Journal of Computers Vol. 37 No. 2, April 2026, pp. 99-112
https://doi.org/10.63367/199115992026043702007

Prediction and Simulation of Elevator Traffic Flow in High-Rise Hospital Inpatient Buildings

Rui Tian[1], Weihong Luo[1], and Weimin Gao[2*]

[1] School of Mechanical and Electrical Engineering, Guangdong Construction Polytechnic, Guangzhou, 510440, China
`{tianrui, luoweihong}@gdcvi.edu.cn`

[2] School of Medical Information and Engineering, Hunan University of Medicine, Huaihua, 418000, China
`gwm@hnmu.edu.cn`

Received 26 May 2025; Revised 2 June 2025; Accepted 2 June 2025

Abstract. This paper employs a random forest regression model to predict elevator traffic flow in hospital inpatient buildings, providing a scientific basis for optimizing the scheduling and management of hospital elevator systems. Through a comprehensive analysis of the traffic flow characteristics of elevators in these buildings and the factors influencing pedestrian flow, we have developed a prediction model that incorporates various elements, including time distribution characteristics, population composition, and spatial distribution. The simulation results demonstrate that the random forest regression model effectively captures the periodic changes and fluctuations in elevator traffic flow, achieving high prediction accuracy. Predict elevator traffic flow using a simulated dataset (500 samples), when compared to support vector machines (SVM) and back propagation (BP) neural networks, the model exhibits lower mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) values, along with higher R^2 values, confirming its superior performance in predicting elevator traffic flow in tertiary hospital inpatient buildings.

Keywords: elevator traffic flow, prediction and simulation, random forest regression model, high rise hospital inpatient buildings

1 Introduction

Since Mr. Otis invented the safety elevator, elevators have increasingly played a crucial role in vertical transportation within high-rise and super high-rise buildings. The evolution of elevator technology has progressed from single elevators to the parallel operation of two elevators, and subsequently to the collective operation of multiple elevators, leading to the development of theories related to elevator group control and traffic configuration.

From the 1920s to the mid-1970s, elevator control mainly used statistical methods to study the performance patterns of elevator traffic; After the 1970s, with the development and application of computers and integrated circuits, models such as neural networks, fuzzy logic, and expert systems were adopted to study the nonlinearity, fuzziness, and uncertainty of elevator transportation systems. In order to improve the prediction accuracy of elevator passenger flow, Litian Yuan, Jie Zhang and others proposed an elevator passenger flow prediction model combining grey prediction and least squares support vector machine. Firstly, gray model and least squares support vector were used to predict the linear and linear variation laws of passenger flow, and then linear regression was used to determine the weights of the two prediction results. Finally, the elevator passenger flow prediction results were obtained based on the weights [1]; Yihui Zhang proposed an elevator passenger flow prediction model combining RBF neural network and ARMA. Firstly, ARIMA was used to predict the linear variation of elevator passenger flow. Then, RBF neural network was used to predict the nonlinear part. Finally, the results of the two were added and the combined model was used for elevator passenger flow prediction [2]; Wei Xie, Lipeng Zhao, Ming Zeng, Xu Luo and others established a maximum 5-minute passenger flow model based on the moving least squares method for surface fitting, which solves the problem of inaccurate elevator stop numbers in traditional algorithms [3].

With the assistance of machine learning algorithms, it is possible to accurately predict a building's energy demand using multi-source data, including historical energy consumption records, real-time weather conditions, and the activities of occupants within the building. Consequently, energy distribution can be dynamically

---

* Corresponding Author

adjusted to optimize energy utilization [4-5]. For instance, the Long Short-Term Memory (LSTM) network is employed to perform both short-term and long-term predictions of a building's power consumption. Based on these predictions, the operation of equipment such as air conditioners and lighting systems can be effectively regulated. Reference [6] introduced a novel method for predicting building loads utilizing deep learning and Non-Intrusive Load Monitoring (NILM) technology, with the goal of enhancing energy management in smart buildings. References [7-8] conducted real-time monitoring and analysis of operational data from building equipment, including elevators, air conditioners, and water pumps. By employing machine learning algorithms to develop fault prediction models, potential equipment failures can be identified in advance, allowing for timely maintenance that reduces both equipment downtime and maintenance costs.

Elevator passenger flow prediction encompasses both short-term and long-term forecasting. This process relies on various observed time series data collected at specific intervals to establish a predictive model that estimates passenger flow at multiple future time points [9]. Factors such as weekday commuting patterns and holiday schedules contribute to the non-linear and periodic characteristics of elevator passenger flow. Traditional forecasting methods, including time series models, Kalman filtering models, exponential smoothing models (ES), auto-regressive models (AR), and moving average models (ARMA), are often inadequate for effectively analyzing and learning the non-linear dynamics of passenger flow [10]. Consequently, advanced algorithms developed in recent years, such as artificial neural networks, support vector machines, generative adversarial networks, and least squares support vector machines, have gained widespread application in this field [11].

The analysis and prediction of elevator traffic flow characteristics have become the most fundamental issue in the entire elevator group control system. It is also an extremely important yet difficult problem in the performance analysis of the elevator group control system. Moreover, the solutions to the other two major problems in the group control system are closely related to it. Therefore, the research on the analysis and prediction of elevator traffic flow characteristics has become the top priority in today's elevator traffic flow research. If we can start from the analysis of traffic flow characteristics, adopt various means, and study the changing characteristics of elevator traffic flow from different perspectives, and then develop corresponding reasonable and effective prediction methods for different flow characteristics to predict elevator traffic flow, it will surely greatly improve the performance of the elevator group control system. In this paper, through an in-depth analysis of the elevator traffic flow characteristics in the hospital inpatient building and the influencing factors of the pedestrian flow in the inpatient building, we have constructed a prediction model based on random forest regression. This model comprehensively considers various influencing factors such as time distribution characteristics, population composition, and spatial distribution characteristics. It can effectively capture the periodic changes and fluctuations of elevator traffic flow, and has relatively high prediction accuracy. It can effectively improve the transportation effect of elevators, enhance the elevator dispatching efficiency, and to the greatest extent, save time and reduce energy waste.

