Eng. Sci., 2024, **29**, 1145

Engineered Science
DOI: https://dx.doi.org/10.30919/es1145

A Novel Approach for Route Prediction in Multimodal Transport Networks: A Monte Carlo Simulation and Long Short-Term Memory-based Model

Surya Prakash* and Bibhya Sharma*

Abstract

This research introduces a pioneering approach for route prediction in multimodal transport networks, leveraging the capabilities of Artificial Intelligence (AI). At the core of the proposed methodology is the integration of Monte Carlo simulations with Long Short-Term Memory (LSTM) networks, aimed at understanding and predicting complex route dynamics in multimodal transport systems. This combination aims to capture the intricate dynamics and uncertainties of multimodal transport, utilizing LSTM's strength in processing temporal sequences and recognizing data dependencies over time. Our obtained results show that the model is predicting feasible routes given starting sequences and those predicted routes are optimal and match the data that the model was trained in. The results also show that the obtained model generates plausible predictions for new, unseen sequences thus demonstrating remarkable adaptability. This demonstrates the model's real learning potential, beyond mere data memorization. The integration of Monte Carlo simulations and LSTM networks represents a significant step forward in multimodal transport route prediction, opening avenues for future enhancements and broader applications in transportation studies, including navigating network obstacles. This research not only contributes to the field by improving route prediction accuracy but also by suggesting a framework for future developments in the domain.

Keywords: Artificial Intelligence; AI; freight; logistics; Monte carlo simulations; Multimodal; Long short-term memory; LSTM; Route selection; Transportation.
Received: 05 February 2024; Revised: 05 April 2024; Accepted: 17 April 2024.
Article type: Research article.

1. Introduction

Multimodal freight transportation is often defined as the transportation of goods carried out by a sequence of at least two different modalities. Multimodal transportation networks are complex systems with a wide variety of interconnected elements and dynamic properties. Predicting accurate routes in these networks is a challenging task due to the inherent variability and unpredictability of factors such as traffic, weather conditions, and user behavior. Route prediction has significant applications in areas such as traffic management,[1] logistics,[2] and autonomous vehicles[3] contributing to efficiency and safety in transportation.

Another term that is used interchangeably with multimodality is intermodality. Multimodal freight transportation involves at least two modes of transit where goods can be transported in boxes, pallets, cages, containers, or as loose goods put directly onto trucks, vessels, or aircraft. This definition is wider and more inclusive since it highlights the major elements of multimodal commodities transit. Intermodal freight transportation, on the other hand, involves transporting commodities via at least two modes without changing the load unit. In intermodal transportation, products are loaded into the load unit (typically a container) at the origin (shipper warehouse). After switching modes, the transportation unit unloads cargo at the delivery destination.[4] For this paper, both the areas of multimodal and intermodal freight transportation are considered since they both involve more than 2 modes; however, we will use the general term of multimodal.

The emergence of deep learning technology has profoundly transformed various application areas, including of route prediction in transportation networks. Long Short-Term Memory (LSTM) networks, which are a sort of recurrent neural network, have demonstrated encouraging outcomes

School of Information Technology, Engineering, Mathematics and Physics, The University of the South Pacific, Suva 1168, Fiji.
* *Email:* Surya.prakash@usp.ac.fj (S. Prakash), Bibhya.sharma@usp.ac.fj (B. Sharma)

because of their capacity to record temporal relationships in sequential data. However, when it comes to applying the technique to anticipate routes, there are still numerous obstacles to overcome, including the need for large amounts of training data and the inability to consider uncertainty and variability in the transportation environment. Various approaches, such as transfer learning,[5] Reinforcement Learning (RL),[6] Decision Trees (DTs),[7] Genetic Algorithms (GAs)[8] etc have been applied to overcome this data need issue. In this paper, Monte Carlo simulation methods (MCS) have been investigated as a way to tackle data need issues to train LSTM models. This choice was made due to several synergistic effects between the stochastic nature of Monte Carlo methods and the predictive capabilities of LSTMs. This combination appears particularly effective for several reasons including:

- Scenario Analysis: MCS enables the exploration of numerous scenarios by repeatedly sampling from the probability distributions of variables affecting route planning. By integrating these simulations with LSTM predictions, it's possible to assess the impact of various scenarios on route efficiency, safety, and reliability. This comprehensive analysis helps in identifying optimal routes that are resilient to different potential future conditions.
- Optimization of Multiple Objectives: Route planning often involves balancing multiple objectives, such as minimizing travel time, reducing fuel consumption, and avoiding risky conditions. The MCS, coupled with LSTM predictions, can optimize routes considering these multiple objectives by evaluating a vast number of possible routes under various conditions and identifying those that best meet the desired criteria.

MCS are extensively utilized across various disciplines due to their capacity to manage intricate systems and measure uncertainty by producing a significant number of potential outcomes based on probability distributions. The diverse applications of MCS include research where MCS is applied to benefit-cost analysis[9] and crash cost estimation.[10] Other examples of its diversity include the application of MCS for selecting risk response strategies in project management,[11] emergency relief transportation strategies[12] and to generate synthetic schedules for large-scale microsimulations of travel behavior.[13] The MCS, renowned in transport research for modeling uncertainty and generating diverse scenarios, has been a focal point of studies that explored its potential to create extensive routing possibilities[14-16] and estimating the reliability of network travel times.[17]

The contributions of this research are detailed as follows:

- Methodological Innovation: Introducing a novel approach that significantly improves the accuracy of route predictions within multimodal transport networks. This method showcases analytical depth by integrating the MCS with the LSTM networks and capturing the dynamic interplay of multiple factors influencing transport routes. This innovation not only enhances prediction accuracy but also broadens the understanding of complex transport dynamics, setting a new benchmark for AI applications in multimodal transportation planning;
- Empirical Validation: Our findings, supported by rigorous empirical analysis, validate the effectiveness of combining the MCS with the LSTM networks for route prediction. They also establish new benchmarks for methodological accuracy in this field.
- Broader Implications: The implications of our research extend beyond transportation. They offer insights into the application of AI in urban planning and infrastructure management, paving the way for more resilient and adaptable urban environments such as smart cities.

In this paper, we present a novel approach that combines the power of the MCS and the LSTM networks to predict routes in multimodal transport networks. The proposed method utilizes the MCS to generate a comprehensive set of training samples, encompassing various possible scenarios in the network. The MCS captures a broader range of scenarios in multimodal transport networks, including rare but critical cases. This approach enriches the training dataset, enabling the model to potentially achieve a more comprehensive understanding of complex transport dynamics. These training datasets are processed and then used to train an LSTM model capable of learning from the temporal dependencies in the data to make accurate predictions. The proposed hybrid approach aims to improve the robustness and accuracy of route prediction models, making a significant contribution to the field of transport network modeling.

