Journal of King Saud University - Computer and Information Sciences 36 (2024) 101916

HOSTED BY

Contents lists available at ScienceDirect

Journal of King Saud University - Computer and Information Sciences

journal homepage: www.sciencedirect.com

Full length article

Advancing speed limit detection in ADAS: A novel data-driven approach using Pareto-GBDTMO

Xu Luo [a,b], Fumin Zou [a,b,*], Qiang Ren [d], Sijie Luo [c], Feng Guo [a,b], Huan Zhong [a,b], Na Jiang [a,b], Xinjian Cai [a,b]

[a] Fujian Key Laboratory for Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou, 350118, Fujian, China
[b] Renewable Energy Technology Research Institute, Fujian University of Technology, Ningde, 352000, Fujian, China
[c] School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu, 611731, Sichuan, China
[d] Chongqing Key Laboratory of Public Big Data Security Technology, Chongqing College of Mobile Communication, Hechuan, 401527, Chongqing, China

ARTICLE INFO

Keywords:
Speed limit information recognition
Advanced driver assistance systems
Pareto
GBDT-MO
ETC date

ABSTRACT

Recognizing speed limit information is crucial for advanced driver assistance systems (ADAS) as it directly affects the safety planning and decision-making process of intelligent driving systems. However, traditional image recognition-based solutions confront inherent restrictions and precision issues due to uncontrolled external factors. This paper endeavors to present a novel, data-driven solution for speed limit information recognition that leverages the stability and maturity of data-driven technologies, overcoming these challenges. We introduce Pareto-GBDTMO, a cutting-edge method that synergistically blends Gradient Boosting Decision Trees for Multiple Output (GBDT-MO) and Fast Pareto Feature Selection (FPFS). This integration is instrumental in discerning salient features to direct and expedite the learning process of GBDT-MO. When coupled with Bayesian optimization, the feature set undergoes dynamic updates at each boosting iteration, allowing GBDT-MO to concentrate on the most prominent features. This adaptive, relevance-guided feature space regularization mechanism enhances the efficiency and precision of speed limit recognition. Fujian Province highway electronic toll collection (ETC) data is used for further validation, and the experimental results emphasize the effectiveness of our model, with an high accuracy of 97%, a low loss rate of 0.7%, and minimal latency. These findings affirm the feasibility and scientific validity of our data-driven approach, offering a reliable and redundant solution for speed limit information recognition in ADAS. This study not only contributes to the practical application of ADAS but also lays the groundwork for future large-scale lane-level data research

1. Introduction

Speed limit information plays a crucial role in the efficacy of Advanced Driver Assistance Systems (ADAS) (Xie et al., 2022), significantly contributing to road safety (Wu, 2022) and traffic management (Wu et al., 2020). Accurate recognition and interpretation of speed limit signs not only enhance compliance with traffic regulations (Dewi et al., 2021b) but also serve as a vital component in reducing the incidence of road accidents (Dewi et al., 2022; Dai et al., 2022). Despite the progress made in ADAS, the reliable detection and interpretation of speed limit information remain a challenging task, especially in varying environmental conditions. This study seeks to advance this field by exploring a novel data-driven approach for speed limit information recognition, aiming to address the challenges faced by traditional image-based methods.

Current approaches to speed limit information recognition primarily rely on image recognition-based technologies. While these methods have demonstrated success (Liu et al., 2021; Puli et al., 2023), they are not without significant challenges (Zhang et al., 2022). The most prominent issues include the accurate detection of salient targets in

* Corresponding author.
E-mail addresses: 2211801028@smail.fjut.edu.cn (X. Luo), fmzou@fjut.edu.cn (F. Zou), 2200601001@smail.fjut.edu.cn (Q. Ren), sjluo@std.uestc.edu.cn (S. Luo), n180310004@fzu.edu.cn (F. Guo), 2221308075@mail.fjut.edu.cn (H. Zhong), 2221901002@smail.fjut.edu.cn (N. Jiang), cxj2001@fjut.edu.cn (X. Cai).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2024.101916
Received 23 November 2023; Received in revised form 25 December 2023; Accepted 4 January 2024
Available online 20 January 2024

complex and dynamic road environments, as well as the recognition of small or distant targets. Furthermore, factors such as poor lighting, weather conditions, and obstructed views often lead to low-precision imaging, thus impeding the effectiveness of these systems. Acknowledging these challenges inherent in image recognition technologies, this study proposes a shift towards data-driven solutions. Capitalizing on the proliferation of big data in transportation, our approach seeks to offer a more robust and adaptable alternative, bypassing the limitations of traditional vision-based systems.

The advent of big data in transportation presents an opportunity to explore data-driven solutions. These solutions have the potential to circumvent the challenges posed by vision-based systems. Traditional methods, while effective in certain scenarios, struggle with the variability and unpredictability inherent in real-world driving conditions. Data-driven approaches, on the other hand, can leverage the vast amount of traffic and sensor data now available, offering a more adaptable and comprehensive framework for speed limit recognition (Rim et al., 2023). This shift in methodology, from a purely vision-centric to a data-centric model, forms the basis of our research, proposing a novel perspective in addressing the complexities of real-world traffic environments.

Our study delves into the realm of multi-label classification tasks (Gao and Zhou, 2021; Wang et al., 2021b; Huang et al., 2023; Ran et al., 2023), a concept not yet fully explored in the context of speed limit information recognition. We posit that the task of speed limit recognition is, in essence, a multi-label classification problem, where each speed limit sign can possess multiple attributes needing simultaneous identification. This perspective is relatively uncharted in transportation research and offers a new lens through which speed limit information can be analyzed. By framing the problem in this manner, we seek to bridge a critical gap in current research, utilizing the principles of multi-label classification to enhance the accuracy and efficiency of speed limit recognition systems.

Hence, this paper aims to develop a novel data-driven approach for speed limit recognition using multi-label classification. Unlike previous image-based solutions, our approach treats speed limit recognition as a multi-output learning problem to consider different speed limits on different roadways. We construct label-specific features from driving behavior and supplement road attributes as common features. To improve the convergence of the training model, we develop a fast Pareto feature selection technique that identifies the most relevant and non-redundant features to guide the learning process of the training model. Bayesian optimization further adjusts the subset of features used in each boosting iteration to achieve optimal regularization, this mechanism improves the efficiency and accuracy of speed limit information recognition.

The principal contributions of this paper are:

1. Proposing a data-driven recognition solution for speed limit information, integrating the advantages of data-driven methodologies to provide more accurate and efficient sensing information for ADAS.
2. Introduce an innovative multi-label classification framework specifically tailored for transportation research, pioneering new methodologies to address complex, layered speed limit data, thereby filling a significant research gap in the field.
3. Addressing the slow convergence speed issue of GBDT-MO, we enhance the GBDT-MO model by applying the Pareto advantage-based feature selection technique, thereby accelerating convergence speed.

The rest of this paper is organized as follows: Section 2 reviews the related research on speed limit information recognition and multi-label classification methods, particularly focusing on the limitations of traditional speed limit information recognition solutions and the current modeling techniques in multi-label classification. In Section 3, our method and model are described in detail. Section 4 outlines our experimental design and execution. In Section 5, we analyze the experimental results and provide a thorough discussion. Finally, Section 6 summarizes our study and proposes future research directions.

2. Literature review

Speed limit information recognition, a critical technology for ADAS, is essential for decision-making in intelligent driving systems. Its importance has been highlighted with the recent advancements in intelligent driving, consequently attracting extensive research attention. Although traffic signs bearing speed limit information have traditionally been the focus of image recognition technologies, this paper aims to proposes a data-driven recognition scheme. This scheme, based on a multi-output learning mechanism, provides safety redundancy for advanced assisted driving systems. In light of this, we present a detailed review of speed limit information recognition research and multi-label classification methods.

2.1. Speed limit information recognition study

The challenge of detecting and recognizing speed limit signs during rapid vehicle travel has prompted extensive research in the field of machine vision. Key areas of investigation include detection speed, recognition accuracy, high-quality data generation, and salient object detection in complex environments.

Detection speed is essential to accommodate the real-time requirements of high-speed vehicular environments. For instance, (Bayoudh et al., 2021) developed a hybrid traffic sign recognition model, Hybrid-TSR, that leverages pre-trained deep 2D CNNs and shallow 3D CNNs to reduce network complexity and improve training speed. Wang et al. (2023) enhanced the Yolo v5 architecture using the AF-FPN structure, thereby improving the feature map's information extraction capability and the detection of multi-scale objects.

Recognition accuracy, a necessary capability for advanced network architectures, is a heavily researched topic in various application scenarios. Bi et al. (2021) designed a compact traffic sign classification network based on the VGG-16 model to ensure real-time performance with high accuracy. Xie et al. (2022) developed a federated learning paradigm spiking neural network (FedSNN) aimed at data privacy protection and meeting the high accuracy requirements in in-vehicle environments.

High-quality data generation is the answer to inconsistent design standards between different countries and difficult data access. Dewi et al. (2021a) tackled this problem by using multiple GANs to synthesize training data, validated through Yolo v4. The results demonstrated that hybrid image data enhances the network's recognition performance. Similarly, (Dewi et al., 2022) utilized Deep Convolutional Generative Adversarial Network (DCGAN) to validate this idea i.e., the mixing of raw and synthesized images enhances the recognition performance of the model.

The challenge of detecting small targets in complex environments has also been explored. Wang et al. (2021c) proposed using low-level information to enrich the feature hierarchy of the network and enhance the feature information characterization of small targets. This was validated using Yolo V4-tiny's lightweight traffic sign recognition algorithm. Subsequently, Dewi et al. (2021b) improved the Yolo V3 architecture by using the SPP principle and designed a robust detection method for small traffic sign recognition.

These efforts have been instrumental in field development, adapting to different application scenarios, and meeting various needs. However, the limitations have also been exposed. For instance, Bi et al. (2021) and Chen et al. (2022) emphasized that high-quality data is necessary for high-performance models, and the lack thereof results in performance degradation. Factors such as lighting conditions, motion blur, occlusion, and long-distance target recognition pose challenges that still

need to be overcome (Zhu and Yan, 2022). Thus, it is necessary to develop another recognition solution to enhance ADAS's safety redundancy, akin to LiDAR application in smart cars, using a multi-source datas fusion solution.

In the era of information technology, data has become a precious resource. Data-driven solutions have found applications in fields like intelligent transportation, fault diagnosis, and automation control, due to their reliability, efficiency, and stability. Unlike the uncontrollable nature of image recognition data collection, IoT device data collection tends to be controlled and stable. Although anomalies datas exist, but data-driven core solutions have mature technology architectures and have achieved considerable success in predicting traffic flow (Ma et al., 2021), location (Liu et al., 2023a), and travel time (Chiabaut and Faitout, 2021). Therefore, the inclusion of data-driven solutions to overcome vision-based solution limitations is a worthwhile exploration.

