

Final Report

Estimation of Recurring Traffic Congestion Using Bayesian Network

Shannon Feng

Course 7525: Cognitive and object-oriented modeling under uncertainties as aspects of artificial intelligence in practical applications

Course 7540: Artificial intelligence in applications. Modeling, Machine Learning and Data Classifier Performance

Abstract

This study employs Bayesian networks to estimate recurring traffic congestion and compares two time slots: during the COVID-19 pandemic and recent post-COVID data. Surprisingly, there are minimal differences in traffic patterns between the two periods, with both achieving 100% accuracy in predicting traffic congestion on the testing dataset.

Using a comprehensive dataset of traffic flow information and historical congestion data, our Bayesian network model captures the complex relationships among variables such as segment, direction, day, time, speed, vehicle count, SPI, V/C, and finally calculate the congestion state.

The findings have important implications for urban planning and traffic management. Despite the pandemic's impact on travel behavior and overall traffic volume, recurring traffic congestion patterns remain consistent. This suggests that existing traffic management strategies can effectively address congestion even in the post-COVID era.

The high accuracy of the Bayesian network model reinforces its potential as a valuable tool for capturing the interactions that influence traffic congestion. Further research is needed to validate the model across different cities and incorporate real-time data and emerging transportation technologies.

In conclusion, this report highlights the effectiveness of a Bayesian network model in estimating recurring traffic congestion. The comparison of pre and post-COVID data reveals minimal differences, underscoring the importance of considering long-term traffic behavior. The model's 100% accuracy further establishes its value in traffic management and urban planning contexts.

1. Introduction

Traffic congestion has become a pervasive issue in urban areas around the world. The rise in population, rapid urbanization, and increased vehicle ownership have all contributed to the growing problem of traffic congestion. The negative impact of congestion on our daily lives, economy, and environment cannot be ignored. As a result, accurately estimating traffic congestion has emerged as a crucial endeavor in creating smarter and safer cities. It is vital for efficient traffic management, enhanced public safety, improved productivity, and sustainable urban living. It allows cities to identify congestion hotspots, optimize traffic flow, reduce delays, prevent accidents, and create a healthier and more sustainable environment.

Accurately estimating traffic congestion is a crucial component of managing urban mobility and ensuring public safety. By estimating congestion levels, cities can optimize traffic flow, additionally, congestion estimates provide valuable data for informed decision-making and strategic urban planning. As cities continue to grow, addressing traffic congestion through accurate estimation becomes increasingly vital to create efficient, safe, and sustainable urban environments for residents and visitors alike.

Traffic congestion is a complex and dynamic phenomenon influenced by various factors such as road conditions, weather, traffic volume, and individual driver behavior. Accurately estimating congestion levels and predicting traffic patterns is a challenging task that requires sophisticated modeling techniques. One such approach that has gained significant traction in recent years is the use of Bayesian networks. Bayesian networks offer a powerful and flexible framework for traffic congestion estimation, providing several advantages over traditional methods.

Bayesian networks can incorporate uncertainty and probabilistic reasoning, accounting for incomplete and uncertain data in traffic congestion estimation; capture the intricate relationships between factors influencing congestion, such as traffic volume, road conditions, and time of day, leading to more accurate predictions; learn from historical data, adapting to changing traffic patterns and conditions over time; seamlessly integrate real-time data on traffic conditions, providing up-to-date congestion estimates for proactive traffic management; enable scenario analysis and "what-if" simulations, assisting decision-makers in evaluating different interventions or policy changes to mitigate congestion effectively. By utilizing Bayesian networks, cities can enhance their traffic management strategies, optimize infrastructure planning, and improve overall mobility for their citizens.

Literature review

This article are mostly refer from *A probabilistic estimation of traffic congestion using Bayesian network (Afrin & Yodo, 2021)*. In this paper, a Bayesian Network (BN) based probabilistic congestion estimation approach is proposed. The proposed BN-based approach considers both speed and volume- related measures and provides a probabilistic estimation of the probable congestion state. For recurring and nonrecurring congestion, two different BN models were developed and implemented in realtime datasets. The case study results showed that the proposed BN models could quantify the probable congestion level in terms of a probability for each state in a variable, at the presence of different combinations of prior variables' state. Further, the proposed BN based approach can be employed in the decision-making process that involves the probabilistic estimation of traffic congestion with a vision of the realtime circumstances. The flow chart of the proposed framework for the BN-based congestion estimation approach is illustrated in Fig. 1. (Afrin & Yodo, 2021)