Route prediction, a key focus in transportation research, has experienced significant advancements. Our research on AI-driven multimodal route prediction shares a foundational concept with the classical Travelling Salesman Problem (TSP) the pursuit of optimal route selection amidst complex constraints. While the TSP aims to find the shortest possible loop that visits each location exactly once, our work extends this quest into the realm of multimodal transportation systems, where the challenge is not only to optimize the route based on distance or time but also to navigate dynamically changing conditions such as traffic, weather, and transportation mode availability.

The domain of multimodal routing problem (MRP) has witnessed substantial evolution, particularly with the advent of multi-objective multimodal transport routing models. This literature review encapsulates the significant contributions to this field. We have chosen to focus on more recent research publications, mainly post 2014. For an extensive review of the literature preceding 2014, please refer to the review paper.[4] Due to the huge coverage of this subject area, the literature review is confined to publications involving more than two modalities.

When more than two modes of transportation are used, operational problems arise because of the need to combine different legs, which are covered by different modes of transportation, in a way that satisfies required conditions including being economical, providing better service to customers and/or being environment friendly.

The process of determining the optimal routes is often approached as a complex issue known as the multimodal multi-commodity flow problem,[18] where it is tackled through a MILP (Mixed Integer Linear Programming) formulation solved via Lagrangian relaxation. This paper focuses on the calculation of transportation costs, considering economies of scale (EOS) and the total weight transported. It proposes a mathematical model for the multi-objective multimodal multicommodity flow problem (MMMFP) with time windows and concave costs, formulated as an NP-hard problem. The algorithm uses relaxation and decomposition techniques for efficient solutions. The work of Xiong and Wang[19] introduced a two-level Taguchi genetic algorithm for a multimodal routing problem with time windows. The model consisted of two optimal objectives: multiple available transportation manners and different demanded delivery times. The upper level finds global Pareto-optimal solutions and provides feasible paths for the lower level. The lower level finds local Pareto-optimal solutions and provides the best combinations of transportation manners. Tao *et al*.[20] focused on a column generation method. It addresses the route planning problem in fourth party logistics (4PL), which involves selecting logistics companies by a 4PL provider to optimize routes of delivering goods through a transportation network. The concept of 4PL emerged in response to the shortfall in service capabilities of traditional third-party logistics and has been proven capable of integrating logistics resources to fulfill complex transportation demands. A mixed-integer programming model is established for the planning problem with setup cost and edge cost discount policies. A column generation approach combined with a graph search heuristic is proposed to efficiently solve the problem. Darayi *et al*.[21] proposed an optimization formulation to maintain the productivity of the desired region's economy by integrating a multi-commodity network flow model with a risk-based economic interdependency model. It captures failure propagation in interconnected industries, defines adaptive capacity for rerouting strategies, and incorporates economic elements to study disruption effects on infrastructure networks from various perspectives. Gendron *et al*.[22] achieved a notable theoretical advancement in multi-commodity network flow challenges by refining the MILP model using discretization techniques. This refinement added an integral flow constraint on network arcs and introduced various effective inequalities, with solutions found through Lagrangian relaxation to produce accurate bounds. This work[23] presents a new method for solving the bi-objective multi-commodity minimum cost flow problem (BMCMCF) problem by integrating Dantzig-Wolfe decomposition with the bi-objective simplex method. The method reformulates the problem into a bi-objective master problem and several linear fractional sub-problems over network constraints. The algorithm iteratively moves from one non-dominated extreme point to the next one by finding entering variables with the maximum ratio of improvement of the second objective over deterioration of the first objective. The master problem iteratively updates cost coefficients for the fractional sub-problems, and based on these cost coefficients, an optimal solution for each sub-problem is obtained. Baykasoğlu *et al*.[24] developed a MILP framework aimed at a multi-objective load planning problem with a sustainability emphasis, suitable for crafting solutions for intermodal transportation involving road, rail, and sea services. Their model considers import/export load flows to satisfy the transport demands of customers and other related issues and uses multiple objective optimization procedures to handle conflicting objectives simultaneously under crisp and fuzzy decision-making environments. An example of enhancing efficiency through coordinating different transportation modes is where the Rotterdam port is analyzed using a comparison between traditional intermodal and the more recent synchromodal approaches.[25] The challenge here is a scheduling problem that leverages synchromodality's benefits to seamlessly transition between modes, reduce idle times, improve customer service, and lower emissions. Behdani *et al*.[26] explored a scheduling issue focusing on the coordination of truck, train, and barge services. This study discussed the integrated design of synchromodal hinterland transport services to improve the performance of freight systems. The ability to utilize multiple modes of transport in an orchestrated manner has a considerable impact on the accessibility of seaports. A robust method for determining optimal routes in multimodal transport networks that can be used amidst disruptions in the network was proposed while using a tiered table, serving as a dynamic, readily available resource for optimal route selection.[27]

In recent years, the focus has shifted towards more integrative and environmentally conscious models. Demir *et al*.[28] introduced a bi-objective linear mixed-integer mathematical formulation for green intermodal service network design, prioritizing cost and CO_2 emission reduction.

Chen *et al*.[29] put forward a multi-objective optimization model employing the normalized normal constraint method (NNCM) for Pareto optimization. The main objectives were to minimize total transportation cost, transport time, and container use cost, maximize profit for freight forwarders, meet shipper time requirements, and reduce enterprise costs. The exploration of fuzzy multi-criteria decision-making approaches has been notable in recent studies. Liaqait *et al*.[30] presented a multi-criteria decision framework for sustainable supplier selection and order allocation using multi-objective optimization and fuzzy approach. The framework consisted of a multi-objective mixed-integer nonlinear programming mathematical model augmented with fuzzy multi-criteria decision-making techniques and forecast demand. Additionally, an integrated framework of fuzzy risk assessment model (FRAM), data envelopment analysis (DEA), and multiple criteria decision-making (MCDM) approaches for route selection in multimodal transportation networks was proposed.[31] The FRAM phase calculates risks through linguistic variables and triangular fuzzy numbers for risk likelihood and severity scales. The Mamdani fuzzy rule-based inference system converts membership degrees for each term of aggregated likelihood and severity scales into those for each term of the risk magnitude scale. In the DEA phase, precise and crisp risk magnitudes are characterized by a new defuzzifier based on the DEA algorithm. The field has also seen the development of algorithms addressing specific challenges in MRP. Zhu *et al*.[32] proposed algorithms for route decision problems under transport speed uncertainty and varying carbon policies. The authors use the Law of Large Numbers to estimate nonlinear uncertainties, and solve the models using the K-shortest paths algorithm and Non-dominated Sorting Genetic Algorithm (NSGA-II). They use Pareto theory to find the optimally symmetrical compromise between objectives and the influence of transport speed uncertainty and carbon emission policies on path decisions. Furthermore, Shao *et al*.[33] employed a decision process based on NSGA-III and a dominance-based rough set approach to optimize cost and enhance punctuality, reliability, and flexibility in freight routing. In their work, they authors propose a decision process to help shippers participate better in routing decisions by considering their requests on transportation cost, timeliness, reliability, and flexibility.