2.2. Research on multi-label classification methods

Multi-label classification (MLC) solutions have primarily been guided by research centered around methodology and data characterization. Some researchers focus on developing state-of-the-art algorithms, while others prioritize enhancing input space performance before applying established algorithms. Both approaches contribute significantly to model performance improvement.

Two central ideas for solving MLC tasks are problem transformation and algorithmic adaptation (Bogatinovski et al., 2022). Problem transformation converts an MLC problem into a traditional binary or multi-class classification problem, leveraging the mature technology of traditional classification. Methods like Binary Relevance (BR), Classifier Chaining (CC), and Label Power (LP) belong to this approach. The difference lies in whether and how label relevance is considered. Algorithmic adaptation, on the other hand, is a design strategy that enables an algorithm to automatically adjust its behavior to adapt to changes in input data or environment.

Several representative methods have been proposed. For instance, Wang et al. (2021a) proposed an active K-label set integration learning paradigm (ACkEL) based on the LP mechanism, capturing information about irrelevant class label subsets from the feature space. Xia et al. (2021) proposed a weighted classifier selection and stacked integration multi-label classification approach (MLWSE), constructing label-specific features and adjusting the weights of base classifiers considering label relevance. Zhang and Jung (2020) developed GBDT-MO, an adaptive algorithm for Gradient Boosting Decision Tree (GBDT) with a multi-output learning mechanism, which considers the correlation between labels.

Moreover, significant research has focused on label-specific features, which are most directly related to the labels. For instance, (Zhang and Li, 2021) proposed a multi-label learning method called LF-LELC, constructing label-specific features for each label, exploring the correlation between labels, and finally building a classification model for each label. Liu et al. (2023b) proposed a new Robust Label and Feature Space Cooperative Learning approach (RLFSCL), using feature learning techniques in both feature space and label space to address high-dimensional data samples' correlation and labels.

For high-dimensional output scenarios, (Liu et al., 2022) proposed a method, MoRE, which learns the low-rank structure of the residuals between the input and output space, reducing the label space dimensionality. Li et al. (2022) argued that focusing only on label-specific features and label correlation may be sub-optimal because common features are also important. Thus, they proposed a new method, CLML, for learning common and label-specific features for multi-label classification using correlation information. Huang et al. (2023) constructed label-specific features directly from social networks and utilized the classical ML-KNN model for training, achieving state-of-the-art performance. This further validates the effectiveness of label-specific features. Paul et al. (2021) proposed a three-stage multi-label online feature selection method, MMOFS, based on the particle swarm optimization technique, acknowledging that real-life data transmission is dynamic. Yu and Zhang (2021) argued that the decoupled nature of the two-stage strategy may result in a sub-optimal model and proposed a wrapper learning approach that jointly performs label-specific feature generation and classification model generalization. Hashemi et al. (2021) modeled the problem of multi-label feature selection as a bi-objective optimization problem concerning feature relevance and redundancy, processed further by taking advantage of Pareto.

While these advances are noteworthy, most research has been limited to basic theory, with little progress in the transportation domain, probably due to the uniqueness of transportation data. This paper aims to fill these research gaps by developing a customized multi-label learning approach for data-driven speed limit identification. By integrating user driving behavior with road attributes for a specific road segment, the proposed integrated Pareto-GBDTMO method achieves excellent performance. Comparison with existing multi-label algorithms further highlights the benefits of specialization in this problem domain.

3. Methodology

3.1. Problem description

The task of speed limit recognition can be framed as a multi-label classification problem. In this study, we examine a traffic dataset $D = \{(\mathbf{x}_i, \mathbf{y}_i)_{i=1}^n\}$ containing n samples, where each $\mathbf{x}_i \in \mathbb{R}^m$ represents a road segment described by an m-dimensional feature vector. Thus, we define $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ where each $\mathbf{x}_i$ is a feature vector corresponding to a road segment. Concurrently, $Y = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n] \in \{0,1\}^{n\times d}$ represents the set of speed limit labels for these n samples, each being a d-dimensional label vector. In this context, a single road segment $\mathbf{x}_i$ can have multiple labels $\mathbf{y}_i$, indicative of different speed limit standards applicable to various vehicle types. The primary objective of this study is to discover a mapping function $f : \mathbb{R}^m \to \mathbb{R}^d$ that associates each input segment $\mathbf{x}_i$ with a corresponding set of output labels $f(\mathbf{x}_i) \subseteq Y$. Given that speed limits are determined by road attributes and prevailing traffic conditions, function f aims to effectively model the correlation between input features and output labels. Multi-label classification, unlike single-label classification, can predict multiple labels for each input instance, making it particularly apt for speed limit recognition where various standards may apply to the same road segment. This approach is more versatile and retains more information than enforcing a singular speed limit output for each road segment.

3.2. Multi-value coding for speed limit labels

The Design Specification for Highway Speed Limit Signs (JTG T 3381-02-2020) categorizes the speed limit standard of highways into four levels: 120 km/h, 110 km/h, 100 km/h, and 80 km/h. Speed limits are set according to spatial, temporal, and combined modes. Spatial modes include single-speed limits, speed limits for specific vehicle types, and lane-specific speed limits. Temporal methods encompass time-dependent and weather-dependent speed limits, while combined methods merge lane-specific and vehicle-specific speed limits. Considering the characteristics of our dataset, this study primarily focuses on single speed limits and speed limits by vehicle type set in a spatial method. Single-speed limits apply a uniform speed limit to all vehicles in a specific area, while vehicle-type speed limits set different speed limits for the main types of vehicles in the area.

Factors such as the road's design standard, road surface condition, vehicle performance, and safety distance are taken into account while setting speed limits. This implies that different road segments may have varying speed limits. Additionally, due to the diverse nature of different vehicle types, distinct speed limit standards may be applicable to different vehicle types on different road segments. Traditional classification tasks can only recognize a single speed limit, losing the speed limit

Table 1
The case of multi-valued code conversion for different speed limit standards.

Speed limit information	Value	80	100	110	120
Single speed limit	120	0	0	0	1
Single speed limit	100	0	1	0	0
Speed limit by vehicle type	120/100	0	1	0	1
Speed limit by vehicle type	110/100	0	1	1	0

information for different vehicle types and ignoring label connections. This limitation is why we chose the multi-value encoding approach, which allows instances to have more than one activation value. If a segment has a single speed limit, it has only one activation value; if it has a vehicle-type speed limit, it has multiple activation values.

Single speed limits encompass four categories [80, 100, 110, 120], and vehicle-type speed limits contain three categories [(100/80), (110/100), (120/100)], partial examples are shown in Table 1. The larger value in the vehicle-type speed limit indicates the speed limit for small passenger vehicles, while the smaller value indicates the speed limit for other vehicles. Unlike traditional one-hot encoding, multi-value encoding allows for multiple activation values in a single row rather than restricting it to one activation value per row. One-hot encoding suits traditional classification problems where each sample can only belong to one class, but highway segments may have multiple speed limit standards concurrently. Although binary coding could also apply to this scenario, it could be less intuitive and might require more bits for more complex combinations.

3.3. Speed limit information feature vector construction

The gantry transaction data records the complete travel trajectory of a vehicle, including key information such as the gantry nodes passed through, the transaction time through the gantry, etc., whereas two neighboring gantries can form a separate independent segment. Thus, the user's complete travel trajectory can be composed of either consecutive gantry nodes or connected segments. The difference between the transaction times of two neighboring gantries is the user's travel time through the segment, and the ratio of segment distance to travel time is the vehicle's travel speed through the segment. Through this series of operations, we can then complete the construction of the user's driving behavior (Afghari et al., 2023).

Extracting label-specific features

Label-specific features are those most pertinent to each label. Since different road segments may have distinct speed limits that apply to various vehicle types, we perform clustering based on vehicle type and extract driving behavior indicators accordingly. For instance, for small passenger vehicles, the driving behavior metrics include maximum, minimum, and average travel speeds, standard deviation of travel speed, and the 85th percentile, 15th percentile, 75th percentile, 25th percentile, and median (50th percentile) of travel speeds. Considering the varying speed limits, we classify vehicle types into small passenger vehicles (car2), small freight vehicles (car5), large freight vehicles (car4), other types of vehicles (car3), and a category that includes all types (car1).

Extracting label common features

Common label features are those that potentially contribute to information gain across all labels. In our study, road information meets this criterion. Therefore, we consider segment length, service area construction, traffic flow composition, and time-domain features as common features. The time-domain feature refers to the average travel speed of all vehicles from 6:00 am to 6:00 pm. Traffic flow is counted based on vehicle type, enabling us to understand the traffic composition of each segment.

By considering the heterogeneity of road speed limit information, we extract label-specific features from users' natural driving behaviors of multiple vehicle types and enhance label-common features with attribute information of roads. Since then, the construction of feature vectors is completed and applied to downstream training and prediction tasks.

3.4. Fast-Pareto feature selection algorithm

The objective of the feature selection method is aimed at selecting a set of features that best reveal the salient characteristics of the data, with the advantage of improving efficiency without impairing model performance. Considering that each sample in this study contains numerous features, there may be a lot of irrelevant or redundant information among these features. Thus, inspired by the efficient optimization method based on Pareto dominance proposed by Hashemi et al. (2021), we model the multi-label feature selection issue as a bi-objective optimization problem concerning feature relevance and redundancy. Due to the high computational complexity of the original non-dominated sorting method, we adopt a well-established fast non-dominated sorting method (Deng et al., 2022) (FNS) to enhance the algorithm's performance and expedite the feature selection process. The steps are as follows:

Correlation calculation: we adopt ridge regression (RR) to obtain the linear relationship between the feature set and the target variable. At the same time to prevent overfitting, RR introduces a regularization term, which can effectively reduce the effect of multicollinearity among features. The calculation is as follows:

$$W = \left(X^T X + \lambda I\right)^{-1} X^T Y \tag{1}$$

Where X denotes the input data, Y is the target variable, λ is the regularization hyperparameter, and I denotes the identity matrix.

Our algorithm aims to optimize two objectives: objective one, maximize the importance of each feature ($F1$), which is expressed by the absolute maximum weight of the ridge regression solution; and objective two, minimize the redundancy between features ($F2$), which is measured by calculating the average pairwise Euclidean distance (ED) between feature weights.

$$F1 = \max(W, axis = 1) \tag{2}$$

$$ED = \sqrt{\sum_{i=1}^{m} \left(W_i - W_j\right)^2} \tag{3}$$

$$F2 = \text{mean}(FD, axis = 1) \tag{4}$$

Where m denotes the feature dimension. ED represents the Euclidean distance between pairs of features, while FD refers to the feature distance, which is defined by a feature distance matrix composed of ED values. The subscript i and j refer to different features in the context of the W matrix.