In summary, our major contributions are:

- Based on the on-site statistics of pedestrian flow in the inpatient building of a tertiary hospital in a certain area, we analyzed the characteristics of pedestrian flow changes in the auxiliary functional areas and nursing units in terms of time and space.
- Based on the changing characteristics of pedestrian flow in various spaces and analyzing the influencing factors of pedestrian flow, we propose an elevator traffic flow prediction model based on forest regression.
- We provide a detailed description of the specific input variables in the elevator traffic flow model. Taking a hospital's inpatient building as an example, we use Python programming language to write code to predict the pedestrian flow in the inpatient auxiliary functional area and nursing unit. We compare the predicted values with the measured values and analyze the accuracy of the prediction model. Through simulation experiments, our proposed method achieves better prediction accuracy than SVM and BP methods.

The remainder of this paper is structured as follows. In Sect. 2, we described the characteristics of elevator traffic flow in the inpatient building. The factors and changing patterns that affect elevator traffic flow are described in Sect. 3. In Sects. 4 and 5, In sections 4 and 5, we proposed a prediction model for elevator traffic flow and a forest regression prediction algorithm, respectively. In Sect. 6, we construct a dataset and compare performance metrics using different algorithms. Finally, we conclude the paper in Sect. 7.

2 Traffic Flow Characteristics of the Inpatient Building in Hospital

In the operation system of modern hospitals, the elevator in the inpatient building is a crucial vertical transportation facility. Its efficient operation not only affects the daily travel experience of patients and medical staff, but also has a profound impact on the overall service quality and operational efficiency of the hospital. Studying the traffic flow characteristics of elevators in inpatient buildings can provide scientific basis for optimizing the configuration and scheduling management of elevator systems, helping hospitals improve their service levels, and alleviate the problem of vertical traffic congestion. The main characteristics of traffic flow in hospital inpatient buildings are as follows:

2.1 Time Distribution Characteristics

Morning and evening rush hour: From 7:00 to 9:00 in the morning, medical staff concentrate on work, patients go out for examination, and family members come to visit, resulting in a sharp increase in demand for elevators, causing serious congestion. From 5:00 to 7:00 in the evening, when medical staff finish work and family members leave the hospital, the downward flow of the elevator significantly increases.

Off peak hours: 9:00 am to 5:00 pm. Except for specific inspection periods, the elevator flow is relatively stable, but there are still small fluctuations due to patient treatment and nursing needs. From late night to early morning, elevator usage significantly decreases, with only a few emergency patients and on duty medical staff using it.

2.2 Flow Characteristics

Upward traffic flow and downward traffic flow differences: The inpatient building is mainly composed of patients and their families, and there is a larger upward traffic flow when patients are admitted and their families visit; When patients are discharged and their families leave, the downward traffic flow dominates. During the examination and treatment period, patients need to travel back and forth between the ward and the examination department, which can increase the two-way flow between specific floors.

Inter-layer traffic flow: As the main entrance and exit of the hospital, the first floor has a large flow of traffic between each floor. The popular floors of the inpatient department, such as the operating room, intensive care unit, and popular department floors, also have relatively high elevator traffic compared to other floors.

2.3 Use Entity Characteristics

Patients and their families: may have limited mobility and need to carry medical equipment, daily necessities, etc. The elevator speed may be slow, which may increase the elevator stopping time. Patients with different medical conditions have different elevator needs, such as emergency patients having higher requirements for elevator timeliness.

Medical staff: Their working hours are relatively fixed, their travel is regular, and they often need to carry medical equipment. Sometimes they also need to push stretcher carts, which have special requirements for elevator space and operational efficiency.

2.4 Congestion Characteristics

Peak hour congestion: During peak hours, the contradiction between elevator supply and demand is prominent, and congestion is prone to occur.

Ward floor congestion: Due to the concentration of patients and their family members on the ward floor, the elevators are used more frequently, which easily leads to congestion.

Elevator hall congestion: In the waiting area of the elevator hall, there is a high density of people, which can easily cause congestion.

Emergency event: In case of emergency or soft pair examination and treatment, the elevator needs to prioritize the passage of emergency patients and medical staff, which may break the normal flow pattern and lead to a sudden increase in flow during specific periods and floors.

Fig. 1 shows the traffic flow statistics of a tertiary hospital's inpatient building during a certain period of time. The traffic flow of the inpatient building consists of three parts: upward traffic flow, inter-layer traffic flow, and downward traffic flow. In the upward traffic flow, passengers take elevators from the lobby to various target floors of the inpatient building; In the downward traffic flow, passengers take elevators from each floor to reach the lobby; In the inter floor traffic flow, passengers take the elevator from their own floor to other floors except for the lobby. The actual traffic pattern for a day is a combination of these three traffic flows.