The literature demonstrates a shift from traditional optimization models to more complex and integrative approaches, incorporating AI and ML techniques, fuzzy logic, and environmental constraints and considerations. Our research contributes to this evolving field by introducing a novel approach that synergizes the MCS with the LSTM networks, offering a unique perspective in addressing the complexities of multimodal transport route prediction.

To the best of our knowledge, this innovative hybrid of the MCS and the LSTM for multimodal transport route prediction presents a novel approach for predicting routes in Multimodal Transport Networks.

Artificial Neural Networks (ANNs) are computing systems inspired by the biological neural networks that constitute animal brains. They are the cornerstone of modern machine learning and form the basis for complex architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory Networks (LSTMs).

ANNs are structured around interconnected nodes or "neurons," arranged in layers: an input layer to receive the data, an output layer to make predictions or decisions, and one or more hidden layers to process the inputs. Each node in a layer is connected with all the nodes in the previous and next layer. Each neuron receives data inputs, processes them using a pre-determined function, and passes the result to the next layer. This function is usually nonlinear and can be as simple as a step function, or more complex such as sigmoid or ReLU (Rectified Linear Unit) functions.[34]

The strength of connections between neurons, termed "weights," play a crucial role in an ANN's operation. During training, the ANN adjusts these weights based on the error of its predictions, a process often referred to as backpropagation. By using a large number of hidden layers and nodes, the ANNs are capable of learning complex patterns and relationships in the data, even when the data is noisy, incomplete, or uncertain. ANNs are versatile and can be used in a wide variety of applications, including image and speech recognition, natural language processing, and medical diagnosis, to name just a few. They can handle both regression and classification tasks, and can be designed to work with a broad range of data types, including numerical, categorical, and text data. ANNs are widely used in diverse fields such as finance, healthcare, and computer vision, enabling tasks like image recognition, speech recognition, and language processing. They form the basis of deep learning, a subset of machine learning, which involves training large neural networks on vast amounts of data.

As mentioned previously, Recurrent Neural Networks (RNNs) are a special type of Artificial Neural Networks (ANNs) designed to recognize patterns in sequences of data, such as text, genomes, handwriting, or time series data. These networks are distinguished from other types of ANNs by their *memory* — the ability to take information from previous steps into account in the current step, which is particularly useful when dealing with sequence data. RNNs are structured similarly to traditional ANNs, with layers of nodes, but they also have loops in the network of nodes, allowing information to be passed from one step in the sequence to the next. This gives them a kind of memory about what has been calculated previously. The output of a certain layer may become input for the same layer in the next time-step, creating a 'recurrence' or loop in the network hence the name. See Fig. 1 which shows typical ANN and RNN designs.

However, RNNs have significant limitations. They suffer from what is known as the 'vanishing gradient' problem, where the contribution of information decays geometrically over time, making it difficult for the network to learn and maintain information over long sequences.[34] Essentially, the network *forgets* information from earlier in the sequence which can make them ineffective when handling long sequences and when trying to learn long-term dependencies. Another challenge is the 'exploding gradient' problem where the value of a gradient can become extremely large, causing drastic changes in network weights during learning. This can cause the learning process to become unstable and the network model may not converge or settle on a good set of weights.

To overcome these issues, variations of RNNs have been developed, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs). These network architectures incorporate mechanisms to allow the network to learn over longer sequences and maintain the information they need, while forgetting the information they don't. They have gating mechanisms to control the flow of information into and out of the memory cells in the network, thereby offering a solution to the vanishing and exploding gradient problem.

Long Short-Term Memory Networks (LSTMs) are a

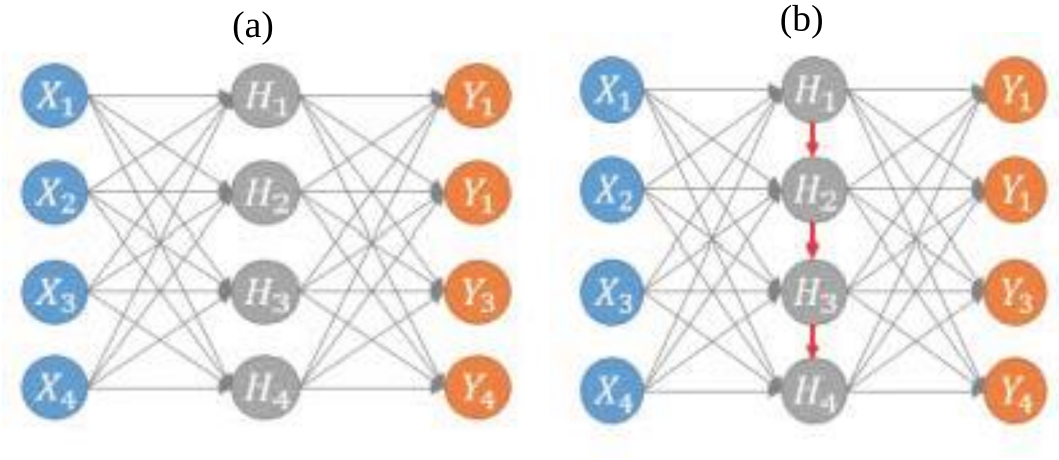

Fig. 1 a) Typical Artificial Neural Network (ANN), b) Recurrent Neural Network (RNN).

special kind of Recurrent Neural Network (RNN) specifically designed to overcome the limitations of standard RNNs, as mentioned above, namely the vanishing and exploding gradient problem, which makes learning from long sequences of data challenging.[35]

The core idea behind LSTMs is the cell state, a horizontal line running through the top of the LSTM's diagram, see Fig. 2. The cell state acts as a "conveyor belt" that carries relevant information across long sequences without the risk of vanishing or exploding gradients. This makes LSTMs highly efficient at learning from data where temporal dependencies are at play.

A critical part of the LSTM's ability to maintain and manipulate its cell state is the use of gates. Gates are neural network layers that have the ability to let information through or not. In LSTMs, there are three types of gates.[36]

- Forget Gate: The forget gate controls the extent to which a value remains in the cell. This gate decides what information should be thrown away or kept. It uses the output from the previous hidden state and the current input to generate a value between 0 and 1. A value close to 0 means 'forget' and a value close to 1 means 'keep'.
- Input Gate: The input gate determines the extent to which a new value flows into the cell. This gate updates the cell state with new information. It has two parts - a sigmoid layer which decides what values to update, and a *tanh* layer which creates a vector of new candidate values that could be added to the state.
- Output Gate: The output gate controls the extent to which the value in the cell is used to compute the output activation of the LSTM unit. This gate decides what the next hidden state should be. It takes the previous hidden state and the current input and decides what parts of the cell state make it to the output.