The Fast Undominated Sort algorithm (Deng et al., 2022) hierarchizes the solution set by grouping undominated solutions at each level. It identifies non-dominated solutions and assigns them to the appropriate level, subsequently removing them from the set. This algorithm performs a fast non-dominated sorting based on two objectives, effectively dividing solutions into non-dominated frontiers. A solution's rank, designated as the Pareto Number (PN), corresponds to the frontier it belongs to. The PN is a crucial metric in this process; it quantifies the level of dominance (or non-dominance) of a solution within the solution space. Solutions with a lower PN are considered superior, indicating fewer conflicting objectives and a closer approximation to the Pareto front. This rank is dynamically assigned as the algorithm progresses through the solution set, categorizing each solution based on its relative performance compared to others.

Calculation of Crowding Distance (Hashemi et al., 2021): The crowding distance is a metric utilized in multi-objective optimization

to assess the density of solutions within the objective space. Its purpose is to maintain diversity among solutions in the Pareto frontier. In the context of multi-objective feature selection, the crowding distance is employed to measure the density of features in a dual objective space defined by correlation and redundancy measures. It enables the differentiation of features with similar correlation and redundancy values. The calculation of the crowding distance involves comparing the objective values of neighboring features. A larger crowding distance indicates a greater distance from neighboring elements, implying that the feature provides unique and valuable information. Therefore, within each non-dominated frontier, we calculate the crowding distance to evaluate the proximity of a solution to its neighbors. The crowding distance is computed as the sum of the normalized distances in the two objective functions.

$$\mathrm{d}_i = (F1_{i+1} - F1_{i-1}) + (F2_{i+1} - F2_{i-1}) \tag{5}$$

Combined Ranking and Crowding Distance: To incorporate both the Pareto Frontier Number (PN) and Crowding Distance into a single metric, we devised a combined score. The PN represents the feature's position within the Pareto Frontier, with lower PN values indicating better positions. On the other hand, the Crowding Distance (d) measures the density of features in the target space, quantifying their isolation from neighboring elements. A larger crowding distance suggests that a feature provides unique and valuable information. By integrating PN and d in the score calculation, our algorithm assigns scores to features, considering their position in the Pareto front and their density in the target space. This approach not only promotes the selection of diverse features but also ensures the inclusion of valuable and distinctive information for multi-label classification tasks.

$$R = PN + \frac{1}{1+d} \tag{6}$$

Sorting by combination scores: we sort the features based on the combination scores to produce an ordered list of feature indexes that optimally balances feature importance and redundancy. In conclusion, the proposed algorithm provides a robust and effective approach for feature selection in multi-label classification problems, enabling researchers to identify the most informative and non-redundant features in high-dimensional datasets.

In this study, we propose a multi-objective feature selection algorithm for multi-label classification problems that combines ridge regression with the concepts of Pareto optimality and crowding distance. The aim is to select a set of optimal features while maximizing the importance of each feature and minimizing the redundancy between them. The pseudo-code of our algorithm is summarized below:

Algorithm 1 Fast-Pareto Feature Selection Algorithm

Require: Feature data matrix X, label data matrix Y
Ensure: Ranking vector $Rank$

$Rank = \emptyset$
$\lambda = 10$
$W = (X * X^T + \lambda I)^{-1} * X^T * Y$ ▷ Calculate the W matrix using Eq. (1)
$F1 = \max(W, 1)$ ▷ Calculate the maximum value of each row of W
Initialize the distance matrix FD to store Euclidean distance between each pair of features
for $i = 1$ to m **do**
  for $j = 1$ to m **do**
    $FD(i, j) = ED(W(:, i), W(:, j))$ ▷ $W(:, i)$ is the ith column in W
  end for
end for
$F2 = \mathrm{mean}(FD, 1)$ ▷ Calculate the average values of each row
$Fe = \{1, 2, \ldots, m\}$ ▷ feature indexes
$k = 1$ ▷ counter
while $Fe \neq \emptyset$ **do** ▷ Perform Fast Non-dominated Sorting using $F1$ and $F2$ as two objective functions
  $[S, front] = FastNonDominatedSort(F1, F2, Fe)$ ▷ Fast non-dominated sort
  for feature in S **do**
    $PN[\text{feature}] = k$ ▷ Non-dominated feature subset
  end for
  $k = k + 1$
  $Fe = Fe - front$ ▷ Remove the non-dominated features from Fe
end while
$d = CalculateCrowdingDistance(front, F1, F2)$ ▷ Calculate the crowding distance of each feature
$R = PN + \frac{1}{1+d}$
$Rank$ = Sort the features based on their values in R in ascending order

3.5. Gradient boosting decision tree with multi-output algorithm

Multi-output gradient boosted decision trees (GBDT-MO) extend traditional GBDTs to handle multiple prediction targets (Zhang and Jung, 2020). As a supervised learning method, GBDT-MO retains key advantages of decision trees while improving performance through boosting: (i) they have decision mainfolds with approximate hyperplane boundaries, which is efficient for tabular data; (ii) they are more interpretable by keeping track of their decision nodes. We define a functional equation f that represents a method for learning a multi-output GBDT in which the leaves of the tree construct predictions for all variables or a subset of automatically selected variables. The function f takes an input vector x and maps it to the corresponding output variables using the mapping function $q(x)$.

$$f(\mathbf{x}) = \mathbf{W}_{q(\mathbf{x})}, \quad q : \mathbb{R}^m \to [1, L], \quad \mathbf{W} \in \mathbb{R}^{L \times d} \tag{7}$$

Since GBDT-MO is an additive model in which each tree contributes to the final prediction by summing its own prediction with those of the previous trees, the prediction of the tth tree is represented by Eq. (8):

$$\hat{\mathbf{y}}_i = \sum_{k=1}^{t} f(\mathbf{x}_i) \tag{8}$$

Traditional strategies aim to learn multiple output GBDTs by constructing a separate tree for each output variable through the idea of problem transformation, but this strategy may result in a learned tree structure that is redundant and ignores the correlations between the output variables. To address this problem, the method constructs each leaf of the GBDT-MO to predict a subset of all output variables or automatically selected variables and simultaneously consider the correlations between them. This is achieved by considering the sum of the target returns of all output variables, which ensures that the tree is optimized for all variables simultaneously. It represents the objective function used in the multi-output approach for learning gradient-enhanced decision trees. We consider the $(t+1)$-th tree given $\hat{y}$, and since each leaf is separable, the objective for a single leaf is as follows:

$$\mathcal{L} = \sum_i l(\hat{\mathbf{y}}_i + \mathbf{w}, \mathbf{y}_i) + \lambda \mathcal{R}(\mathbf{w}) \tag{9}$$

The objective function, denoted by $\mathcal{L}$, is the sum of all the training instances applied to the predicted and true outputs. Where the predicted output is a vector containing the predicted values of all output variables, and $R(W)$ is a vector of weights learned during training. The loss function is regularized by a penalty term that encourages the weights to become smaller, where λ is a hyperparameter that controls the strength of the regularization. Assuming that the loss function l is a second-order differentiable function, which can be approximated by the second-order Taylor expansion of and setting $R(w) = \frac{1}{2}\|\mathbf{w}\|_2^2$, Eq. (9) can be derived from Eq. (10):

$$\mathcal{L} = \sum_i \left\{ l(\hat{\mathbf{y}}_i, \mathbf{y}_i) + (\mathbf{g})_i^T \mathbf{w} + \frac{1}{2}\mathbf{w}^T (\mathbf{H})_i \mathbf{w} \right\} + \frac{\lambda}{2}\|\mathbf{w}\|_2^2 \tag{10}$$

where $(\mathbf{g})_i = \frac{\partial l}{\partial \hat{\mathbf{y}}_i}$, $(\mathbf{H})_i = \frac{\partial^2 l}{\partial \hat{\mathbf{y}}_i^2}$. Overall, the objective function aims to minimize the sum of the loss function, the contribution of each leaf node, the curvature of the loss function and the regularization term. To avoid symbols conflicting with the subscripts of vectors or matrices, the symbol $(\cdot)_i$ is used to indicate that the object belongs to the ith sample. If there is no ambiguity, this symbol is omitted. By setting $\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{0}$, we obtain the optimal leaf values:

$$\mathbf{w}^* = -\left(\sum_i (\mathbf{H})_i + \lambda \mathbf{I}\right)^{-1} \left(\sum_i (\mathbf{g})_i\right) \tag{11}$$

The optimal leaf values are computed using the inverse matrix of the matrix, which is used to update the predictions of the GBDT-MO model at each iteration of the boosting process. Substituting w^* into eq(10) and ignoring the constant term $l(\hat{\mathbf{y}}_i, \mathbf{y}_i)$, the optimal objective is as follows:

$$\mathcal{L}^* = -\frac{1}{2}\left(\sum_i (\mathbf{g})_i\right)^T \left(\sum_i (\mathbf{H})_i + \lambda \mathbf{I}\right)^{-1} \left(\sum_i (\mathbf{g})_i\right) \tag{12}$$

As a result, we obtain the optimal leaf value and optimal objective for multiple outputs. After determining the computation of the optimal leaf value and optimal objective, the histogram is constructed using the histogram optimization algorithm and the gradient statistic of all samples i. Further, we compute the sum $G_t(k, j)$ and the sum of squares $H_t(k, j)$ of the gradient statistics for each bin k in the histogram with the following equation:

$$G_t(k, j) = \sum_{i \in \text{bin } k} g_t(i, j) \tag{13}$$

$$H_t(k, j) = \sum_{i \in \text{bin } k} g_t(i, j)^2 \tag{14}$$

Compute the segmentation gain at bin k. For each bin k, we can compute the segmentation gain $Gain_t(k, j)$ at that location, which is calculated using the sum of the gradient statistics $G_t(k, j)$, the sum of squares $H_t(k, j)$, and a regularization parameter λ. The formula is as follows:

$$Gain_t(k, j) = \frac{1}{2}\left[\frac{G_t(k, j)^2}{H_t(k, j) + \lambda} - \frac{G_t(k, j-1)^2}{H_t(k, j-1) + \lambda} - \frac{G_t(k, j+1)^2}{H_t(k, j+1) + \lambda}\right] \tag{15}$$

Find the bin k^* with the largest segmentation gain and compute the segmentation threshold $s_t(j)$:

$$s_t(j) = \frac{H_t(k^*, j)}{G_t(k^*, j)}\left[\frac{G_t(k^*, j-1)}{H_t(k^*, j-1)} + \frac{G_t(k^*, j)}{H_t(k^*, j)}\right]/2 \tag{16}$$

The samples are divided into two parts according to the threshold value $s_t(j)$, which in turn calculates the predicted value of each part. We then calculate the predictive contribution of each sample on the new decision tree:

$$L_t(j) = \{i | X(i, j) \le s_t(j)\}. \tag{17}$$

$$R_t(j) = \{i | X(i, j) > s_t(j)\} \tag{18}$$

Compute new predictive values: using the split samples and the corresponding negative gradient values, we can construct a new regression tree model $f_t j$ and compute the new predicted values $F_t(i, j)$:

$$F_t(i, j) = F_{t-1}(i, j) + \eta \cdot f_t j(i, j) \tag{19}$$

where η is the learning rate, is used to control the step size of each iteration of the model update. After T rounds of iterations, we will get T decision trees, and the final model predictions are obtained by weighted average of the predictions of these decision trees. i.e:

$$F_T(i, j) = \sum_{t=1}^{T} \eta \cdot f_t j(i, j) \tag{20}$$

Where $f_t j(i, j)$ denotes the predicted value of sample i in variable j in the tth decision tree, and $F_T(i, j)$ is the final prediction result. In this way, through iterative optimization, the model can gradually reduce the prediction error and improve the prediction accuracy during multiple rounds of learning.