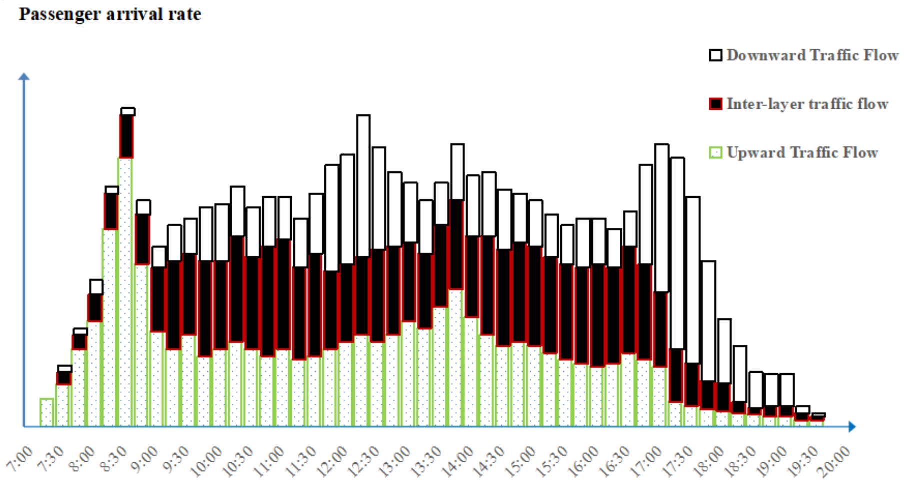


Fig. 1. The traffic flow statistics of a tertiary hospital's inpatient building

This paper analyzes two traffic modes: upward peak traffic and inter-floor peak traffic in high-rise hospital buildings. It also presents elevator scheduling schemes tailored to each traffic mode, aimed at improving elevator operational efficiency and alleviating vertical traffic congestion in hospitals.

3 Analysis of Influencing Factors and Changing Patterns of Pedestrian Flow in Inpatient Buildings

The personnel in the inpatient building include inpatients, caregivers, visitors, doctors, nurses, cleaners, security personnel, etc. Different types of personnel have different activity flow lines and stay times at various nodes in the inpatient building [12, 13].

According to survey statistics, the number of hospitalized patients accounts for approximately 28%-47% of the total number of hospitalizations. Inpatients are classified into severe and mild cases based on the severity of their condition. Severe illness refers to patients who are unable to take care of themselves and require 24-hour companionship. They spend most of the day receiving treatment and are unable to go out for activities. Mild cases refer to patients who can take care of themselves and usually receive treatment in the ward in the morning, and can move freely in the afternoon. According to the age of the patients, they are divided into elderly, middle-aged, teenagers, and infants. People of different age groups may have different types of illnesses, family and social identities, and possible accompanying visitation personnel, as shown in Table 1. The visitation time and number of visitors from different social identities vary. For example, colleagues and classmates usually come to visit in the evening or on weekends, and there are more visitors at once; However, close relatives such as grandparents and parents may visit at any time, and the number is relatively small.

Table 1. Different inpatients of visiting and accompanying persons

Type of patient	Possible disease	Family identity	Social identity	Possible visitors	Possible caregivers
Old people	Diabetes, cardiovascular disease, etc	Elder	Retired and elderly personnel	Children, relatives and friends	Life partner, children, or full-time caregivers
Middle-aged person	Various diseases	Workforce	Office staff	Parents, children, relatives and friends	Spouse

3.1 Factors Affecting the Flow of People in Inpatient Buildings

There are four main imaging factors affecting the flow of people in inpatient buildings.

(1) Patient Related Factors

① Disease type and length of hospital stay: Different disease types determine the length of hospital stay for patients. For example, patients undergoing major surgeries require a long period of recovery after surgery and may frequently use elevators for examination and treatment; Patients with the common cold have a shorter hospital stay and use elevators relatively less frequently. At the same time, certain seasonal high incidence diseases, such as winter flu and summer intestinal diseases, can lead to an increase in hospitalizations during specific periods, thereby affecting elevator passenger flow.

② Patient rehabilitation stage: Patients in the early stages of rehabilitation may need to frequently take elevators accompanied by medical staff to the rehabilitation treatment room and examination department; As the rehabilitation process progresses, their frequency of going out gradually decreases. For example, patients with fractures need to undergo regular X-ray examinations using elevators in the early postoperative period, while in the later stages of rehabilitation, fewer examinations are required.

(2) Factors Affecting Medical Staff

① Scheduling system: The work arrangement of medical staff is usually divided into different shifts, and there may be a large number of medical staff taking the elevator together during shift handover. For example, during the handover of morning and evening shifts, hundreds of medical staff need to arrive at their respective work floors in a short period of time, greatly increasing the frequency of elevator usage.

② Ward rounds and treatment work: Medical staff need to conduct daily rounds and provide treatment services to patients, which makes them frequently travel between different floors. Especially during peak ward rounds, the flow of elevator passengers significantly increases and is mostly concentrated between the floors of the wards.

(3) Visitor Factors

① Visiting hours: The hospital’s designated visiting hours have a significant impact on the flow of elevator passengers. During the allowed visitation period, a large number of visitors flooded into the inpatient building, leading to a surge in demand for elevator usage. For example, most hospitals set the afternoon and evening as visitation hours, during which the elevator is crowded and waiting times are extended.

② Number of visitors: The severity of the patient’s condition will affect the number of visitors. Severely ill patients often attract more relatives and friends to visit, increasing the passenger flow of elevators. In addition, the accompanying habits of patients’ family members also vary, and some family members may spend a long time with the bed, further increasing the burden of elevator usage.

(4) Hospital Management Factors

① Department distribution: Different departments are located on different floors. If the relevant departments are far apart, it may cause patients, medical staff, and visitors to frequently shuttle between floors. For example, separating the operating room from the intensive care unit (ICU) will increase the frequency of elevator usage for postoperative patient transportation and medical staff travel.

② Special event: Large scale free clinics, health lectures and other activities carried out by hospitals will attract a large number of people to enter the inpatient building, causing a sharp increase in elevator traffic during a specific period of time and putting significant pressure on elevator operation.

3.2 The Changing Patterns of Influencing Factors

According to the analysis of statistical results regarding the number of personnel in various hospital inpatient buildings, the flow of nursing staff exhibits both temporal and spatial distribution characteristics.

(1) Time Dimension

The temporal characteristics are categorized into three segments: daily variation, weekly variation, and monthly variation.