Each of these gates employs a sigmoid activation function, which outputs a value between 0 and 1, indicating the amount

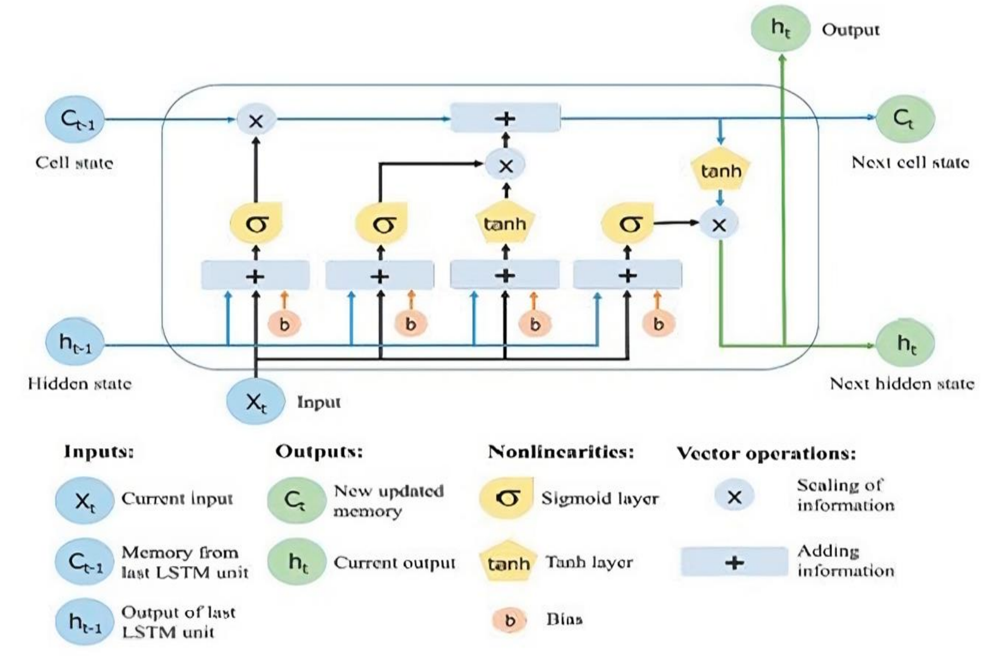

Fig. 2 The structure of the Long Short-Term Memory (LSTM) neural network.

of information to allow through. A value of 0 signifies "let nothing through," while a value of 1 indicates "let everything through." By selectively remembering and forgetting information through these gates, LSTMs can effectively learn long-term dependencies while mitigating issues like vanishing gradients. This ability to capture long-term dependencies is what distinguishes LSTMs and makes them attractive for sequence prediction problems.

The power of LSTMs comes from their general applicability. They have been used for generating human-like text, translating languages, generating captions for images, and much more. They are one of the reasons why AI has made such significant strides over the last decade.

2. Experimental

As noted earlier, one of the challenges of deep learning approach is the lack of vast quantities of data hence we employ the MCS to generate a large number of training samples, providing a rich diversity of possible routes in a multimodal transport network. This obtained data is then used to train the LSTM model to solve path find problems. The reason to utilize the LSTM is that it is well-suited for sequence prediction problems. Described in this section is the proposed approach to produce training data to train the LSTM model and later the LSTM model itself is described, which uses this training data to yield results.

The initial phase of the methodology entails acquiring training data through the application of the MCS, a computational algorithm that relies on repeated random sampling to produce numerical results. This method allows to overcome the limitations often faced with historical data availability. The primary advantage of using simulated data lies in its ability to encompass a broader range of scenarios, including rare but critical cases that are typically underrepresented in historical datasets. This enriched dataset enables the model to potentially achieve a more comprehensive understanding of the complexities inherent in multimodal transport networks, leading to improved prediction performance in both typical and atypical situations.

The MCS generates a diverse set of possible routes between the origin and destination within the multimodal transport network. By generating many samples, the simulation captures a wide range of potential scenarios, including available transport modalities and available vertices to traverse. Each simulation generates a sequence of events, representing a particular route through the network.

The data obtained from the MCS is refined by the applying the Pareto optimality approach.[37] Multi-objective optimization problems involve optimizing multiple objective functions simultaneously. Usually, there is no single optimal solution that optimizes all of the objective functions, and Pareto optimal solutions are used to represent the solution set. A Pareto optimal solution cannot be improved in any of the objectives without degrading at least one of the other objectives. That is, a Pareto optimal solution is one in which improving one objective would lead to the degradation of at least one other. This captures the trade-off nature inherent in multi-objective optimization problems, where it is generally impossible to improve all objectives simultaneously. Pareto optimality provides a way to navigate through the complexity of multi-objective optimization problems by giving decision-makers a set of optimal trade-off solutions to choose from, depending on their specific priorities and needs.

The set of Pareto optimal outcomes is known as the Pareto front or Pareto boundary as shown in Fig. 3.

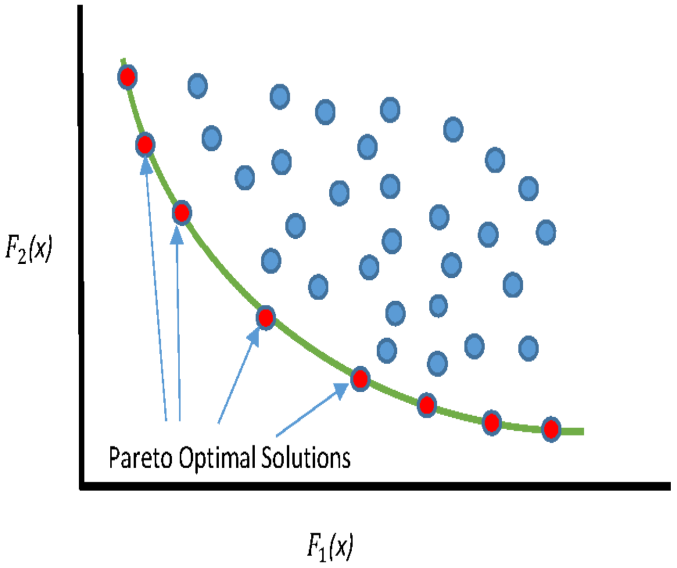

Fig. 3 Pareto front and optimal solutions.

Having obtained the data set as described in Section 2, the next step is to train the LSTM model. Categorical cross-entropy is used as the loss function, given the multi-class nature of the prediction problem. The goal of training is to adjust the model's parameters to minimize the loss function, in other words, to make the model's predictions as close as possible to the true next steps in the route.

For the problem at hand, random sampling approach was used to generate training samples as follows, for each generated sequence. Next is described how the training data is rewritten in a format suitable for LSTM training, then to extract the subsequences for training purpose, noting that for each sequence, the start of each sample would be the start point or origin, A_O since all the sequences would start from A_O and end with destination A_D.

For each sequence of length, r:

1. Generate random sample (sub-sequence) length: Random numbers are generated between 1 and $r-1$ (inclusive) and used as sample length.
2. Form sequence: Starting from the start point A_O, extract a sample using the random number obtained (in step 2 above) from the input sequence data together with the label (output data).
3. Repeat the process until desired amount of training samples as required are obtained.

The LSTM network is at the core of the proposed model. As a recurrent neural network (RNN) variant, it has the unique ability to learn and remember over long sequences, making it particularly suited for route prediction problems,[35] as discussed in detail in Section 1.