3.6. Speed limit information recognition framework

To address speed limit information recognition, this paper develops an integrated framework (referenced in Fig. 1) that employs ridge regression to compute the weight matrix between features and labels, and models the two objectives of relevance and redundancy. It leverages Feature Pareto Frontier Selection (FPFS) for efficient Pareto-based feature selection to regulate and steer the Gradient Boosting Decision Tree Multi-Objective (GBDT-MO) learning process towards faster convergence. The number of top-ranked FPFS features is dynamically optimized as a hyperparameter within GBDT-MO via Bayesian methods, focusing training on the most relevant features at each boosting iteration. This approach provides adaptive, relevance-guided feature space regularization for GBDT-MO, enhancing the model's efficiency and accuracy in predicting speed limits.

4. Case study

4.1. Data presentation

The experimental dataset of this study was provided by Fujian Provincial Highway Information Technology Co. The dataset contains all the data for the month of May 2021. Citing the travel speed collection conditions of the Design Specification for Highway Speed Limit Signs: 1. To avoid the interference and fluctuation caused by the surge of travel on weekends, the transaction data from Monday to Friday should be selected; 2. To avoid interference caused by too early or too late, the normal work and rest time from 6:00 a.m. to 6:00 p.m. should be selected; therefore, this study selects the data from the time range of Monday, May 10, 2021, to Friday, May 14, 2021, 6:00 a.m. to 6:00 p.m.

This dataset is a comprehensive amalgamation of three specialized data sources: Electronic Toll Collection (ETC) Gantry Transaction Dataset: As the cornerstone dataset, it comprises approximately 27 million records of vehicular movements through gantries. Its extensive nature is critical in extracting label-specific features that are essential for our model's training. Road Segment Information Dataset: This dataset includes 2,950 records providing detailed information about road segments such as length, service areas, and network interconnections. It is instrumental in offering common features that bolster the predictive capabilities of our model. Speed Limit Information Dataset: Comprising 855 records, this dataset informs the model about speed regulations across various segments. It serves as labeled data for training the model in recognizing and categorizing speed limit adherence. It details speed limits by gantry numbers, distinguishing between 635 segments with vehicle-type-specific limits and 220 segments with uniform limits applicable to all vehicles. For a clear and concise representation of this information, Table 2 illustrates the breakdown of these datasets, while Fig. 2 showcases the spatial distribution of the gantry nodes.

All model training and evaluation were conducted using Python 3.9 on a system equipped with an Intel i9-10900K CPU, Nvidia GeForce RTX3070 GPU, 64 GB RAM, and CentOS 7.9 OS. The division of training and test sets adheres to a standard machine learning protocol, partitioning the data into a 7:3 ratio. This split ensures a robust training process and allows for a reliable evaluation of the model's predictive performance. Notably, both the baseline and proposed models utilize the same inputs from these datasets, ensuring consistency and objectivity in our experimental results.

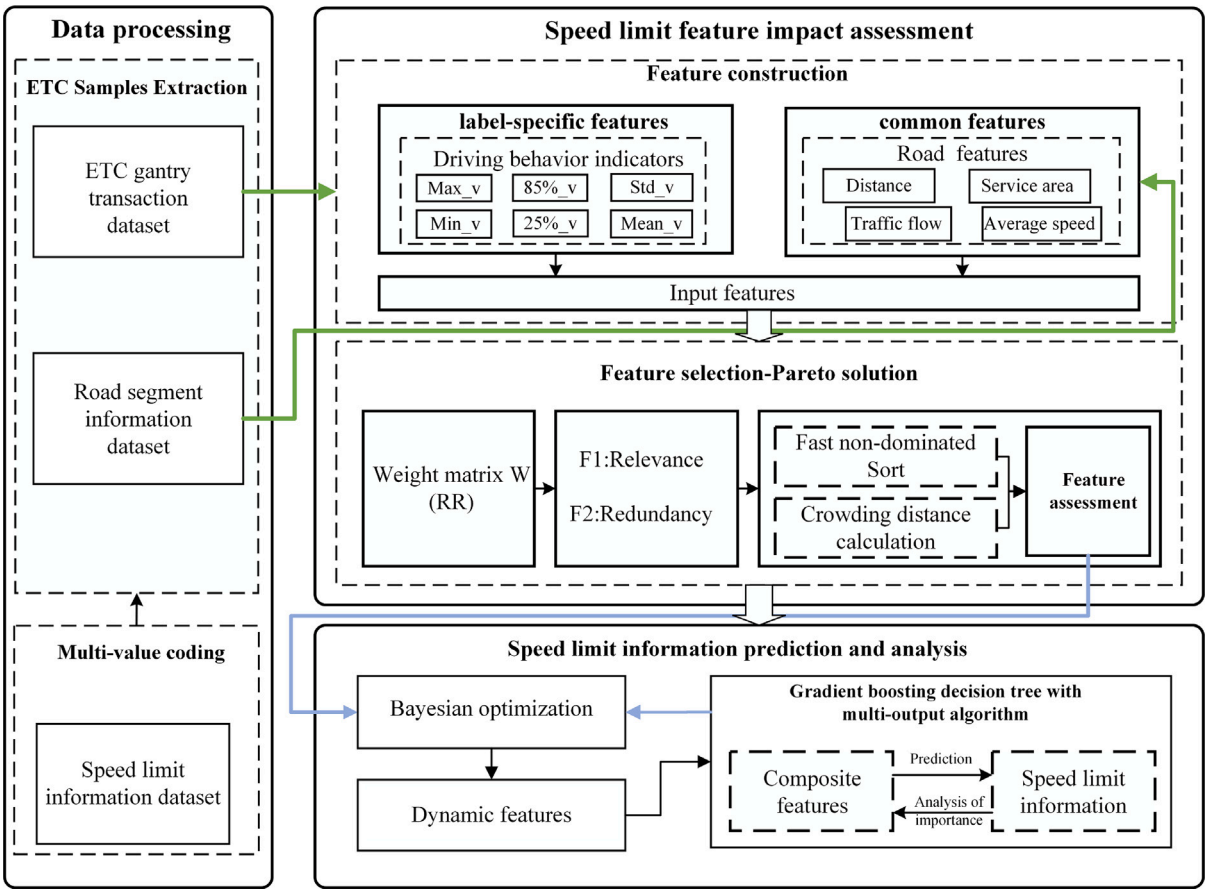


Fig. 1. An integrated framework based on Data-Driven speed limit information recognition Models.

Table 2
Partial field description of the dataset.

Dataset	Fields	Description	Format
ETC gantry transaction dataset	tradetime	Vehicle and gantry interaction time	Datetime
	obuplate	License plate number	String
	entime	Time for vehicles to enter the highway	Datetime
	enstation	Name of toll station	String
	flagid	Number of the gantry	String
	vehclass	Vehicle Class	Decimal
Road segment information dataset	section_id	Consists of numbers entering the gantry and leaving the gantry	String
	distance	length of road segments	Float
	service	Service Area Information	String
Speed limit information dataset	ennodeid	Number of the gantry entered	String
	exnodeid	Number of the leaving gantry	String
	limit_speed	Speed limit information for sections	String

4.2. Data processing

This study aims to develop a data-driven approach for recognizing speed limit information. Given the differing speed limit standards across various road segments, it is crucial to create a segment-based speed limit information feature library. We construct this feature information base by segmenting and clustering driving behavior data, detailed in Section 3.3. The present subsection outlines how we handle data noise during this feature construction process. Duplicate data generated during data fusion is dealt with by retaining a single data entry and discarding the rest. The ensuing discussion elaborates on the treatment of anomalous and missing data.

Abnormal data handling

In the highway environment, vehicle speed has actual physical significance and has a remarkable impact on driving safety and traffic efficiency. However, anomalous data can be generated due to measurement errors, recording errors, or other unforeseen reasons, which can further lead to distorted analysis results and negatively impact the decision-making process.

In order to effectively identify anomalous data, a series of effective strategies are employed. First, considering the constraints of vehicles traveling in one direction, it is impossible for the driving speed to be negative. Thus, when analyzing the user's driving behavior, we constructed the first screening mechanism based on this to eliminate the retrograde data that does not match the actual situation. Second, in further optimizing the data screening process, we refer to statistical principles, especially the Central Limit Theorem (CLT). According to the theorem, in the case of a large number of independent random variables, the sample mean will converge to a normal distribution. Therefore, we use the sample mean plus or minus 2.62 times the standard deviation as the upper and lower boundaries of the data (Liao

Table 3
Some examples of missing common features (Unit: km/h).

Road segments	Hour6	Hour7	Hour8	Hour9	.....	Hour15	Hour16	Hour17
34***7-34***9	NaN	NaN	89.979479	93.874206	.....	90.904301	92.504815	95.160878
34***B-34***D	43.582353	NaN	42.73125	NaN	.....	41.332252	42.994231	42.161013
35***7-35***5	84.499835	82.847865	83.229129	NaN	.....	NaN	86.785834	86.545572
35***D-35***B	96.418211	100.685371	98.969502	96.231214	.....	NaN	NaN	NaN

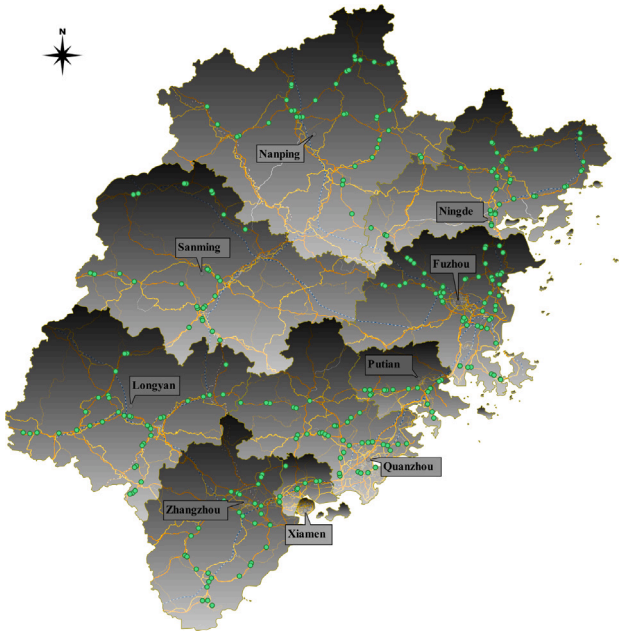


Fig. 2. Schematic diagram of distribution nodes of highway gantries in Fujian Province.

et al., 2015) to determine the range of anomalous data. We set the following boundary values.

$$\text{lowerbound} = \text{mu} - 2.62 * \text{sigma} \tag{21}$$

$$\text{upperbound} = \text{mu} + 2.62 * \text{sigma} \tag{22}$$

Hence, data beyond the range of boundary values will be regarded as anomalous data, which requires in-depth investigation and processing to avoid misleading the analysis results. We utilized the Cumulative Distribution Function (CDF) to estimate data coverage within specific intervals, determining that approximately 99.12% of data samples fall within our calculated range. This indicates a high level of confidence in our method for identifying and treating anomalous data. By adopting these two screening mechanisms, we are able to identify and process anomalous data more effectively, thus improving the accuracy and reliability of model training and prediction. This anomalous data processing method is highly flexible and scalable and can be widely applied to a variety of transportation scenarios.