From a daily cycle perspective, the flow of people in the inpatient building exhibits a distinct “double peak” pattern. The first peak occurs between 7:30 and 9:30 a.m. on weekdays. During this time, newly admitted patients proceed to their wards after completing necessary procedures, while medical staff focus on their duties, including ward rounds and treatment activities. Concurrently, some patients leave their wards for examinations. This combination of factors results in a rapid increase in elevator traffic. The second peak occurs from 3:00 p.m. to 5:00 p.m., primarily due to the commencement of visitation hours, which sees a significant influx of visitors into the inpatient building. Additionally, medical staff engage in afternoon treatment activities, and patients return to their wards, further elevating the demand for elevator usage. At night, the flow of people is relatively low, and the frequency of elevator use decreases significantly, primarily due to the duty patrols of medical staff and the urgent needs of a limited number of patients.

From a weekly perspective, foot traffic from Monday to Friday is significantly higher than on weekends. During weekdays, various medical activities in the hospital are conducted regularly, with frequent patient admissions, treatments, examinations, and a relatively large number of visitors. On weekends, however, the schedule for elective surgeries and examinations is reduced, leading to a decrease in the number of newly admitted patients. Additionally, some medical staff take turns off, further contributing to the decline in foot traffic. Nevertheless, the number of visitors on weekends should not be underestimated, particularly on Sunday afternoons, which typically represent the peak period for visits.

From the perspective of monthly and annual cycles, hospitals experience certain seasonal and cyclical fluctuations in their business volume. During specific periods, such as the peak influenza season and outbreaks of infectious diseases, the number of hospitalized patients significantly increases, leading to a sustained high volume of traffic in the inpatient facilities. Additionally, at the end of the year and the beginning of the new year, many patients tend to concentrate on completing admission and discharge procedures for reimbursement and other reasons, which can also result in notable fluctuations in patient flow during this time.

(2) Space Dimension

In terms of spatial distribution, the flow of individuals across various floors of the inpatient building is uneven. The floors connected to outpatient services and medical technology departments, as well as those housing critical areas such as intensive care units, operating rooms, and obstetrics and gynecology, experience relatively high foot traffic. These areas have a significant demand for elevator usage due to the concentrated entry and exit of patients, frequent movement of medical staff, and the presence of visitors.

4 Elevator Traffic Flow Prediction Model

The random forest regression algorithm is an ensemble learning method based on decision trees that enhances the accuracy and robustness of the model by combining the predictions of multiple decision trees. Its core principle, encapsulated in the saying “Three stinky cobblers, one Zhuge Liang”, two key characteristics: ensemble learning and randomness [14, 15]. Random forest is categorized under the “bagging” method in ensemble learning, which involves training a large number of distinct decision trees to make independent predictions. The final output is obtained by averaging all predicted results (in the case of regression), akin to “a group of people voting to determine the result”. collective intelligence to minimize the errors associated with a single model [16]. During the training process of each tree, two types of randomness are introduced: sample randomness and feature randomness. This ensures that each tree possesses unique characteristics, preventing all trees from making the same errors and resulting in a more stable overall prediction.

4.1 Data Representation

Assuming we have a training data-set:

$$\mathrm{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(\mathrm{N})}, y^{(\mathrm{N})})\} \quad \mathbf{(1)}$$

Among them, $x^{(i)}$ is the feature vector of the i-th sample, $y^{(i)}$ corresponds to the elevator traffic flow.

4.2 Random Forest Construction

Random forest consists of B decision trees, each of $\mathrm{T}_b(x;\Theta_b)$ is constructed through bootstrap sampling and feature subset selection on data-set D.

4.3 Decision Tree Construction

Each tree T_b can be represented as a series of decision rules that divide the feature space into multiple regions and provide predicted values on each region. It can be expressed mathematically as:

$$T_b(x;\Theta_b) = \sum_{m-1}^{M_b} c_{b,m}\Pi(x \in R_{b,m}) \tag{2}$$

Among them, Θ_b is a parameter of tree b, $R_{b,m}$ is the m-th region defined by the decision rule, $c_{b,m}$ is the predicted value for this region, $\prod$ is an indicator function.

4.4 Random Forest Prediction

The prediction of a random forest is the average of the predictions of all trees. Prediction of Random Forest for a New Input Sample x:

$$\hat{y}_{RF}(x) = \tfrac{1}{B}\sum_{b-1}^{B}(\sum_{m-1}^{M_b} c_{b,m}\Pi(x \in R_{b,m})) \tag{3}$$

4.5 Training Process

During the training process, for each tree T_b, we will perform the following steps:
(1) Perform self sampling from dataset D to obtain a subset D_b;
(2) For each node, randomly select a subset of features;
(3) Find the optimal splitting point and divide the nodes accordingly;
Repeat the above process until the termination condition of the tree is reached.

4.6 Prediction Process

In the prediction process, for a new input sample x, we perform the following steps:
(1) For each tree in the forest , find the leaf node where x falls based on its characteristics;
(2) Output the predicted value of the leaf node;
(3) Calculate the average of the predicted values of all trees to obtain the prediction for the random forest.

An abstract representation of the random forest regression model, practical implementation will involve more details such as tree construction, selection of splitting rules, calculation of predicted values, etc [17].

5 Random Forest Regression Elevator Traffic Flow Prediction Algorithm

The elevator traffic flow prediction based on the random forest regression model is shown in Algorithm 1, which mainly consists of the following steps:

(1) Problem definition. Our goal is to predict the traffic flow of elevators during a specific time period, which usually involves input features such as time (hour, day of the week, etc.), weather conditions, special events, building types, traffic flow patterns, etc [18, 19]. The specific features used as input to the random forest model mainly include:

Algorithm 1. Random forest regression elevator traffic flow prediction algorithm

```
Input: D,B,M,N
# D = Training data X_train (features X,Y)
# B = The number of decision trees
# M = The number of features considered during each segmentation
# N = The number of samples in the training data-set
Output：random_forest
random_forest = []
# For each tree b, from 1 to B
for b in range(1, B + 1):
    # Self sampling from training dataset
    Db = bootstrap_sample(D)
    Tb = create_tree()
        # Select the best segmentation point and recursively construct a tree
    build_tree(Tb, Db, M)
        # Add the constructed tree to the random forest
    random_forest.append(Tb)
# Define the function for building a tree
function build_tree(tree, data, M):
# If the stopping condition is met (e.g. node purity is high enough or sample size less than the threshold) return
    if stopping_condition(data):
        set_leaf_node(tree, data)
        return
        # Randomly select M features
    features = random_select_features(data.features, M)
        # Find the best segmentation point
    best_split = find_best_split(data, features)
        # Split dataset
    left_data, right_data = split_data(data, best_split)
       build_tree(tree.left, left_data, M)
       build_tree(tree.right, right_data, M)
# Define prediction function
function predict(random_forest, X):
    predictions = []
    for tree in random_forest:
        prediction = tree_predict(tree, X)
        predictions.append(prediction)
        # Calculate the average of all tree predictions
    mean_prediction = mean(predictions)
    return mean_prediction
# Define the prediction function for a single tree
function tree_predict(tree, X):
    # Starting from the root node of the tree, recursively traverse the tree until reaching the leaf nodes
    while not is_leaf_node(tree):
        if X[tree.feature] <= tree.split_point:
            tree = tree.left
        else:
            tree = tree.right
        # Return the value of the leaf node (predicted result)
    return tree.value
```

- Hour: represents the number of hours in a day, with a value range of 0 to 23. In the scenario of predicting pedestrian flow in elevators, different hours may correspond to different pedestrian flow patterns. For example, the frequency of elevator usage may be higher during the morning rush hour (such as 7-9 am) and evening rush hour (such as 17-19 pm), while the frequency of elevator usage may be lower during late night hours (such as 0-4 pm). Design consideration: By using hours as a feature, it is possible to capture the impact of different time periods of the day on elevator pedestrian flow.

- Is_weekday: This is a binary feature with a value of 0 or 1. Among them, 0 represents non working days (such as weekends), and 1 represents working days. There is usually a significant difference in personnel activity patterns between workdays and non workdays. On workdays, people mainly engage in work-related activities, while on weekends, they tend to engage in more leisure activities. Design consideration: Introducing this feature can distinguish the impact of working days and non working days on elevator pedestrian flow.
- Is_visiting_hour: It is also a binary feature, where 0 represents non visitation time and 1 represents visitation time. In hospitals and other places, there will be more personnel flow during visitation hours, which will directly affect the use of elevators. Design consideration: This feature is used to capture the impact of the special factor of visitation time on elevator pedestrian flow.
- Numb_of_inpatients: represents the number of hospitalized patients, ranging from 100 to 500. The number of hospitalized patients will affect the flow of personnel within the hospital, and the activities of patients and their families will use elevators. Therefore, the more hospitalized patients there are, the greater the demand for elevator use may be. Design consideration: Using the number of hospitalized patients as a feature can reflect the impact of patient related activities within the hospital on elevator pedestrian flow.
- Is_peak_interfloor: binary feature, 0 represents non inter layer peak period, 1 represents inter layer peak period. The peak period between floors may be related to medical activities of medical staff or patients between different floors, resulting in frequent movement of personnel between floors, and the use of elevators will be more frequent at this time. Design consideration: This feature is used to capture the impact of the peak period between floors on elevator pedestrian flow.

(2) Data collection. Collect historical elevator usage data, including passenger flow records of elevators and related factors such as time and environment.

(3) Data preprocessing. Including data cleaning, feature engineering, and data segmentation.

(4) Construction of Random Forest Model. Including: self sampling, tree construction, forest integration.

(5) Model training. Train a random forest model using the training set through the above process. Each tree learns how to predict elevator traffic flow based on input features during the training process.

(6) Model validation. Evaluate the performance of the model using a test set and calculate performance metrics such as mean square error (MSE), root mean square error (RMSE), or coefficient of determination (R^2).

(7) Model prediction. For new input data, each tree will provide a predicted value. The final prediction of a random forest is the average of the predicted values of all trees.

This piece of pseudo code implements a random forest regression model. Random forest is an ensemble learning method that consists of multiple decision trees. Each decision tree is trained on a dataset obtained through bootstrap sampling, and a subset of features is randomly selected when splitting each node. The final prediction is the average of the predictions from all decision trees. By utilizing bootstrap sampling and random feature selection, the random forest can reduce over-fitting, enhance the model's generalization ability, and increase both stability and accuracy [20-22]. The specific implementation process is as follows:

5.1 The Number of Tree

In this code, the number of trees is controlled by the hyper parameter n_estimators. In the hyper parameter tuning section, different values were set for n_estimators using the param_dest dictionary, specifically [50, 100, 200, 300]. RandomizedSearchCV will randomly select 10 sets of n_iter from these candidate values to try and find the optimal n_estimators value.

```
param_dist = {
'n_estimators': [50, 100, 200, 300],
'max_depth': [5, 10, 15, 20, None],
'min_samples_split': [2, 5, 10],
'min_samples_leaf': [1, 2, 4]
}
```

5.2 Tuned the Parameters

(1) Max_depth: It specifies the maximum depth of the decision tree. If set to None, the decision tree will continue to grow until all leaf nodes are pure nodes or reach the minimum sample size limit. The candidate values for x_depth in the code are [5, 10, 15, 20, None].

(2) Min_Samples_split: This parameter represents the minimum number of samples required to split internal nodes. When the sample size of a node is less than this value, it will no longer be split. The candidate values for min_stamples_split in the code are [2, 5, 10].

(3) Min_Samples_leaf: It refers to the minimum number of samples required for leaf nodes. If the sample size of a leaf node is less than this value, it may be pruned. The candidate values for min_stamples_leaf in the code are [1, 2, 4].

5.3 Cross Validation Strategy

The code uses a combination of random search and cross validation to find the optimal hyperparameters. The specific cross validation strategy used is the cv parameter in RandomizedSearchCV, which is set to cv=5, that is, 5-fold cross validation.