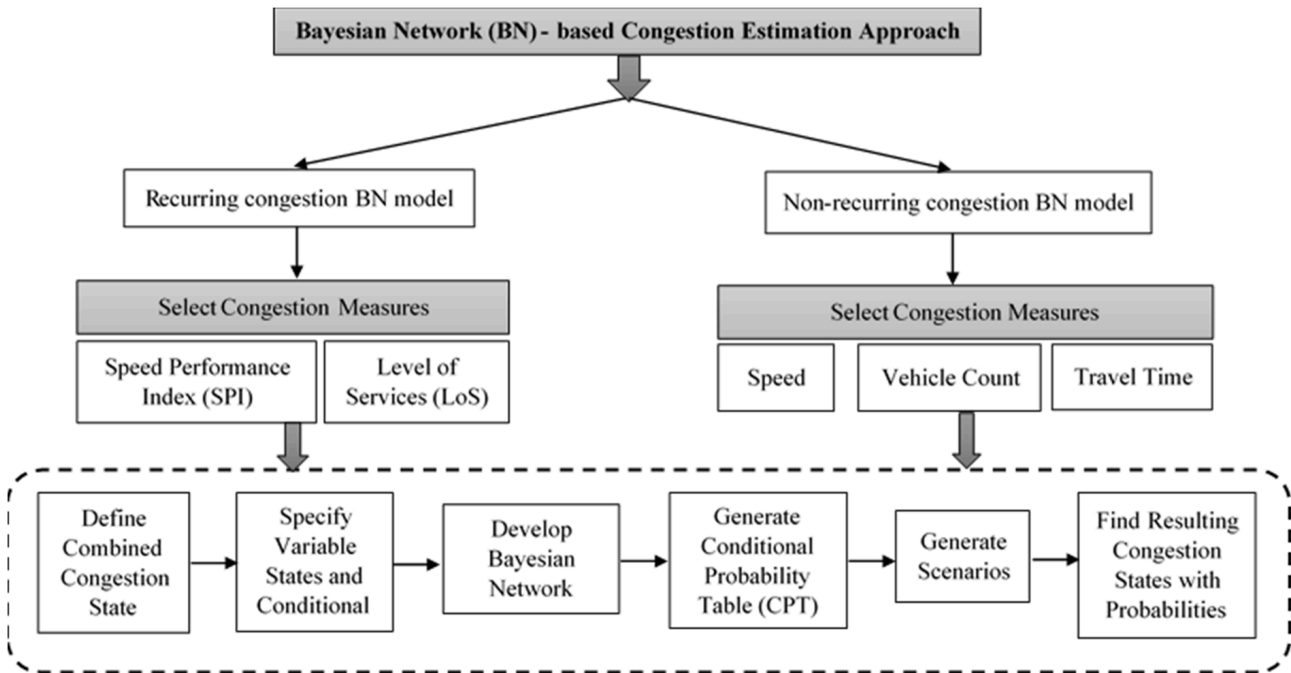


Fig. 1. Bayesian Network (BN) – based congestion estimation framework. (Afrin & Yodo, 2021)

This paper proposed 2 models, including recurring congestion BN model for recurring state, and non-recurring congestion BN model for an unusual kind of congestion that generally occurs due to unpredictable events. However, this report only conducted and analyzed the recent data on recurring congestion state.

2. Bayesian network

Bayesian networks have emerged as a valuable tool in the field of probabilistic reasoning and decision-making. Developed from the principles of Bayesian statistics, Bayesian networks provide a graphical representation of probabilistic relationships among variables.

A Bayesian network (BN) is a directed acyclic graph (DAG) with a collection of nodes and arcs that represents probabilistic relationships between different variables. These variables and the dependency relationships between them are represented by the nodes and the arcs, respectively. The directions of arcs connecting pairs of nodes represent the type of dependencies between the variables. Consider a graph $G=(V,E)$ as a BN with a set of nodes (variables) $V=\{X_1, X_2, \dots, X_n\}$ and a set of arcs (links) E as shown in Fig. 2. A link is directed from node X_i to X_j , indicates the states of X_j are dependent on the states of X_i . In this case, X_i is the parent of X_j , and X_j is the child of X_i . The set of all parents of X_i could be defined as $\text{par}(X_i)$. Nodes that do not have any child are called the leaf node and do not have any parent are called root vertex. The BN is also known as a belief network. (Afrin & Yodo, 2021)