Missing value handling

To extract the common features of the road segments, we counted the average speed of all vehicles in the road segments in a one-hour sliding window, which could be a potential speed limit feature information. However, due to roadway attributes, not all areas will have the same high traffic volumes as the urban neighborhood highways. Our statistics on the experimental data from May 10 to May 14, 2021, indicate that the minimum traffic flow on all roadways is about 300 vehicles, the maximum traffic flow is about 72,000 vehicles, and the average traffic flow is about 9,500 vehicles. It is clear that the distribution of traffic flow is not balanced. This may lead to a remote area in some time periods and no vehicles through, which also means that the road segment in the corresponding time period does not have vehicle travel speed information, which will lead to some of the common features of the section extraction failure, see Table 3.

For the treatment of containing missing values, we mainly divided into two cases: for the road segments with a large amount of missing data, we did not take the corresponding complementary measures, but deleted them. For road segments with a small amount of missing data, we supplemented them using the average values of different road segments in the same time domain.

Imbalance data processing

Dataset imbalance is a common challenge in categorization tasks, often mirroring real-world scenarios rather than being a characteristic of the data itself. This phenomenon is also observed in our study, which evaluates road speed limits influenced by multiple factors. Consequently, not all roads have uniform speed limit information. In our analysis, sub-model speed limits account for approximately 72.3% of the data, whereas single speed limits comprise about 27.7%. More specifically, for sub-model speed limits, all recorded speeds are 110/100. For single speed limits, the distribution is as follows: 1.8% at 80 km/h, 20.5% at 100 km/h, 3.4% at 110 km/h, and 2% at 120 km/h.

To handle the problem of unbalanced experimental datasets, inspired by a recent review on unbalanced multi-label classification methods (Tarekegn et al., 2021).MLC unbalanced methods can be categorized into four groups: resampling methods, classifier adaptive methods, integration methods, and cost-sensitive methods. Resampling methods have the advantage of being applicable to any MLC classifier but at the same time may introduce noise; classifier adaptation requires extensive knowledge of the particular classifier and problem domain; integration methods reduce overfitting but have excessive computational complexity; and cost-sensitive methods are computationally efficient but often the actual cost value is unknown. After considering the computational cost and complexity, we choose the resampling method. The synthetic minority class oversampling heuristic generation technique is a resampling method that aims to balance the dataset by considering the similarity between minority class samples and then generating new synthetic samples. It increases the diversity of the minority class samples and captures the distribution of features of the minority class samples.

4.3. Competitive algorithm

In this study, we selected seven multi-label classification models including ML-KNN, ML-RF, RFDTBR, RFPCT, TabNet, 1D CNN, and GBDT-Sparse. These models were chosen to compare the methods because of their representativeness and excellent performance in the domain of multi-label classification. Note that we did not optimize all the parameters of the models, but selected the more important ones, and the rest of the parameter settings follow the recommended parameters.

ML-KNN (Zhang and Zhou, 2007): one of the most classical learning methods applied to the domain of multi-label classification, adapted from the traditional KNN machine learning method. By using k nearest neighbor instances for classification, it is able to effectively handle the case of multiple related labels. [Parameter configuration: k = 10].

ML-RF (Tang et al., 2023): An adaptation of the Random Forest algorithm for multi-label classification tasks that applies the idea of algorithmic adaptivity to train a Random Forest model that considers

all labels simultaneously. [Parameter configuration: max_depth = None, min_sample_split = 2, n_estimators = 200]

Binary Relevance with Random Forest of Decision Trees (RFDTBR) (Bogatinovski et al., 2022): uses the problem transformation idea in conjunction with a random forest of decision trees to train a separate random forest for each label in a multi-label classification problem. It is the best-performing algorithm in the most recent review of multi-label classification methods. [Parameter configuration: max_depth = 20, min_sample_split = 2, n_estimators = 200]

Random Forest of Predictive Clustering Trees (RF-PCT) (Bogatinovski et al., 2022): a multi-label classification method that employs the integration of algorithmic adaptive methods, using the PCT as a basic learner that samples the instance space and the feature space. It is the next best-performing algorithm in the latest reviews on multi-label classification methods. [Parameter configuration: n_estimators = 100].

TabNet (Arik and Pfister, 2021): it is a deep learning model introduced by researchers at Google Cloud AI. It is designed for structured data. We made a small adaptation to make it suitable for multi-label categorization tasks. [Parameter configuration: default hyperparameters recommended by the original authors]

1D Convolutional Neural Networks (1D CNN): from a multi-label classification competition organized by the kaggle community, this scheme achieved the optimal performance by reshaping structured data into a multi-channel image format. [Parameter configuration: n_epochs = 200, batch_size = 64, learning_rate = 0.001]

GBDT-Sparse (Si et al., 2017): a gradient-enhanced decision tree variant where the output space is high-dimensional sparse. Considering the experimental data and the nature of the task in this study, we also use this algorithm as one of the comparison methods. In addition, the method proposed in this study belongs to the same family as this algorithm. [Parameter configuration: max_depth = 17, learn_rate = 0.1160].

In the next subsection, we will describe in detail the parameter optimization method of the model and the reasons for choosing the optimization algorithm.

4.4. Hyperparameter optimization

Hyperparameter optimization is an important aspect of training machine learning models as it involves searching for the best set of hyperparameters that maximizes the model performance. Commonly used optimization methods include but are not limited to, the following techniques, grid search, random search, Bayesian optimization, evolutionary algorithms, meta-heuristic algorithms, and so on. Different optimization methods are usually applied in different application scenarios. For example, grid search is a relatively simple and straightforward method that defines the search space of hyperparameters and exhaustively evaluates all possible combinations of hyperparameters. Although grid search methods are able to find the optimal set of hyperparameters, they can be computationally expensive when dealing with a large number of hyperparameters. Randomized search is employed randomly from a defined search space and it does not exhaustively search for every combination of hyperparameters but focuses on randomly selected points. Bayesian optimization is a sequential model-based optimization technique that uses an agent model to approximate the objective function and iteratively explores the optimization space by evaluating the objective function at selected points, which is suitable for scenarios where the evaluation of the objective function is more time-consuming. We will not mention other optimization techniques in detail and can refer to the cited literature.

In this study, we applied the Bayesian optimization technique (Shahriari et al., 2015) instead of the grid search technique or other optimization methods. In the subsection of speed limit information feature evaluation, in order to improve the performance of the training model and to speed up the convergence of the training model. We applied a feature selection algorithm (FPFS) based on the Pareto dominance to get the importance ranking of the feature information. However, the number of selected features is required to be actively chosen by us, and different combinations of features with different hyper-parameter configurations will have different performances. If the grid search technique is used, the time complexity of the model is too great, the computational cost is large, and the implementation of other optimization techniques is relatively complex. On the contrary, Bayesian optimization has good performance in terms of adaptability and processing efficiency in dealing with different feature sets.

4.5. Evaluation metrics

Understanding the performance of predictive models is a key aspect of research experiments, and this can only be achieved through the use of a set of scientifically based assessment metrics. These metrics provide quantifiable measures of model effectiveness and sophistication and help to compare and contrast the relative capabilities and shortcomings of the designed models with state-of-the-art alternatives. In this study, we draw on the most recent review article (Tarekegn et al., 2021) to establish an evaluation system for models from three different perspectives in order to properly assess the comprehensive performance of the models.

Example-based evaluation metrics: Such metrics analyze individual instances and their associated predicted results, providing a basic view of model performance. The family includes:

Hamming loss quantifies the discrepancy between a model's predicted labels and the actual labels by calculating the average number of labels incorrectly predicted for each sample. This metric employs Δ, representing the symmetric difference between two label sets, to measure this divergence.

$$HammingLoss = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{d}|f(x_i)\Delta y_i| \tag{23}$$

Accuracy: This quantifies the proportion of true-positive and true-negative predictions out of all predictions. Accuracy scores close to 1 indicate a higher number of correct predictions. Where δ is an indicator function returning 1 for correct predictions.

$$Accuracy = \frac{1}{d}\sum_{i=1}^{n}\sum_{j=1}^{d}\delta(y_{ij} = f(x_{ij})) \tag{24}$$

Jaccard Index: also known as concurrent intersection, this quantifies the size of the overlap between predicted and actual labels divided by the concurrent set size. A higher Jaccard score is better because it indicates a higher degree of similarity between predicted and actual labels.

$$JaccardIndex = \frac{1}{n}\sum_{i=1}^{n}\frac{|f(x_i)\cap \mathbf{y}_i|}{|f(x_i)\cup \mathbf{y}_i|} \tag{25}$$

Label-based evaluation metrics (Li et al., 2022): these metrics scrutinize the model's performance from the perspective of each label in a multi-label setup. They help to understand how the model performs in each individual category. They include:

F1 Macro: This metric calculates the F1 score (the reconciled average of precision and recall) for each label independently and then averages it. This treats all categories equally, regardless of their frequency.

$$MacroF_1 = \frac{1}{d}\sum_{i=1}^{d}\frac{2p_i^d r_i^d}{p_i^d + r_i^d} \tag{26}$$

F1 Micro: The global F1 score is calculated by counting the total number of true positives, false negatives, and false positives. It is particularly useful when class imbalances need to be taken into account, as it weights each class according to its presence in the dataset.

$$MicroF_1 = \frac{2\sum_{j=1}^{d}\sum_{i=1}^{n} y_{ij} f\left(x_{ij}\right)}{\sum_{j=1}^{d}\sum_{i=1}^{n} y_{ij} + \sum_{j=1}^{d}\sum_{i=1}^{n} f\left(x_{ij}\right)} \tag{27}$$

Performance-based assessment metrics: These metrics focus on the efficiency of the model in terms of computational resources, thus enabling an understanding of the actual feasibility of the model. They include:

Training time: this is the time taken by the model to learn from the training data. A model with a shorter training time is better because it is more efficient.