The specific method of 5-fold cross validation is to divide the training dataset into 5 subsets of similar size, selecting 1 subset as the validation set and the remaining 4 subsets as the training set each time. This will repeat the training and validation process 5 times. Finally, the optimal hyper parameter combination is determined by evaluating the average performance of these 5 validation results.

The time complexity of the random forest algorithm mainly depends on the construction and prediction processes of the decision trees. In the worst-case scenario, the depth of a decision tree may reach O(N). At this time, the time complexity of constructing the random forest is [Here you didn’t provide the complete complexity expression, it should be something like $O(BMN^2)$ according to the previous analysis, you can fill it in as appropriate], and the time complexity of prediction is O(BN). However, in practical applications, the depth of the decision tree is usually much smaller than the number of samples, so the actual time complexity will be much lower than the worst-case scenario.

6 Simulation and Evaluation

The hardware environment of the experiment adopts a computer equipped with Intel Core I7 processor and 16GB memory to ensure the efficiency of data processing and model training. In terms of software [23-24], Python programming language is used, and powerful open source libraries such as Scikit learn, Pandas, Numpy are utilized for data processing, model building, and evaluation.

6.1 Experimental Dataset and Partitioning

The experimental dataset is derived from the historical operational data of the elevator control system in the hospital’s inpatient building. This dataset records essential information, including elevator operating times, floor stopping conditions, and the number of passengers entering and exiting. The collected elevator traffic flow data is normalized and divided into three subsets: a training set, a validation set, and a test set, with proportions of 70%, 15%, and 15%, respectively. The training set is utilized to train a random forest regression model and to identify potential patterns within the data. The validation set is employed to fine-tune the model’s hyper-parameters, thereby preventing over-fitting. Finally, the test set is used to assess the model’s generalization ability and to ensure its predictive accuracy on previously unseen data.

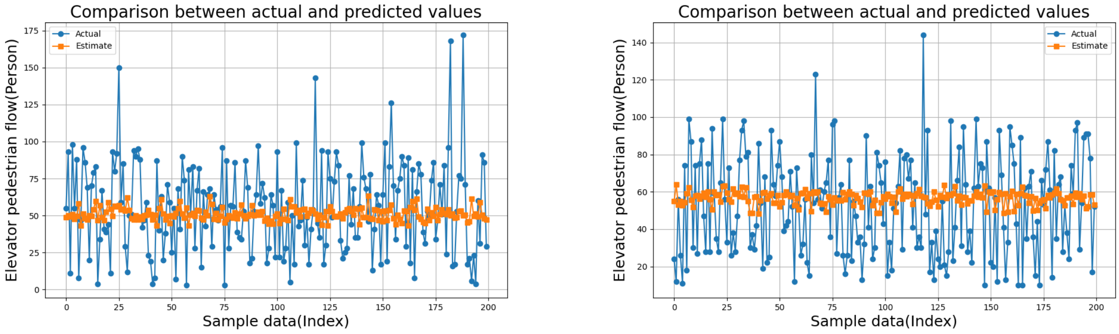
6.2 Parameter Settings

When constructing a random forest regression model, it is necessary to set the key hyper-parameters of the model. The number of decision trees (n_estimators) is set to 100 to balance the complexity and computational efficiency of the model; Set the maximum depth (x_depth) to 10 to prevent over-fitting caused by the decision tree being too deep; The minimum number of sample partitions (min_stamples_split) is set to 2, and the minimum number of sample leaf nodes (min_stamples_leaf) is set to 1. In addition, a combination of grid search and cross validation is used to optimize hyper-parameters and find the optimal parameter combination.

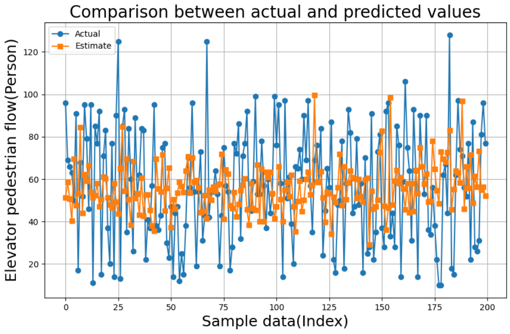
6.3 Model Evaluation and Result Analysis

Selection of Evaluation Indicators. To comprehensively evaluate the predictive performance of the random forest regression model, mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) were selected as evaluation indicators. MSE and RMSE are used to measure the average sum of squared errors between predicted and true values, MAE reflects the average absolute error between predicted and true values, and R^2 is used to evaluate the model's goodness of fit to the data.

Comparison and Analysis of Experimental Results. We will compare the prediction results of the Random Forest Regression (RDF) model with other mainstream prediction models such as SVM, BP neural network, etc. for the two traffic modes of upward peak traffic mode and inter floor peak traffic mode in high-rise hospital buildings. The experimental results are shown in Fig. 2-Fig. 3 and Table 2-Table 3. Through the analysis of experimental results, it was found that the random forest regression model performed well in all evaluation indicators, with lower MSE, RMSE, and MAE values, as well as higher R^2 values, proving the superiority of the model in predicting elevator traffic flow in inpatient buildings of tertiary hospitals.