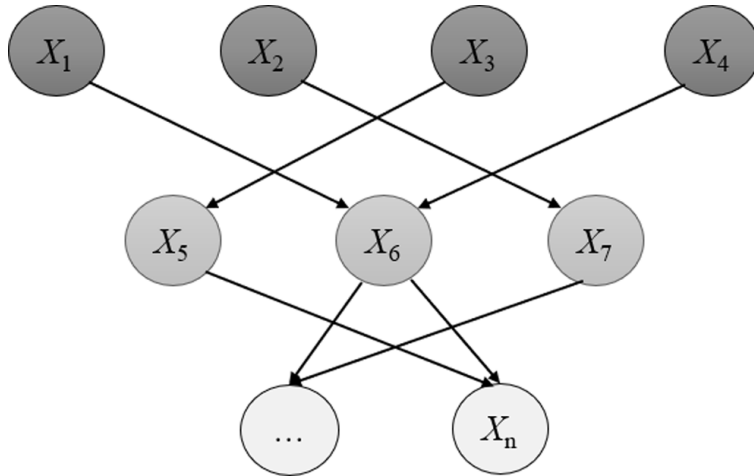


Fig. 2. A general structure of a Bayesian Network with n nodes. (Afrin & Yodo, 2021)

As BN is capable of modeling relationships between variables in complex systems efficiently, it is being used in many real-world applications, including forecasting, predictive analysis, risk management, and many other applications. This paper is mainly focused on employing BN models in probabilistic analysis in transportation and logistics research areas.

3. Estimation recurring congestion using Bayesian network-based model

3.1 Recurring congestion BN network

In this report, I applied the model from the same paper, using Bayesian network to develop estimating recurring congestion. So except for the parameters from the data, also applied speed performance index (SPI), and level of services (Los) in the BN model. These two measures consider the most basic standard parameters, which are speed and volume, that can represent the traffic condition in the most significant way. In general, for road traffic to be considered in the congestion state, the average traffic speed is significantly lower, and volume is higher compared to the normal traffic condition. The justification behind employing both SPI and LoS rather than just either one of them is that the combination of these measures is a better representation of traffic state since the root cause of congestion involves both speed and vehicle volume. Both speed and volume are crucial measures to determine the congestion state as using only one would not be sufficient for this purpose. Moreover, these measures have been used in different works of literature as well as in annual traffic reports. On the basis of adhering with the measurements employed in formal traffic reports, these two measures were selected and deemed to be sufficient to define various traffic states. But there is still no standard for determining the scope of parameter values for different traffic levels. (Afrin & Yodo, 2021)

Speed Performance Index (SPI): SPI is the ratio between average vehicle speed and the maximum permissible speed, as shown in Eq. (1), where SPI denotes the speed performance index, V_{avg} indicates the average travel speed, and V_{max} denotes the maximum permissible road speed. The value of SPI ranges from 0 to 100. The traffic state level can be classified with three threshold values (25, 50, and 75) and four levels, as shown in Table 1. (Afrin & Yodo, 2021)

$$SPI = (V_{avg}/V_{max}) \times 100 \quad (1)$$

Table 1. SPI with traffic state.

SPI	Traffic state level	Description of traffic state
[0, 25]	Heavy	Low speed, poor state
(25, 50]	Mild	Lower speed, weak state
(50, 75]	Smooth	Higher speed, better state
(75, 100]	Very smooth	High speed, good state

Level of Services (LoS): The LoS approach was introduced in the Highway Capacity Manual (HCM). LoS is a popular method in determining traffic states due to its simplicity. The volume-to-capacity ratio (V/C) is one of the methods used to estimate the LoS of a roadway. The scale intervals of the volume-to-capacity ratio (V/C) are shown in Table 2. The V/C ratio can be calculated by Eq (2), where N_v is the spatial mean volume, and N_{max} denotes the maximum number of vehicles that a segment can contain as the capacity. It can be further quantified as Eq (3), where L_s is the spatial segment length, L_v is the average vehicle length occupancy, and N_l is the number of lanes. L_v includes vehicle length and safety distance. In general, it is assumed that vehicle length is about 14 ft. (approx. 4.27 m), and safety distance is about 15 ft. (approx. 4.57 m) (Afrin & Yodo, 2021)

$$V/C = N_v/N_{max} \quad (2)$$

$$N_{max} = (L_s/L_v) \times N_l \quad (3)$$

Table 2. LoS based on the corresponding V/C ratio and operating conditions.