The LSTM network consists of an input layer, a certain number of hidden LSTM layers, and an output layer. The LSTM layers are designed to capture the temporal dependencies in the sequences of routes and transportation modes generated by the MCS. In the proposed LSTM architecture, we implement an input layer that matches the shape of the data. Following the input layer, several hidden LSTM layers are stacked. The hidden LSTM layers allow the model to learn higher-order temporal dependencies. To prevent overfitting, dropout layers were incorporated in the model. The output layer uses a softmax activation function to produce a probability distribution over all possible next steps in the route. This distribution represents the model's predictions for the next stage of the route, given the previous stages. An LSTM layer contains memory cells that are designed for retaining information over long periods. In the context of the route prediction model at hand, the LSTM layer operates on sequences of data, where each data point is a feature vector representing an event in a route. It uses its internal memory to keep track of long-term dependencies in these sequences, which helps it to make more accurate predictions about future events. This sequence of LSTM layers, combined with dropout layers for regularization, allows the network to learn complex patterns in the sequence of transportation modes and routes. Finally, an output layer with a softmax activation function is used which provides us a probability distribution over the possible next steps in the route. This architecture allows the LSTM network to effectively learn from the sequences generated by the MCS and make accurate predictions about future routes.

The proposed approach is best summarized in a block diagram as shown in Fig. 4.

As a demonstration of the proposed approach to obtain an optimal (or near) freight route selection in a multimodal network, the proposed method is applied to the problem of finding the optimal route in a multimodal network consisting of 35 vertices and 136 edges, as illustrated in Fig. 5. Data about this case study and detailed results will be shared on request.

The goal is to travel from vertex 1 to vertex 35. Each edge in the graph corresponds to a possible mode of transportation, namely A (road), B (rail), or C (water), which are also represented as modes 1, 2, and 3 respectively.
The approach as described in Section 2 is used with the objective functions as provided below:

$$F(x) = [F_1(x), F_2(x)] \tag{1}$$

where:
Cost function:

$$F_1(x) = \mu \sum_{h=1}^{t-1} \left[\delta\left(A_h^i, A_{h+1}^i\right)\alpha^j + \phi\left(A_h^i, A_{h+1}^i\right)\right] \tag{2}$$

Time function:

$$F_2(x) = \mu \sum_{h=1}^{t-1} \left[\frac{\delta\left(A_h^i, A_{h+1}^i\right)}{\vartheta^j} + \theta\left(A_h^i, A_{h+1}^i\right)\right] \tag{3}$$

and

- $x = (A_O, A_2^m, \dots, A_D^m)$ are the vertices to be traversed to reach destination, $A_D = 35$ from the origin, $A_O = 1$ and via mode: $m \in \{1{:}road, 2{:}rail, 3{:}water\}$;
- $t > 0$ is the number of traversed vertex and $A_O = 1$ and $A_t = 35$;
- $i, j = \{1{:}road, 2{:}rail, 3{:}water\}$ are the modes of transportation used to arrive at respective vertices;
- $\delta(u^i, v^j)$ is the distance of two vertices u and v traversed using mode j;
- speed (unit distance per unit time): $\vartheta^j = \begin{cases} 4.5, & if\ j = 1 \\ 3, & if\ j = 2 \\ 1, & if\ j = 3 \end{cases}$
- cost ($ per unit load): $\vartheta^j = \begin{cases} 6, & if\ j = 1 \\ 3, & if\ j = 2 \\ 1, & if\ j = 3 \end{cases}$
- cost transfer rate ($): $\phi(u^i, v^j) = \begin{cases} 0, & if\ i = j \\ 4, & if\ i = 1, j = 2\ or\ i = 2, j = 1 \\ 7, & if\ i = 1, j = 3\ or\ i = 2, j = 3 \end{cases}$
- time transfer rate (unit time): $\theta(u^i, v^j) = \begin{cases} 0, & if\ i = j \\ 0.0067, & if\ i = 1, j = 2\ or\ i = 2, j = 1 \\ 0.0113, & if\ i = 1, j = 3\ or\ i = 2, j = 3 \end{cases}$

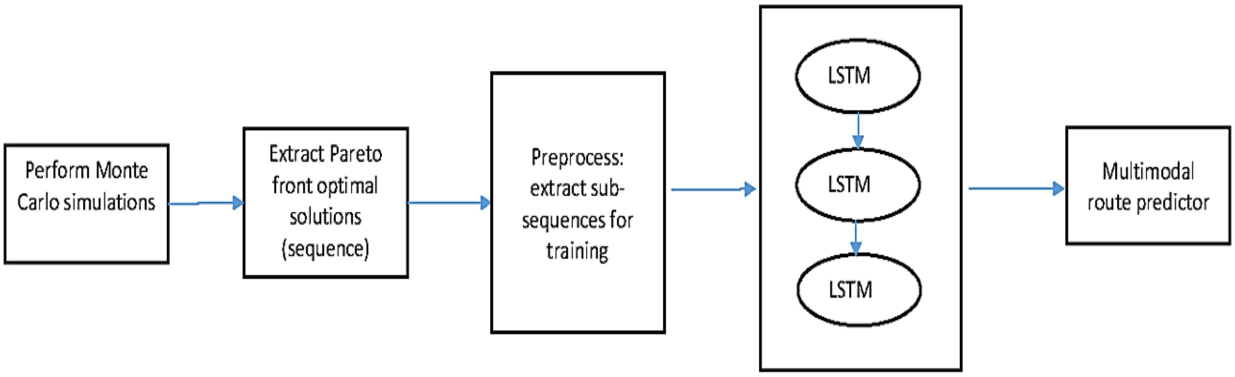

Fig. 4 Block diagram illustrating proposed approach.

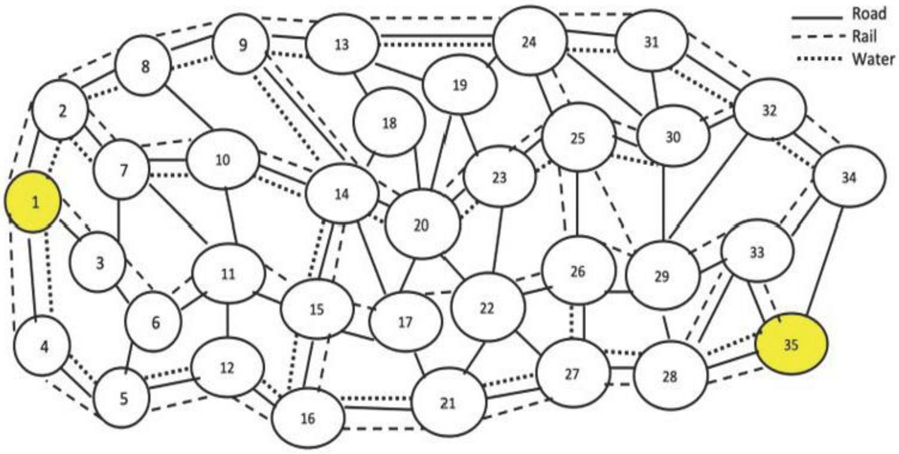

Fig. 5 Example multimodal transport network.