Testing time: this measures the time it takes for the model to predict the results of the test data. A shorter test time is favorable because it indicates a faster model.

This holistic evaluation system ensures a robust assessment of the model's functionality by providing a balanced view of instance-level, label-level, and performance-level attributes. It is this holistic assessment approach that provides a more accurate understanding of the overall strengths and weaknesses of the model in question, thus providing an effective roadmap for further refinement and optimization.

5. Results and discussion

5.1. Performance evaluation based on data-driven methods

In order to validate our proposed Pareto-GBDTMO approach, we evaluated the performance with seven state-of-the-art data-driven multi-label classifiers on a highway dataset. The selected methods include both classical methods such as ML-KNN and ML-RF, as well as state-of-the-art techniques such as RFDTBR, RFPCT, TabNet, 1D CNN and GBDT-Sparse. This broad comparison provides an encompassing benchmark. The performance of each classifier was measured within our comprehensive evaluation system, as shown in Table 4, where the best results under different metrics are highlighted in bold.

As shown in Table 4, our Pareto-GBDTMO model outperforms all other techniques in terms of instance-based Hamming loss, accuracy, and Jaccard similarity metrics. Specifically, we have the lowest Hamming loss, the only one below 0.1%, which indicates that our approach significantly reduces mislabeled predictions. Our accuracy of 97.19% and Jaccard score of 0.9859 are significantly higher than our competitors, demonstrating our excellent overall performance. Analyzing the label-based metrics, our method again performs best in F1 macro and F1 micro with 0.9907 and 0.9883, respectively, demonstrating that Pareto-GBDTMO is able to balance the performance across all labels without favoring the higher frequency categories.

Among competing algorithms, ML-KNN, a classical multi-label classification method, exhibits the lowest metrics. RFDTBR and RFPCT, as state-of-the-art algorithms, present strong performance, and GBDT-Sparse also yields positive experimental results. The deep learning-based methods TabNet and 1D CNN show divergent performances, with TabNet being superior. Several factors could account for this, such as TabNet being designed for tabular data, while 1D CNN is more suitable for unstructured data. Nonetheless, it demonstrates that deep learning methods can be applicable. RFDTBR, RFPCT, GBDT-Sparse, and our model are all decision tree-based algorithms and show high performance, indicating that decision trees effectively capture the complex feature dependencies and intersections between multiple output labels in our problem. Though ML-RF is also a decision tree-based model, it can sometimes overfit, and high feature correlation might degrade its performance. Lastly, ML-KNN, an inert learning mechanism, might not be well suited to handle tasks with high dimensionality and complex class boundaries.

Performance evaluation based on time

Real-time processing in in-vehicle environments is a significant challenge for image-based recognition solutions, which we incorporated into our program's metrics. We separately tested the training time and testing time of all methods, as shown in Fig. 3. Regarding the time-performance balance, our method stands out as optimal. While the training time of our method is not the shortest, it consumes relatively little time. Moreover, our method exhibits the best testing time

Table 4
Instance-based and label-based model performance.

Algorithm	F1 macro	F1 micro	Hamming loss	Accuracy	Jaccard
ML-KNN	0.9194	0.9120	0.0508	0.8781	0.9156
ML-RF	0.9696	0.9601	0.0234	0.9062	0.9516
RFDTBR	0.9856	0.9817	0.0109	0.9563	0.9781
RFPCT	0.9854	0.9818	0.0109	0.9563	0.9781
TabNet	0.9790	0.9727	0.0163	0.9363	0.9675
1D-CNN	0.9633	0.9542	0.0273	0.9000	0.9448
GBDT-Sparse	0.9816	0.9777	0.0133	0.9531	0.9750
OURs	**0.9907**	**0.9883**	**0.0070**	**0.9719**	**0.9859**

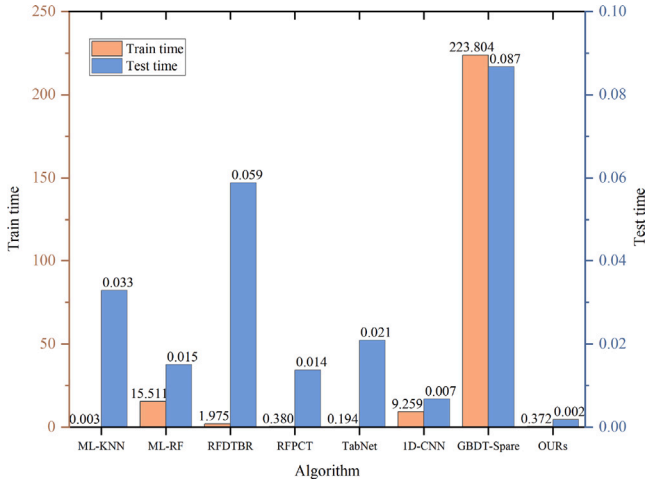


Fig. 3. Time performance based on data-driven methods.

performance among all the approaches. Hence, taking both training and testing times into account, our method demonstrates optimal time performance.

On the contrary, the homology model GBDT-Sparse has the most time-consuming training time and testing time. This is due to the fact that the method itself is designed for extreme multi-label classification problems and the method is not heavily optimized for efficiency. As the state-of-the-art multi-label classification models: RFDTBR and RFPCT, also have excellent performance in terms of time performance. The reason that RFDTBR has a longer test time may be that the models constructed by applying BR require more trees, which tends to require more computation time. The most classical ML-KNN model has the shortest training time, but not the testing time. The reason for this is that ML-KNN is an inert learning algorithm, which means that it does not really train the model during the training phase, it just stores and indexes the training data points, which leads to very fast training. But the testing process requires actual consumption of neighbor search, distance calculation, etc.

Analysis of the causes of incorrect forecasts

Despite our model outperforming contemporary state-of-the-art counterparts, it became necessary to delve into the underlying causes of any mispredictions. Fig. 4 provides a detailed analysis of these mispredictions. Notably, the classification outcomes for both Label 1 and Label 4 were exceptional, underscoring our model's adeptness in accurately detecting these distinct speed limit indicators.

However, challenges emerged in the classifications for Label 2 and Label 3. For Label 2, which represents a speed limit of 100 km/h, there were two instances of misclassification. Similarly, for Label 3 (indicating a 110 km/h limit), four instances were incorrectly categorized as alternative labels. These discrepancies can be attributed to the multifaceted nature of speed limit data for Labels 2 and 3. Unlike single speed limit indicators, these labels encompass layered vehicular speed

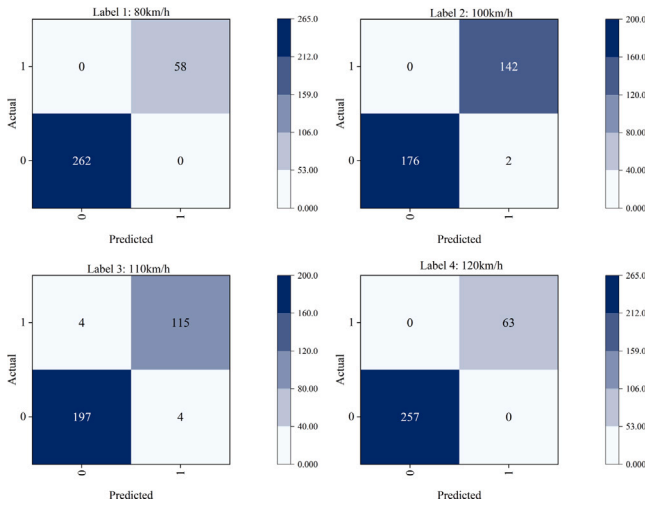


Fig. 4. Analysis of the causes of incorrect forecasts.

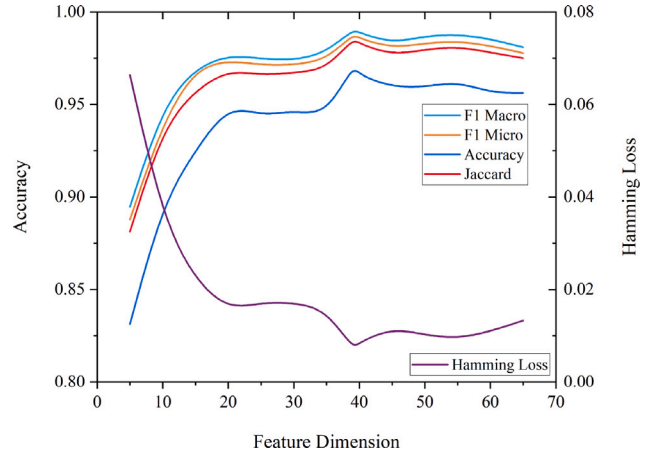


Fig. 5. Relationship between feature dimensions and model performance.

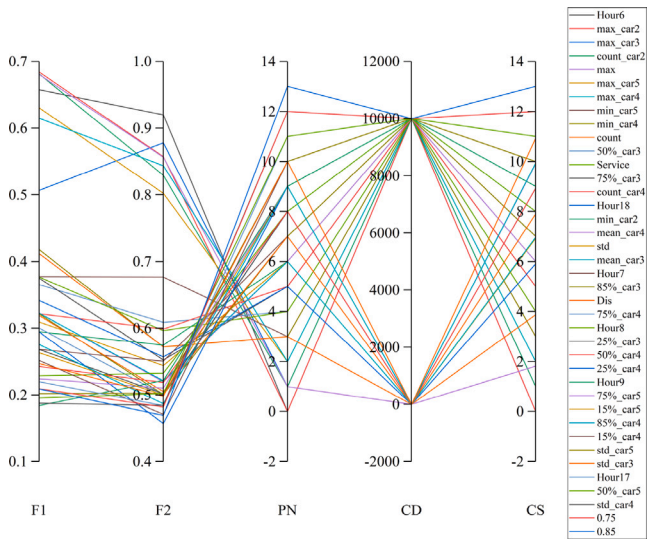


Fig. 6. Characterization analysis of the effective features.