LoS	Traffic state and condition	V/C ratio
A	Free flow	0–0.60
B	Stable flow with unaffected speed	0.61–0.70
C	Stable flow but speed is affected	0.71–0.80
D	High-density but the stable flow	0.81–0.90
E	Traffic volume near or at capacity level with low speed	0.91–1.00
F	Breakdown flow	>1.00

Define Combined Congestion State: To determine the recurring congestion state, V/C and SPI levels are combined, as shown in Table 3. For each level of a measure (V/C and SPI), a value is assigned in ascending order. The values assigned for six degrees of services (LoS): A, B, C, D, E, F based on the V/C ratio are 1, 2, 3, 4, 5, and 6, respectively. A value 1 indicates free flow, and 6 shows breakdown flow. Similarly, the values assigned to the SPI levels: very smooth, smooth, mild, and heavy are 1, 2, 3, and 4, respectively. To find the final congestion state measure, these values are added. The new range for congestion states is defined as smooth (1–4), mild (5–7), and heavy (8–10). (Afrin & Yodo, 2021)

Table 3. The new range for recurring congestion states.

LoS (V/C ratio)	+	SPI	=	Congestion state	
Level	Value	Level	Value	Level	Value
A (Free flow)	1	Very smooth	1	Smooth	1-4
B	2	Smooth	2	Mild	5-7
C	3	Mild	3	Heavy	8-10
D	4	Heavy	4		
E	5				
F (Breakdown)	6				

3.1 Data analysis

The proposed BN model was implemented with an open-source dataset from the Chicago Traffic Tracker (<https://catalog.data.gov/dataset/chicago-traffic-tracker-historical-congestion-estimates-by-segment-2018-current>), and the parameters descriptions could be refer to the webpage (<https://data.cityofchicago.org/Transportation/Chicago-Traffic-Tracker-Historical-Congestion-Esti/sxs8-h27x>). This dataset contains the historical estimation of congestion for over 1000 traffic segments, starting from 17:01 28th February 2018. The dataset I download are contain with the data from 28th February 2018 to 5th June 2023, totally 409884 data. In this report, four segments: G1 (Indiana to I-94 Expressway), G2 (I-94 Expressway to Indiana), G3 (Lawrence to Kennedy Expressway), G4 (Kennedy Expressway to Lawrence), and four directions: EB (Eastbound), WB (Westbound), NB (Northbound), and SB (Southbound) were considered; Day was separated into Weekday (Monday to Friday), and Weekend (Saturday and Sunday); Time was separated into AM Peak (6 - 9 am weekdays), PM Peak (4 -7 pm weekdays), and Off-peak; Speed was separated into Low (0 - 15 milds), Medium (16 - 25 miles), and High (> 26 miles), and in the dataset, a value of -1 in Speed means no estimation is available, here just deleted all that data; Vehicle counts (Bus count + Message count) was separated into Low (0 - 10), Medium (11 - 20), and High (> 21), here is only detects the car that equipped with GPS, and there are 93% of cars have it for average by some surveys, so I calibrated the car values by divided 0.93; SPI and V/C were categories by Table 1 and Table 2; finally our target variable congestion state was categories with Smooth, Mild, and Heavy, please refer to Table 4.

Table 4. Variables and state definitions for the BN model for recurring congestion.

Level 1: Attribute		Level 2: Parameters		Level 3: Congestion level	
Variables	State	Variables	State	Variables	State
Segment	G1: Indiana to I-94 Expressway	Speed	Low	SPI	Heavy
	G2: I-94 Expressway to Indiana		Medium		Mild
	G3: Lawrence to Kennedy Expressway		High		Smooth
	G4: Kennedy Expressway to Lawrence		Low		Very smooth
Direction	EB: Eastbound	Vehicle count	Medium	V/C	A - F
	WB: Westbound		High		Smooth
	NB: Northbound				Mild
Day	Weekday			Congestion state (Target variable)	Heavy
	Weekend				
Time	AM Peak (6 – 9 am weekdays)				
	PM Peak (4 – 7 pm weekdays)				
	Off-Peak				

The variables selected from this dataset for analysis are time, segment, direction, speed, bus count, message count, segment length, from street, and end street, the details . By using these variables, SPI and V/C were calculated using Eqs. (1), (2), and (3). For Nv (car volume) I assumed vehicle length 4.27 m to add on safety distance 4.57 m is 8.84; for bus I assumed the length is 18.288 m to add on safety distance 4.57 m is 22.858 m. Vmax for Chicago streets is 30 miles per hour. For parameter learning and probabilistic inference, Hugin software was used.