The MCS were executed for a total of 10^6 iterations, generating a comprehensive set of potential route scenarios. It was observed that the MCS were able to capture a wide range of scenarios, including those that were less frequent but held significance in real-world transport contexts. This is especially crucial for training the LSTM model as it provides a broad understanding of the transport network's dynamics.

The obtained dataset consists of sequences of locations that represent different routes in the multimodal transport network. Each sequence corresponds to a complete route from an origin to a destination, with different modes of transportation annotated. For input to the LSTM model, from all the obtained routes obtained, pareto front solutions are extracted as described in Section 2. The pareto optimal solutions with corresponding objective function values (cost/time) are depicted in Fig. 6. These pareto front solutions are then divided into subsequences as described in Section 2 for training the LSTM.

Table 1 provides examples of subsequences extracted for LSTM training. The sub-sequences presented below which are extracted from the pareto front solution:

Sequence:

['1','barg','4','barg','5','barg','12','road','16','road','21','road','17','rail','22','rail','26','rail','29','rail','33','rail','35]

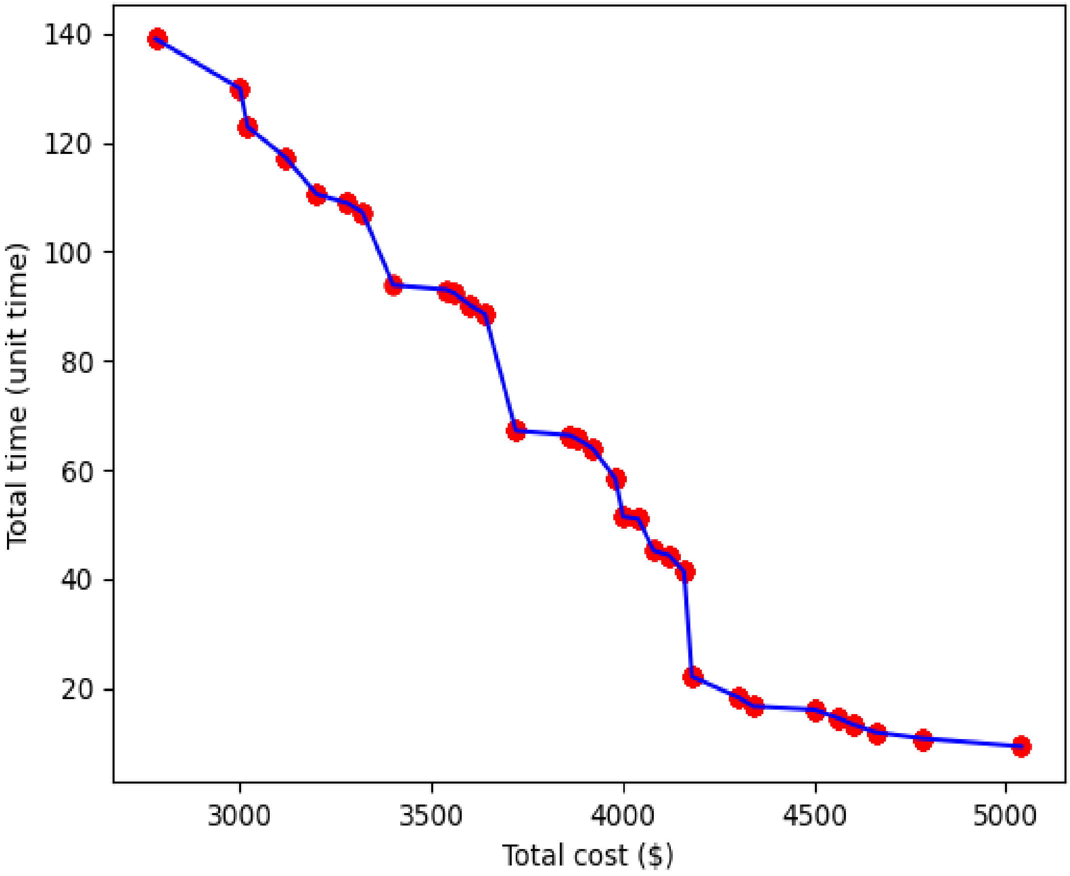

Fig. 6 Obtained pareto front solutions.

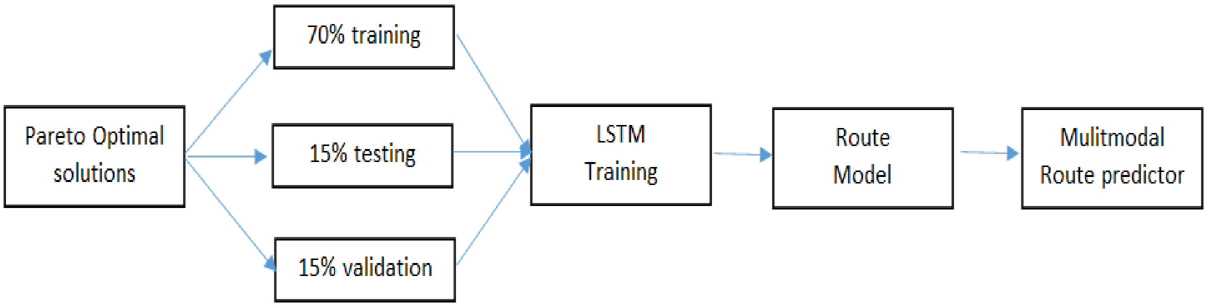

Fig. 7 LSTM training model and output.

Table 1. A sample of sub-sequences.

Random length	Sub-sequence	Label
3	['1', 'barg', '4']	['barg']
7	['1', 'barg', '4', 'barg', '5', 'barg', '12']	['road']
5	['1', 'barg', '4', 'barg', '5']	['barg']

3. Results and discussion

The LSTM network was trained using the samples generated by the MCS as outlined in Fig. 7.

The dataset is initially divided into training and testing sets. 70% of the data is used for training and the remaining 30% is further split into two equal parts of 15% each for validation and testing. The *early stop* callback was used to monitor the validation loss during training and stop the training process if there is no improvement in validation loss for 20 consecutive epochs. The training process ended after 125 Epochs.

Training Performance: The model achieved a training accuracy of 85.97% and a validation accuracy of 86.24% after 125 epochs. This indicates a robust learning from the diverse set of training scenarios. The validation accuracy being is higher than the training accuracy is a good sign that the model is generalizing well and not just memorizing the training data. The model's loss converged to a value of 0.5103 after 125 epochs, confirming the efficiency and effectiveness of the network in learning the route patterns and dependencies. The trained LSTM model's capability to predict full route sequences from given starting sub-sequences was evaluated. This process helps in assessing both the model's generalization capabilities and its adherence to known optimal solutions when they exist.

The output presented in Table 2 matched perfectly with the training data and represents one of the Pareto optimal solutions.

Table 2. Starting with a Single Vertex (Test 1).

