17] Z. Fang and A. M. Lopez, "Intention recognition of pedestrians and cyclists by 2D pose estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4773–4783, 2020.

[18] R. D. Brehar, M. P. Muresan, T. Mariţa, C. C. Vancea, M. Negru, and S. Nedevschi, "Pedestrian street-cross action recognition in monocular far infrared sequences," *IEEE Access*, vol. 9, pp. 74302–74324, 2021.

[19] A. M. Căilean, C. Beguni, S. A. Avătămăniţei, M. Dimian, and V. Popa, "Design, implementation and experimental investigation of a pedestrian street crossing assistance system based on visible light communications," *Sensors*, vol. 22, no. 15, p. 5481, 2022.

[20] B. Völz, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "Inferring pedestrian motions at urban crosswalks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 544–555, 2019.

[21] F. Camara, N. Merat, and C. Fox, "A heuristic model for pedestrian intention estimation," in *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference*, pp. 3708–3713, Auckland, New Zealand, 27-30 October 2019.

[22] J. Zhao, Y. Li, H. Xu, and H. Liu, "Probabilistic prediction of pedestrian crossing intention using roadside LiDAR data," *IEEE Access*, vol. 7, pp. 93781–93790, 2019.

[23] H. Zhang, Y. Liu, C. Wang, R. Fu, Q. Sun, and Z. Li, "Research on a pedestrian crossing intention recognition model based on natural observation data," *Sensors*, vol. 20, no. 6, p. 1776, 2020.

[24] O. Ghori, R. Mackowiak, M. Bautista et al., "Learning to forecast pedestrian intention from pose dynamics," in *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium*, pp. 1277–1284, Changshu, China, 26-30 June 2018.

[25] A. Schulz and R. Stiefelhagen, "A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction," in *Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Gran Canaria, Spain, 15-18 September 2015.

[26] N. Brouwer, H. Kloeden, and C. Stiller, "Comparison and evaluation of pedestrian motion models for vehicle safety systems," in *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems*, Rio de Janeiro, Brazil, 01-04 November 2016.

[27] Y. Hashimoto, Y. Gu, M. Iryo-Asano, and S. Kamijo, "A probabilistic model of pedestrian crossing behavior at signalized intersections for connected vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 164–181, 2016.

[28] F. Schneemann and P. Heinemann, "Context-based detection of pedestrian crossing intention for autonomous driving in urban environments," in *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Daejeon, South Korea, 09-14 October 2016.

[29] T. Dietterich, "Ensemble learning," *The handbook of brain theory and neural networks*, vol. 2, no. 1, pp. 110–125, 2002.

[30] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[31] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, pp. 771–780, 1999.

[32] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[33] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[34] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[36] A. Graves, S. Fernández, and J. Schmidhuber, "Bidirectional LSTM networks for improved phoneme classification and recognition," in *International Conference on Artificial Neural Networks*, pp. 799–804, Springer, Berlin, Heidelberg, 2005.

[37] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 57, p. 758, 2017.

[38] Y. Pei and S. Feng, "Research on design speed of urban pedestrian crossing," *Journal of Highway and Transportation Research and Development*, vol. 23, no. 9, pp. 104–107, 2006.

[39] Y. Guo, "Road traffic safety and business management manual," *Science and Technology Press*, vol. 689, p. 26, 2002.

[40] K. V. R. Ravishankar and P. M. Nair, "Pedestrian risk analysis at uncontrolled midblock and unsignalised intersections," *Journal of Traffic and Transportation Engineering*, vol. 5, no. 2, pp. 137–147, 2018.

[41] X. Zhuang and C. Wu, "Modeling pedestrian crossing paths at unmarked roadways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1438–1448, 2013.

[42] J. Zhao, Y. Tang, and Y. Han, "Gap acceptance probability model for pedestrians at unsignalized mid-block crosswalks based on logistic regression," *Accident Analysis & Prevention*, vol. 129, pp. 76–83, 2019.