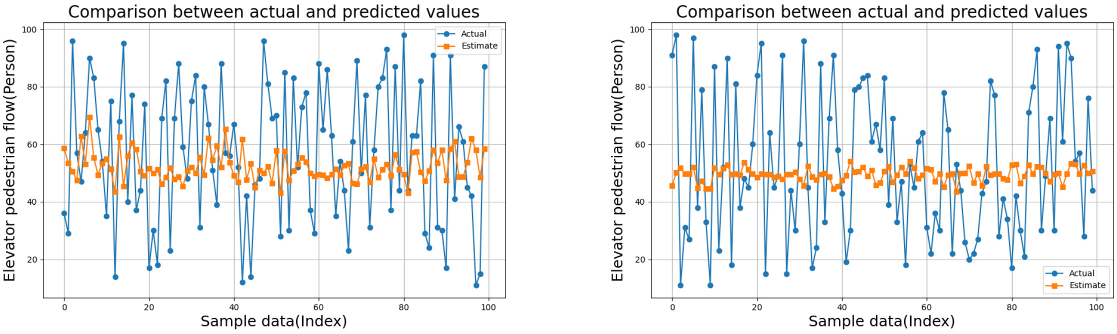


(a) BP-NN (b) SVM

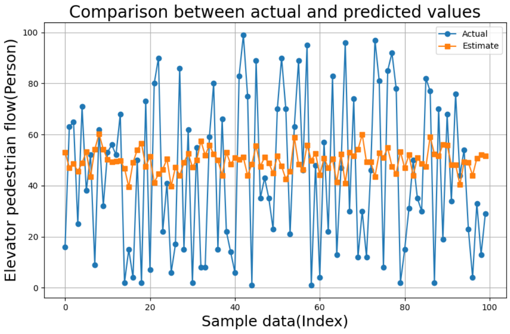


(c) RDF

Fig. 2. Comparison of prediction results of different algorithms during the upward traffic peak period



(a) BP-NN (b) SVM



(c) RDF

Fig. 3. Comparison of Prediction results of different algorithms during the inter-layer peak period

Table 2. Comparison of performance indicators for predicting elevator traffic flow during peak hours

Method Metrics	MSE	RMSE	MAE	R²
SVM	32.56	27.69	23.82	0.18
BP-NN	29.81	24.36	23.64	0.103
Random_forest	8.44	2.91	2.43	0.241

Table 3. Comparison of performance indicators for predicting peak elevator traffic flow between floors

Method Metrics	MSE	RMSE	MAE	R²
SVM	59.6	30.74	26.81	0.053
BP-NN	49.6	28.5	24.2	0.038
Random_forest	36.2	24.71	21.19	0.078

7 Conclusion

Conducting research on the prediction and simulation of elevator traffic flow in high-rise inpatient buildings of hospitals is of paramount importance for enhancing the operational efficiency and service quality of hospital elevator systems. Extensive studies have revealed that multiple factors significantly impact elevator traffic flow. The number of hospitalizations directly correlates with the demand for vertical transportation, as admitted patients frequently require movement between different floors for medical treatments, examinations, and visits.

Outpatient visits also play a crucial role, as the influx of patients coming for consultations, tests, and follow-ups can create peak traffic periods, especially during specific times of the day or days of the week. Moreover, the duration of these visits affects the traffic flow dynamics; longer stays may lead to a more dispersed demand throughout the day, while shorter, concentrated visits can result in sudden surges in traffic.

By leveraging advanced data analytic and mathematical models to accurately predict elevator traffic patterns and meticulously analyzing these key influencing factors, hospitals can formulate highly targeted elevator scheduling strategies. For instance, during peak admission seasons, hospitals can allocate more elevators to serve the inpatient floors, reducing the congestion and wait times for patients and their families. Similarly, during busy outpatient hours, elevators can be programmed to prioritize movement between the ground floor and outpatient departments. This approach not only empowers hospitals to efficiently meet the passenger demand during peak hours, thereby significantly reducing wait times and substantially improving passenger satisfaction but also aligns seamlessly with the hospital's efforts towards refined management practices. It allows for better resource allocation, optimized staff movement, and enhanced overall hospital operations.

In terms of engineering deployment, APIs are developed using robust Flask or Django frameworks. These frameworks offer flexibility and scalability, making them ideal for integrating with the hospital's complex IoT systems. Once integrated, the APIs enable the seamless and real-time acquisition of sensor data from various sources, including elevator control systems, floor access sensors, and patient movement tracking devices. This data is then processed to generate accurate predictions of elevator traffic flow. To further enhance the utility of these predictions, feedback mechanisms are incorporated, allowing the system to continuously refine its forecasting models based on actual traffic patterns.

In addition, powerful data visualization tools such as Tableau and Power BI are employed to create an intuitive dashboard. This dashboard provides a visual comparison of predicted and actual elevator traffic flow across different time periods, such as hours of the day, days of the week, and months of the year. The visual representation makes it easy for management to quickly identify trends, anomalies, and areas of improvement. For example, if there is a significant deviation between the predicted and actual traffic during a particular time slot, management can investigate the root cause, whether it's an unexpected event, a malfunctioning sensor, or an inefficient scheduling algorithm. This dashboard also offers real-time monitoring services for management, ensuring that they are constantly informed about the status of the elevator system. This deployment approach equips management personnel with the ability to promptly identify potential issues in the elevator scheduling system, providing a solid quantitative foundation and comprehensive technical support for the continuous optimization and enhancement of the elevator system in the hospital's inpatient building. As a result, the hospital can ensure smoother operations, improved patient experience, and more efficient use of its vertical transportation resources.

8 Acknowledgement

It is supported by Hunan University of Medicine Doctoral Research Initiation Fund (202407), Special Fund Project of Science and Technology Innovation Strategy of Guangdong Province (Grant number pdjh2024b565, pdjh2025bg326), and Guangdong Provincial Universities Research Platforms and Projects (Grant number 2025ZDZX1067).

References

[1] L.T. Yuan, J. Zhang, Simulation Research on Forecast Method of Elevator Traffic Flow, Computer Simulation 29(10) (2012) 236-239.
https://doi.org/10.3969/j.issn.1006-9348.2012.10.057

[2] Y.H. Zhang, Simulation Research on Elevator Traffic Flow Prediction Based on Hybrid Model, Computer Simulation 28(11)(2011) 192-195.
https://doi.org/10.3969/j.issn.1006-9348.2011.11.047

[3] W. Xie, K.P. Zhao, M. Zeng, X. Luo, Analysis of Elevator Traffic Based on Moving Least-Square, Journal of Beijing Jiaotong University 36(5)(2012) 139-142.