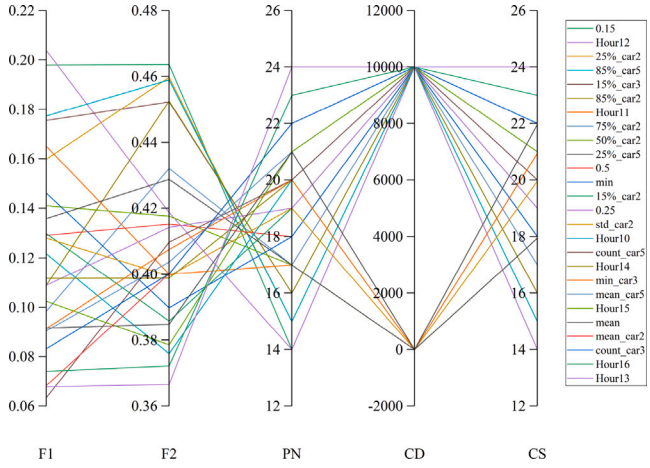


Fig. 7. Characterization analysis of the interference features.

limit details. Our model's occasional struggles in comprehending such intricate feature interactions—particularly when these features are not markedly distinct—might be a contributing factor. Another potential influence could be the paucity of data in specific segments, which might impede the model's ability for robust feature extraction, thereby affecting its accuracy.

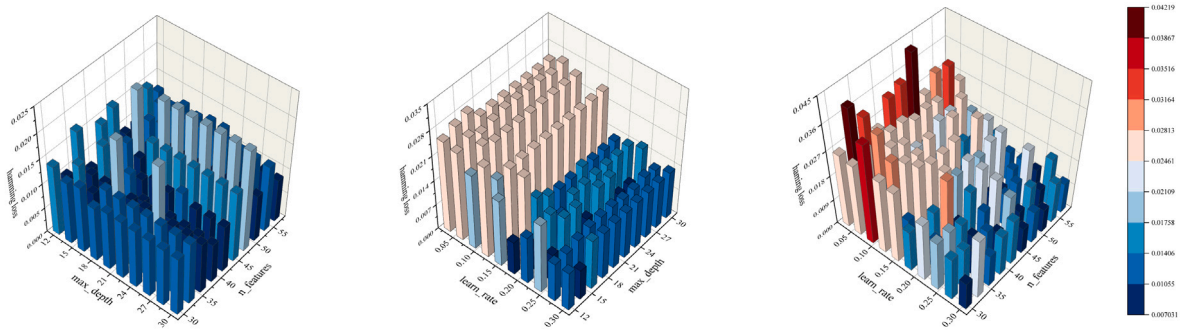
5.2. Analysis of the importance of speed limit information features

To quantify the impact of feature selection, we evaluate the model performance by progressively introducing more features based on the FPFS ranking. The experiments on feature dimension-hamming loss further corroborate our assumptions, see Fig. 5, and provide a constructive basis for distinguishing effective features from interfering ones. It is clear that as the feature dimension increases, all the metrics except the hamming-loss metric increase dramatically. When the feature dimension reaches 39, the model achieves the best performance. However, with another increase of feature dimension, the performance of the model did not improve again as we expected, but instead started to decline.

In order to gain insight into this process, we analyze the characteristics of the top-ranked "effective features" versus the discarded "interfering features". The visualization results are shown in Figs. 6 and 7, and a detailed rationale explanation of the legend can be found in Section 3.4. In this section we only briefly describe that F1 denotes the degree of correlation of the features; F2 measures the redundancy of the features; PN is the number of Pareto fronts assigned to each feature (the lower the value the higher the degree of nondominance); CD denotes the crowding distance, a measure of the solution space (the higher the value, the better); and CS (combined score): is the overall measure of feature quality (lower values are better).

The visualizations in Fig. 6 show that effective features have two trends. One is to have both high relevance and high redundancy, and these features tend to have high non-dominance and high crowding distance. The reason may be that multi-labels with correlation need to share more meta-features. The other is having low relevance and redundancy, which are features with relatively low non-dominance. Some have high crowding distances and some have crowding distances that are low. This may be due to the fact that these low redundancy features all have potentially gainful information that may be relatively small but they are unique. On the contrary, ideal features with both high relevance and low redundancy do not seem to be very common.

In our research, the following features are identified as the most pertinent: the average speed of all vehicles at 6:00 a.m., the maximum speed values of small cars and other vehicle types, the traffic volume of small cars, and the max speed for all vehicles on the roadway. These features are ranked at the top in our composite feature scores, with their specific rankings detailed in Fig. 6.



(a) max_depth ∈ {12,14,...,30} and n_features ∈ {30,33,...,57} (λ = 10, learn_rate = 0.25)

(b) max_depth ∈ {12,14,...,30} and learn_rate ∈ {0.03,0.06,...,0.3} (λ = 10, n_features = 39)

(c) learn_rate ∈ {0.03,0.06,...,0.3} and n_features ∈ {30,33,...,57} (λ = 10, max_depth = 13)

Fig. 8. Model performance with different parameter configurations (evaluation metric: hamming loss).

Table 5
Model performance testing under various conditions.

Type	F1 macro	F1 micro	Hamming loss	Accuracy	Jaccard
Type1 (imbalance)	0.4804	0.9568	0.0393	0.8652	0.9213
Type2 (imbalance-features)	0.4840	0.9632	0.0337	0.8876	0.9326
Type3 (balance)	0.9809	0.9778	0.0133	0.9563	0.9750
Type4 (balance-features)	0.9907	0.9883	0.0070	0.9719	0.9859

In contrast, the interference features shown in Fig. 7 are cluttered, but they all share the common characteristic of lower correlation. Although they also have low redundancy, they are all less non-dominated. This indicates that these interference features are able to be dominated by other features independent of their crowding distance. In other words, the gain information provided by these features can be replaced by other features.

5.3. Verification and testing of various functions

The effectiveness of modeling is closely related to the data set and the nature of the task. In this study, in order to balance the minority class samples, we apply the technique of minority class sample resampling for data imbalance to achieve the balance between data categories and thus improve the model performance. Secondly, the pareto dominance based feature selection technique aims to reduce the feature dimensionality to enhance the training performance.

Therefore, in order to distinguish the effectiveness of the application of each component, four sets of experiments were conducted. They are type1: performance under data imbalance condition; type2: performance under data imbalance condition applying feature selection technique; type3: performance under data balance condition; and type4: performance under data balance condition applying feature selection technique, and the results of the experiments are shown in Table 5.

The experimental results strongly indicate that the nature of the data distribution (balance or imbalance) greatly affects the performance of the multi-label classification model. The transition from an unbalanced to a balanced dataset resulted in a significant enhancement of all the performance metrics evaluated, as evidenced by an increase in the F1 Macro, F1 Micro, Accuracy, and Jaccard scores, as well as a desirable reduction in the Hamming loss. This transition held true for models with and without feature selection (i.e., from type 1 to type 3 and from type 2 to type 4), thus illustrating the critical importance of data balancing for model validity.

In contrast, the introduction of feature selection techniques resulted in a modest enhancement of model performance, as noted when comparing models without feature selection to models with feature selection (i.e., from type 1 to type 2 and from type 3 to type 4). Interestingly, the positive effect of feature selection was slightly more prominent when data were balanced, suggesting a potential interaction between data balancing and feature selection. However, as mentioned earlier, the application of the feature selection technique aims to speed up the convergence of the trained model. For this purpose, we repeat the training model five times with the optimal parameter configuration of the model and take the average value, and compare the performance of the enhancement after applying the feature selection technique in data balance and data imbalance, respectively, as shown in Table 6. The experimental results show that the application of the feature selection technique can accelerate the convergence speed of the training model by 44.28% under the condition of data imbalance, and under the condition of data balance, the application of the feature selection technique can accelerate the convergence speed of the training model by 31.79%.

5.4. Parameter sensitivity analysis

In this study, we develop an integrated framework that utilizes FPFS feature selection to regulate and guide the GBDT-MO learning process to accelerate convergence. And the number of frontier FPFS features is dynamically optimized as hyperparameters of GBDT-MO to focus on the most relevant features in each boosting iteration. Therefore, our hyperparameters mainly come from the feature selection technique FPFS and the training model GBDT-MO, which are λ, n_features, learning_rate, and max_depth, respectively. where learning_rate and max_depth control the learning rate of the training model and the maximal tree depth, respectively, and n _features represents the number of dynamic features input to the training model. λ is a hyperparameter of FPFS used to control the amount of regularization of the coefficients in ridge regression, which determines the balance between fitting the training data well and keeping the coefficients small to avoid overfitting. We followed the original authors' suggestion that $\lambda = 10$ is the optimal value in a bi-objective optimization problem.

To validate and analyze the optimal hyperparameter settings for our proposed method Pareto-GBDTMO, we compare the performance of different learning_rate, max_depth and n_features. The specific parameters are configured as follows: max_depth $\in \{12, 14, \ldots, 30\}$ with a step size of 2; learn_rate $\in \{0.03, 0.06, \ldots, 0.3\}$ with a step size of 0.03; n_features $\in \{30, 33, \ldots, 57\}$ with a step size of 3. We fixed one hyperparameter to be optimal while dynamically adjusting the remaining two hyperparameters, see Fig. 8.

Fig. 8(a) shows that the model's performance is generally superior when the learning rate is optimal, indicating that the learning rate is the primary parameter influencing the model's performance. Moreover, when n_features remain consistent, the model's performance is nearly unchanged, suggesting that max_depth has a lesser influence. Conversely, with a constant max_depth, altering n_features leads to

Table 6
Model training time for balanced vs. unbalanced datasets (Unit: s).

Epoch	Imbalance	Imbalance-features	Enhancement effect	Balanced	Balanced-features	Enhancement effect
1	0.384828	0.222792	0.421060	0.561464	0.381746	0.3201
2	0.385863	0.214888	0.443098	0.567615	0.381655	0.3276
3	0.384982	0.214534	0.442744	0.541453	0.383606	0.2915
4	0.392472	0.214263	0.454067	0.557042	0.385382	0.3082
5	0.389095	0.212831	0.453009	0.571994	0.376420	0.3419

significant performance fluctuation. Fig. 8(b) reveals that the model's performance is sub-optimal when using optimal n_features. With a consistent learning rate, performance remains unusually stable, further indicating that the maximum depth is not a sensitive parameter. However, with a constant max_depth parameter, the model consistently performs well as the learning rate progressively increases. Lastly, Fig. 8(c) emphasizes that the model's performance is poorest when max_depth is optimal. For the same n_features, performance is almost positively correlated with the learning rate. The worst performance always seems to occur for smaller learning_rate values.

6. Conclusion

In this paper, we propose a novel data-driven method for speed limit recognition using multi-label classification. Unlike previous image-based solutions, our approach treats speed limit recognition as a multi-output learning problem to account for different speed limits on different road segments. We construct label-specific features from user driving behavior and supplement road attributes as common features.

To improve the convergence of the model, we develop a fast Pareto feature selection technique that identifies the most relevant and non-redundant features to guide the GBDT-MO training process. Bayesian optimization further adjusts the subset of features used in each boosting iteration to achieve optimal regularization. Experiments demonstrate that our Pareto-GBDTMO method achieves 97% accuracy, 0.7% loss rate, and optimal time performance on highway data, which greatly outperforms existing multi-label classifiers.