Input Sub-sequence	Predicted Route from LSTM Model
['1']	['1', 'rail', '4', 'rail', '5', 'rail', '12', 'rail', '16', 'rail', '21', 'barg', '27', 'barg', '28', 'barg', '35']

This prediction as per Table 3 is also one of the Pareto optimal solutions confirming the model's ability to continue a known optimal route when given part of it as a start sequence.

Table 3. Starting with a Partial Route (Test 2).

Input Sub-sequence	Predicted Route from LSTM Model
['1', 'road', '4', 'road', '5']	['1', 'road', '4', 'road', '5', 'road', '12', 'road', '16', 'road', '21', 'road', '27', 'road', '28', 'road', '35']

While this predicted sequence, as shown in Table 4 is not part of the training data, it remains a feasible route. Its structure is reminiscent of optimal routes, showcasing the model's ability to adapt and generate practical routes even when faced with unfamiliar input sequences. The representation of this route is illustrated in Fig. 8. As is evident here that the resultant route is not from the training set but its objective function values are sitting closely to that of the training pareto set.

Table 4. Testing Model Generalization with unknown starting sequence (Test 3).

Input Sub-sequence	Predicted Route from LSTM Model
['1', 'barg', '4', 'road', '5']	['1', 'barg', '4', 'road', '5', 'road', '12', 'road', '16', 'rail', '21', 'barg', '27', 'barg', '28', 'barg', '35']

This paper introduces an LSTM-based approach for route selection in multimodal transport networks, emphasizing the unique application of the MCS-generated data for training. The pivotal aspect of this approach lies in the Value of Simulated Data: by leveraging simulations, the proposed model gains exposure to a wide spectrum of scenarios. This includes not only common occurrences in transport networks but also rare yet significant situations that are often not captured in historical datasets. Such enriched training enhances the model's ability to handle complex and unforeseen routing scenarios, surpassing the capabilities of models trained solely with historical data.

Our tests demonstrate that the LSTM model, trained on data derived from the MCS, can reliably predict viable routes

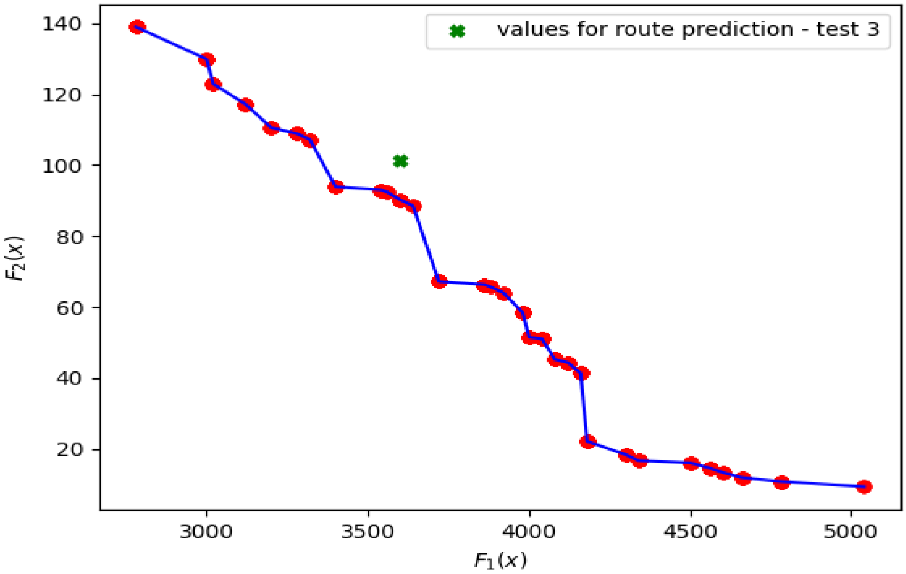

Fig. 8 Test 3 - predicted route values.

in a multimodal transport network. Notably, the model adheres to know optimal solutions from its training set and shows remarkable adaptability in generating plausible predictions for new, unseen sequences. This finding is significant as it indicates the model's genuine learning ability, rather than mere memorization of training data, thus underscoring the effectiveness of the proposed methodological approach.

One of the limitations of this model is that the user, upon giving a start sequence is unable to identify that the required prediction needs to avoid certain vertices for some reason. We intend to address this in our future research.

4. Conclusion

The results obtained from this study affirm the potential of the proposed methodology in the area of route prediction for multimodal transport networks. The MCS-LSTM-based hybrid approach paves the way for more dynamic and accurate predictions in complex transportation environments. The model's success in adhering to optimal solutions and its adaptability in handling new route sequences showcase its potential as a robust tool in transport network analysis.

Looking ahead, there are several avenues for further enhancing this model's capabilities. Future research directions include:

- Integrating Historical Data: Combining the rich scenarios from simulations with real-world historical data could further enhance the model's accuracy and generalizability.
- Comparing performance and accuracy with other machine learning techniques.
- Expanding to Diverse Transport Environments: Applying the model to varied geographical and infrastructural contexts to test its versatility and adaptability, such as introducing obstacles and using sensors in the pathway.
- Real-Time Prediction Capabilities: Developing the model for real-time route prediction, which can significantly aid in dynamic transport management and decision-making process.

By pursuing these directions, the authors aim to contribute more comprehensively to the field of intelligent transport systems, addressing the evolving challenges and demands of modern transportation networks.

Conflict of Interest

There is no conflict of interest.

Supporting Information

Not applicable.

References

[1] J. Li, D. Fu, Q. Yuan, H. Zhang, K. Chen, S. Yang, F. Yang, A traffic prediction enabled double rewarded value iteration network for route planning, *IEEE Transactions on Vehicular Technology*, 2019, **68**, 4170-4181, doi: 10.1109/TVT.2019.2893173.

[2] Y. Issaoui, A. Khiat, K. Haricha, A. Bahnasse, H. Ouajji, An advanced system to enhance and optimize delivery operations in a smart logistics environment, *IEEE Access*, 2022, **10**, 6175-6193, doi: 10.1109/ACCESS.2022.3141311.

[3] H. Wang, B. Lu, J. Li, T. Liu, Y. Xing, C. Lv, D. Cao, J. Li, J. Zhang, E. Hashemi, Risk assessment and mitigation in local path planning for autonomous vehicles with LSTM based predictive model, *IEEE Transactions on Automation Science and Engineering*, 2022, **19**, 2738-2749, doi: 10.1109/TASE.2021.3075773.

[4] M. SteadieSeifi, N. P. Dellaert, W. Nuijten, T. Van Woensel, R. Raoufi, Multimodal freight transportation planning: a literature review, *European Journal of Operational Research*, 2014, **233**, 1-15, doi: 10.1016/j.ejor.2013.06.055.

[5] X. Wan, H. Liu, H. Xu, X. Zhang, Network traffic

prediction based on LSTM and transfer learning, *IEEE Access*, 2022, **10**, 86181-86190, doi: 10.1109/ACCESS.2022.3199372.

[6] B. Bakker, Reinforcement learning with long short-term memory, Advances in Neural Information Processing Systems, 2001.