[4] M. Huotari, A. Malhi, K. Främling, Machine Learning Applications for Smart Building Energy Utilization: A Survey, Archives of Computational Methods in Engineering 31(5)(2024) 2537-2556.
https://doi.org/10.1007/s11831-023-10054-7

[5] H.P. Das, Y.-W. Lin, U. Agwan, L. Spangher, A. Devonport, Y. Yang, J. Drgoňa, A. Chong, S. Schiavon, C.J. Spanos, Machine Learning for Smart and Energy-Efficient Buildings, Environmental Data Science 3(2024) e1. https://doi.org/10.1017/eds.2023.43

[6] Z. Cheng, Z. Yao, A novel approach to predict buildings load based on deep learning and non-intrusive load monitoring technique, toward smart building, energy 312(2024) 133456. https://doi.org/10.1016/j.energy.2024.133456

[7] A. Shokrollahi, J.A. Persson, R. Malekian, A. Sarkheyli-Hägele, F. Karlsson, Passive Infrared Sensor-Based Occupancy Monitoring in Smart Buildings: A Review of Methodologies and Machine Learning Approaches, Sensors 24(5)(2024) 1533. https://doi.org/10.3390/s24051533

[8] L. Yi, X. Xu, Y. Wang, J. Liu, Y. Gao, X. Liu, Non-Intrusive Load Decomposition for Smart Buildings Based on Parallel Connectivity Networks and Attention Mechanism, Journal of Electrical Engineering & Technology 19(8) (2024) 4833-4845. https://doi.org/10.1007/s42835-024-01939-z

[9] C.G. Ying, Based on the Analysis of Vertical Passenger Traffic Optimization of Elevator Group Control, [dissertation] Hangzhou: Zhejiang University of Technology, 2012. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD2012&filename=1012323307.nh

[10] L.S. Yang, Design and Analysis of Elevator Group Control System Using FPGA, [dissertation] Shanghai: Shanghai Jiao Tong University, 2007. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD2007&filename=2007154654.nh

[11] Z. Zhang, Y. Wang, Elevator Selection Factors under Vertical Traffic Pressure in a General Hospital in Shenzhen, Chinese Hospital Architecture & Equipment 25(12)(2024) 9-13. https://doi.org/10.3969/j.issn.1671-9174.2024.12.002

[12] C. Ren, P. Jia, Short-Time Traffic Flow Prediction Based on a Combined GAT-VMD-LightGBM Prediction Model, in: Proc. Sixth International Conference on Traffic Engineering and Transportation System (ICTETS 2022), 2022. https://doi.org/10.1117/12.2668496

[13] G.C. Barney, S. Dos Santos, Elevator Traffic Analysis, Design and Control, London: Peregrinus on behalf of the Institution of Electrical Engineers, 1985.

[14] L.X. Ji, X.G. Fu, The Application of LS-SVM to the Prediction of the Elevator Transportation Flow, Journal of Shanghai Dianji University 9(3)(2006) 62-64.

[15] Q. Zong, Y. Yue, A Traffic Flow Forecasting Method of Elevator Group Control System, Systems Engineering and Electronics 23(7)(2001) 103-105. https://doi.org/10.3321/j.issn:1001-506X.2001.07.030

[16] J. Liu, W. Guan, A Summary of Traffic Flow Forecasting Methods, Journal of Highway and Transportation Research and Development 21(3)(2004) 82-85. https://doi.org/10.3969/j.issn.1002-0268.2004.03.022

[17] R. Xu, D. Wunsch, Survey of clustering algorithms, IEEE Transactions on Neural Networks 16(3)(2005) 645-678. https://doi.org/10.1109/TNN.2005.845141

[18] G.Z. Tang, G.M. Zhang, W. Zhu, EGCS Traffic-Pattern Identification, Elevator World (12)(2005) 116-120.

[19] M. Shi, E. Sun, X. Xu, Y. Choi, Prediction and Analysis of Elevator Traffic Flow under the LSTM Neural Network, Intelligent Control and Automation 15(2)(2024) 63-82. https://doi.org/10.4236/ica.2024.152004

[20] Y. Wu, J. Yang, Directional Optimization of Elevator Scheduling Algorithms in Complex Traffic Patterns, Applied Soft Computing 158(2024) 111567. https://doi.org/10.1016/j.asoc.2024.111567

[21] S. Khosa, F. Rustam, A. Mehmood, G.S. Choi, I. Ashraf, Incorporating Word Embedding and Hybrid Model Random Forest Softmax Regression for Predicting News Categories, Multimedia Tools and Applications 83(11)(2024) 31279-31295. https://doi.org/10.1007/s11042-023-16491-7

[22] A.J. Ramadhan, S.K. Priya, N. Naranammal, Suman, P. Lal, P. Mishra, M. Abotaleb, H. Alkattan, Use of Random Forest Regression Model for Forecasting Food and Commercial Crops of India, in Proc. Fifth International Scientific Conference of Alkafeel University (ISCKU 2024), 2024. https://doi.org/10.1051/bioconf/20249700130

[23] Y.O. Ouma, B. Nkwae, P. Odirile, D.B. Moalafhi, G. Anderson, B. Parida, J. Qi, Land-Use Change Prediction in Dam Catchment Using Logistic Regression-Ca, Ann-Ca and Random Forest Regression and Implications for Sustainable Land–Water Nexus, Sustainability 16(4)(2024) 1699. https://doi.org/10.3390/su16041699

[24] M. Wahba, R. Essam, M. El-Rawy, N. Al-Arifi, F. Abdalla, W. M. Elsadek, Forecasting of Flash Flood Susceptibility Mapping Using Random Forest Regression Model and Geographic Information Systems, Heliyon 10(13)(2024) e33982. https://doi.org/10.1016/j.heliyon.2024.e33982