This research has three main contributions. First, we demonstrate the feasibility of using vehicle-level data for reliable and efficient speed limit recognition in a data-driven framework. Second, we introduce multi-label classification in a traffic study and demonstrate its effectiveness in modeling different speed limits. Finally, our proposed feature selection and model regularization techniques advance multi-label learning for unbalanced, high-dimensional traffic data.

Limitations of this work include recognizing lane-level speed limits and developing a more lightweight architecture. As future work, we aim to collect and evaluate higher granularity data across lanes and optimize our models for computational efficiency. We will also explore mechanisms to integrate these data-driven recognizers with the existing ADAS pipeline to enable real-time speed limit alerts. By increasing redundancy, our approach has the potential to overcome the inherent challenges of image-based recognition and improve the safety of intelligent driving systems.

CRediT authorship contribution statement

Xu Luo: Conceptualization, Methodology, Software, Writing – original draft. **Fumin Zou:** Conceptualization, Methodology, Validation, Writing – review & editing. **Qiang Ren:** Formal analysis, Validation, Writing – review & editing. **Sijie Luo:** Methodology, Software, Validation, Writing – review & editing. **Feng Guo:** Methodology, Data curation, Writing – review & editing. **Huan Zhong:** Methodology, Writing – review & editing. **Na Jiang:** Validation, Writing – review & editing. **Xinjian Cai:** Data curation, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is partially supported by the Renewable Energy Technology Research institution of Fujan University of Technology Ningde, China (Funding number: KY310338), the 2020 Fujian Province "Belt and Road" Technology Innovation Platform (Funding number: 2020D002), the Provincial Candidates for the Hundred, Thousand and Ten Thousand Talent of Fujian (Funding number: GY-Z19113), the Patent Grant project (Funding number: GY-Z18081, GY-Z19099, GY-Z20074), Horizontal projects (Funding number: GY-H-20077), Municipal level science and technology projects (Funding number: GY-Z-22006, GY-Z-220230), Fujian Provincial Department of Science and Technology Foreign Cooperation Project (Funding number: 2023I0024, the Open Fund project (Funding number: KF-X19002, KF-19-22001).

References

Afghari, A.P., Vos, J., Farah, H., Papadimitriou, E., 2023. "I did not see that coming": A latent variable structural equation model for understanding the effect of road predictability on crashes along horizontal curves. Accid. Anal. Prev. 187, 107075.

Arik, S.Ö., Pfister, T., 2021. Tabnet: Attentive interpretable tabular learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35, no. 8. pp. 6679–6687.

Bayoudh, K., Hamdaoui, F., Mtibaa, A., 2021. Transfer learning based hybrid 2D-3D CNN for traffic sign recognition and semantic road detection applied in advanced driver assistance systems. Appl. Intell. 51, 124–142.

Bi, Z., Yu, L., Gao, H., Zhou, P., Yao, H., 2021. Improved VGG model-based efficient traffic sign recognition for safe driving in 5G scenarios. Int. J. Mach. Learn. Cybern. 12, 3069–3080.

Bogatinovski, J., Todorovski, L., Džeroski, S., Kocev, D., 2022. Comprehensive comparative study of multi-label classification methods. Expert Syst. Appl. 203, 117215.

Chen, J., Jia, K., Chen, W., Lv, Z., Zhang, R., 2022. A real-time and high-precision method for small traffic-signs recognition. Neural Comput. Appl. 1–13.

Chiabaut, N., Faitout, R., 2021. Traffic congestion and travel time prediction based on historical congestion maps and identification of consensual days. Transp. Res. C 124, 102920.

Dai, S.-S., Wu, Y.-D., Xiong, K., Xiao, J.-W., 2022. A real-time speed limit sign recognition algorithm based on network. Telecommun. Eng. 62 (10), 1427–1432.

Deng, W., Zhang, X., Zhou, Y., Liu, Y., Zhou, X., Chen, H., Zhao, H., 2022. An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems. Inform. Sci. 585, 441–453.

Dewi, C., Chen, R.-C., Liu, Y.-T., Jiang, X., Hartomo, K.D., 2021a. Yolo V4 for advanced traffic sign recognition with synthetic training data generated by various GAN. IEEE Access 9, 97228–97242.

Dewi, C., Chen, R.-C., Liu, Y.-T., Tai, S.-K., 2022. Synthetic data generation using DCGAN for improved traffic sign recognition. Neural Comput. Appl. 34 (24), 21465–21480.

Dewi, C., Chen, R.-C., Yu, H., Jiang, X., 2021b. Robust detection method for improving small traffic sign recognition based on spatial pyramid pooling. J. Ambient Intell. Humaniz. Comput. 1–18.

Gao, B.-B., Zhou, H.-Y., 2021. Learning to discover multi-class attentional regions for multi-label image recognition. IEEE Trans. Image Process. 30, 5920–5932.

Hashemi, A., Dowlatshahi, M.B., Nezamabadi-pour, H., 2021. An efficient Pareto-based feature selection algorithm for multi-label classification. Inform. Sci. 581, 428–447.

Huang, A., Xu, R., Chen, Y., Guo, M., 2023. Research on multi-label user classification of social media based on ML-KNN algorithm. Technol. Forecast. Soc. Change 188, 122271.

Li, J., Li, P., Hu, X., Yu, K., 2022. Learning common and label-specific features for multi-label classification with correlation information. Pattern Recognit. 121, 108259.

Liao, L.-c., Jiang, X.-h., Lin, M.-z., Zou, F.-m., 2015. Recognition method of road speed limit information based on data mining of traffic trajectory. J. Traff. Transp. Eng. 15 (05), 118–126.

Liu, J., Chen, Y., Huang, X., Li, J., Min, G., 2023a. GNN-based long and short term preference modeling for next-location prediction. Inform. Sci. 629, 1–14.

Liu, Z., Qi, M., Shen, C., Fang, Y., Zhao, X., 2021. Cascade saccade machine learning network with hierarchical classes for traffic sign detection. Sustainable Cities Soc. 67, 102700.

Liu, S., Song, X., Ma, Z., Ganaa, E.D., Shen, X., 2022. MoRE: Multi-output residual embedding for multi-label classification. Pattern Recognit. 126, 108584.

Liu, Z., Tang, C., Abhadiomhen, S.E., Shen, X.-J., Li, Y., 2023b. Robust label and feature space co-learning for multi-label classification. IEEE Trans. Knowl. Data Eng..

Ma, C., Dai, G., Zhou, J., 2021. Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM_BILSTM method. IEEE Trans. Intell. Transp. Syst. 23 (6), 5615–5624.

Paul, D., Jain, A., Saha, S., Mathew, J., 2021. Multi-objective PSO based online feature selection for multi-label classification. Knowl.-Based Syst. 222, 106966.

Puli, M.S., Sunitha, M., Aluri, O.S.B., Jain, D.R., Rayabharapu, M., Venkatesh, M., 2023. Deep learning-based framework for robust traffic sign detection under challenging weather conditions. J. Surv. Fish. Sci. 2650–2657.

Ran, S., Li, X., Zhao, B., Jiang, Y., Yang, X., Cheng, C., 2023. Label correlation embedding guided network for multi-label ECG arrhythmia diagnosis. Knowl.-Based Syst. 270, 110545.

Rim, H., Abdel-Aty, M., Mahmoud, N., 2023. Multi-vehicle safety functions for freeway weaving segments using lane-level traffic data. Accid. Anal. Prev. 188, 107113.

Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N., 2015. Taking the human out of the loop: A review of Bayesian optimization. Proc. IEEE 104 (1), 148–175.

Si, S., Zhang, H., Keerthi, S.S., Mahajan, D., Dhillon, I.S., Hsieh, C.-J., 2017. Gradient boosted decision trees for high dimensional sparse output. In: International Conference on Machine Learning. PMLR, pp. 3182–3190.

Tang, M., Liang, Z., Ji, D., Yi, J., Peng, Z., Huang, Y., Wang, J., Chen, D., 2023. Inadequate load output diagnosis of ultra-supercritical thermal power units based on MIWOA multi-label random forest. Appl. Therm. Eng. 227, 120386.

Tarekegn, A.N., Giacobini, M., Michalak, K., 2021. A review of methods for imbalanced multi-label classification. Pattern Recognit. 118, 107965.

Wang, J., Chen, Y., Dong, Z., Gao, M., 2023. Improved YOLOv5 network for real-time multi-scale traffic sign detection. Neural Comput. Appl. 35 (10), 7853–7865.

Wang, R., Kwong, S., Wang, X., Jia, Y., 2021a. Active k-labelsets ensemble for multi-label classification. Pattern Recognit. 109, 107583.

Wang, R., Ridley, R., Qu, W., Dai, X., et al., 2021b. A novel reasoning mechanism for multi-label text classification. Inf. Process. Manage. 58 (2), 102441.

Wang, L., Zhou, K., Chu, A., Wang, G., Wang, L., 2021c. An improved light-weight traffic sign recognition algorithm based on YOLOv4-tiny. IEEE Access 9, 124963–124971.

Wu, Y., 2022. Research and Implementation of Speed Limit Sign Recognition Algorithm Based on Deep Learning. Chongqing University of Posts and Telecommunications.

Wu, Y., Tan, H., Qin, L., Ran, B., 2020. Differential variable speed limits control for freeway recurrent bottlenecks via deep actor-critic algorithm. Transp. Res. Part C: Emerg. Technol. 117, 102649.

Xia, Y., Chen, K., Yang, Y., 2021. Multi-label classification with weighted classifier selection and stacked ensemble. Inform. Sci. 557, 421–442.

Xie, K., Zhang, Z., Li, B., Kang, J., Niyato, D., Xie, S., Wu, Y., 2022. Efficient federated learning with spike neural networks for traffic sign recognition. IEEE Trans. Veh. Technol. 71 (9), 9980–9992.

Yu, Z.-B., Zhang, M.-L., 2021. Multi-label classification with label-specific feature generation: A wrapped approach. IEEE Trans. Pattern Anal. Mach. Intell. 44 (9), 5199–5210.

Zhang, Z., Jung, C., 2020. GBDT-MO: Gradient-boosted decision trees for multiple outputs. IEEE Trans. Neural Netw. Learn. Syst. 32 (7), 3156–3167.

Zhang, C., Li, Z., 2021. Multi-label learning with label-specific features via weighting and label entropy guided clustering ensemble. Neurocomputing 419, 59–69.

Zhang, M.-L., Zhou, Z.-H., 2007. ML-KNN: A lazy learning approach to multi-label learning. Patt. Recogn. 40 (7), 2038–2048.

Zhang, J., Zou, X., Kuang, L.-D., Wang, J., Sherratt, R.S., Yu, X., 2022. CCTSDB 2021: A more comprehensive traffic sign detection benchmark. Hum.-Centric Comput. Inform. Sci. 12.

Zhu, Y., Yan, W.Q., 2022. Traffic sign recognition based on deep learning. Multimedia Tools Appl. 81 (13), 17779–17791.