[7] P. Shi, A. Gangopadhyay, C. Owens, B. Blunt, C. Grogan, A hybrid model using LSTM and decision tree for mortality prediction and its application in provider performance evaluation, 2019 IEEE International Conference on Big Data (Big Data). Los Angeles, CA, USA. IEEE, 2019.

[8] K. T. Chui, B. B. Gupta, P. Vasant, A genetic algorithm optimized RNN-LSTM model for remaining useful life prediction of turbofan engine, *Electronics*, 2021, **10**, 285, doi: 10.3390/electronics10030285.

[9] S. Prakash, D. Mitchell, Probabilistic benefit cost ratio – A case study. ATRF 2015 - Australasian Transport Research Forum 2015, Proceedings. 2015.

[10] S. Prakash, Alternative approach to estimating crash costs for cost-benefit analysis using Monte Carlo simulation. ATRF 2018 - Australasian Transport Research Forum 2018, Proceedings. 2018.

[11] S. Prakash, A. Jokhan, Monte Carlo for selecting risk response strategies. ATRF 2017 - Australasian Transport Research Forum 2017, Proceedings. 2017.

[12] S. Prakash, Emergency relief goods transportation strategies – A Monte Carlo simulation approach. ATRF 2019 - Australasian Transport Research Forum, ATRF 2019 - Proceedings. 2019.

[13] M. Balmer, K. Meister, M. Rieser, K. Nagel, K. Axhausen, Agent-based simulation of travel demand: structure and computational performance of MATSim-T, 2008

[14] S. Prakash, D. Mitchell, Estimating freight movements using Dijkstra's algorithm. ATRF 2018 - Australasian Transport Research Forum 2018, Proceedings. 2018.

[15] S. Prakash, A. Jokhan, An optimal cane delivery scheduling using the Monte Carlo method. ATRF 2016 - Australasian Transport Research Forum 2016, Proceedings. 2016.

[16] S. Prakash, B. Sharma, An optimized hybrid approach for path planning: a combination of Lyapunov functions and high-level planning algorithms, S. Das, S. Saha, C. A. Coello Coello, J. C. Bansal, International Conference on Advances in Data-driven Computing and Intelligent Systems. Singapore: Springer, 2024: 425-436, doi: 10.1007/978-981-99-9524-0_32.

[17] M. J. Maher, P. C. Hughes, A probit-based stochastic user equilibrium assignment model, *Transportation Research Part B: Methodological*, 1997, **31**, 341-355, doi: 10.1016/s0191-2615(96)00028-8.

[18] T.-S. Chang, Best routes selection in international intermodal networks, *Computers & Operations Research*, 2008, **35**, 2877-2891, doi: 10.1016/j.cor.2006.12.025.

[19] G. Xiong, Y. Wang, Best routes selection in multimodal networks using multi-objective genetic algorithm, *Journal of Combinatorial Optimization*, 2014, **28**, 65-673, doi: 10.1007/s10878-012-9574-8.

[20] Y. Tao, E. P. Chew, L. H. Lee, Y. Shi, A column generation approach for the route planning problem in fourth party logistics, *Journal of the Operational Research Society*, 2017, **68**, 165-181, doi: 10.1057/s41274-016-0024-3.

[21] M. Darayi, K. Barker, C. D. Nicholson, A multi-industry economic impact perspective on adaptive capacity planning in a freight transportation network, *International Journal of Production Economics*, 2019, **208**, 356-368, doi: 10.1016/j.ijpe.2018.12.008.

[22] B. Gendron, L. Gouveia, Reformulations by discretization for piecewise linear integer multicommodity network flow problems, *Transportation Science*, 2017, **51**, 629-649, doi: 10.1287/trsc.2015.0634.

[23] S. Moradi, A. Raith, M. Ehrgott, A bi-objective column generation algorithm for the multi-commodity minimum cost flow problem, *European Journal of Operational Research*, 2015, **244**, 369-378, doi: 10.1016/j.ejor.2015.01.021.

[24] A. Baykasoğlu, K. Subulan, A multi-objective sustainable load planning model for intermodal transportation networks with a real-life application, *Transportation Research Part E: Logistics and Transportation Review*, 2016, **95**, 207-247, doi: 10.1016/j.tre.2016.09.011.

[25] M. Zhang, A. J. Pel, Synchromodal hinterland freight transport: model study for the port of Rotterdam, *Journal of Transport Geography*, 2016, **52**, 1-10, doi: 10.1016/j.jtrangeo.2016.02.007.

[26] B. Behdani, Y. Fan, B. Wiegmans, R. Zuidwijk, Multimodal schedule design for synchromodal freight transport systems, *European Journal of Transport and Infrastructure Research*, 2016, **16**, 424–444, doi: 10.18757/ejtir.2016.16.3.3151.

[27] S. Prakash, R. G. Thompson, C. Prakash, Route selection in multimodal transport networks incorporating disruption. Uddin MS, Bansal JC, International Joint Conference on Advances in Computational Intelligence. Singapore: Springer, 2024.

[28] E. Demir, M. Hrušovský, W. Jammernegg, T. Van Woensel, Green intermodal freight transportation: bi-objective modelling and analysis, *International Journal of Production Research*, 2019, **57**, 6162-6180, doi: 10.1080/00207543.2019.1620363.

[29] D. Chen, Y. Zhang, L. Gao, R. G. Thompson, Optimizing multimodal transportation routes considering container use, *Sustainability*, 2019, **11**, 5320, doi: 10.3390/su11195320.

[30] R. A. Liaqait, S. S. Warsi, M. H. Agha, T. Zahid, T. Becker, A multi-criteria decision framework for sustainable supplier selection and order allocation using multi-objective optimization and fuzzy approach, *Engineering Optimization*, 2022, **54**, 928-948, doi: 10.1080/0305215x.2021.1901898.

[31] N. Koohathongsumrit, W. Meethom, An integrated approach of fuzzy risk assessment model and data envelopment analysis for route selection in multimodal transportation networks, *Expert Systems with Applications*, 2021, **171**, 114342, doi: 10.1016/j.eswa.2020.114342.

[32] C. Zhu, X. Zhu, Multi-objective path-decision model of

multimodal transport considering uncertain conditions and carbon emission policies, *Symmetry*, 2022, **14**, 221, doi: 10.3390/sym14020221.

[33] C. Shao, H. Wang, M. Yu, Multi-objective optimization of customer-centered intermodal freight routing problem based on the combination of DRSA and NSGA-III, *Sustainability*, 2022, **14**, 2985, doi: 10.3390/su14052985.

[34] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[35] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Computation*, 1997, **9**, 1735-1780, doi: 10.1162/neco.1997.9.8.1735.

[36] Le, Ho, Lee, Jung, Application of long short-term memory (LSTM) neural network for flood forecasting, *Water*, 2019, **11**, 1387, doi: 10.3390/w11071387.

[37] R. T. Marler, J. S. Arora, Survey of multi-objective optimization methods for engineering, *Structural and Multidisciplinary Optimization*, 2004, **26**, 369-395, doi: 10.1007/s00158-003-0368-6.