The development of BN model for measuring recurring congestion involved nine variables selected as the nodes. These variables are segment, direction, day, time, speed, vehicle count, SPI, V/C, and congestion state. The states of these variables are described in Table 4. All these variables are divided into three levels considering the variable relationships. Level 1: segment, direction, day, time, Level 2: speed, vehicle count, and Level 3: SPI, V/C, congestion state. The congestion state is the target variable.

Once the states of the variable are specified, the next step is to determine the conditional relationships between variables. In this paper, the structure of the BN model was manually defined based on experts' domain knowledge. The proposed BN model for recurring congestion is shown in

Fig. 3. The variables in Level 2: speed, vehicle count, and Level 3: SPI, V/C, congestion state is conditionally dependent on the variables in Level 1: segment, direction, day, time. There is also a conditional dependency between Level 2 and Level 3 variables. Besides, the target variable congestion state depends on the V/C and SPI states. The connections (links/lines) between the nodes represent the conditional dependency between the variables. (Afrin & Yodo, 2021)

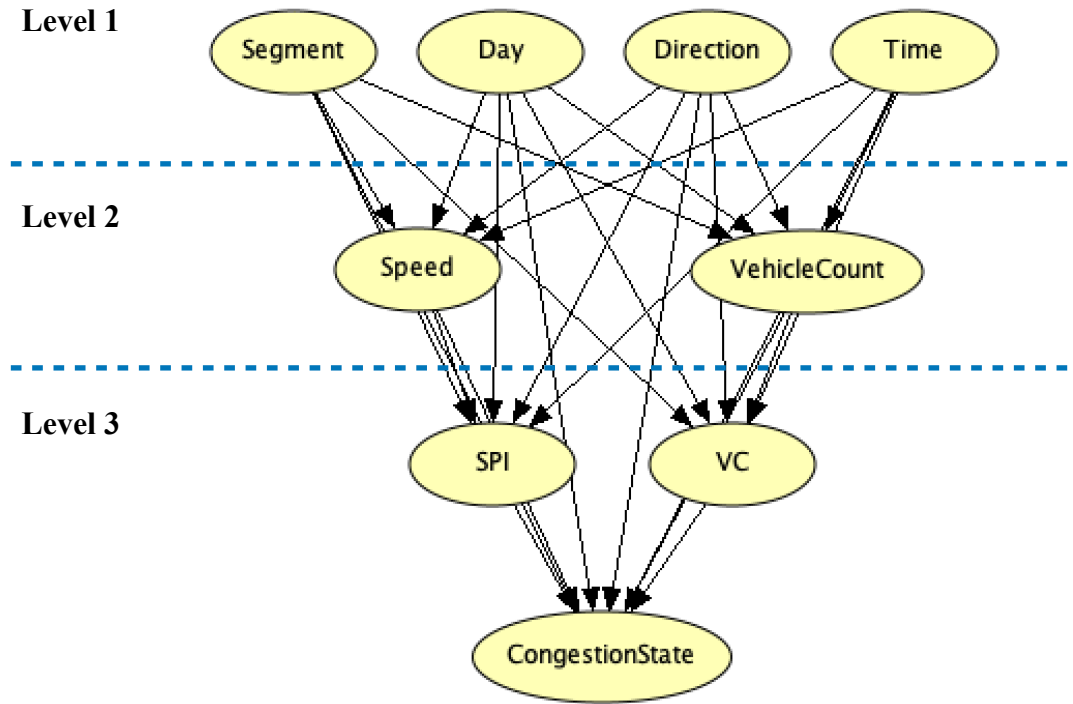


Fig. 3. Graph representation of the proposed BN model for recurring congestion.

For this report, I would like to know is there any recurring congestion differences between Covid-19 period and recent data. Therefore, I divided the time slot into 2 different slot, one is during 2020, the other is 2023. The data set are all contained training set and testing set for 3 months data and 1 month data respectively.

3.2.1 During Covid-19 period

Each of the variable values was taken for every 10 minutes for seven consecutive days from June 1, 2020, to August 30, 2020 as a training set, totally 19232 data; August 31, 2020 to September 27, 2020 as testing set, totally 6454 data.

The proposed Bayesian network was implemented, and the probability distribution was estimated based on the training dataset during Covid-19 period. The estimated probability results are shown in

Fig. 4. The probability distribution of segment, direction, day, and time of the current dataset provides the actual state of the given dataset.

In Level 1, the probability distribution of a given variables x was calculated as, $P(A = x) = (\text{No. of events } A = x) / \text{No. of all possible events from the dataset employed}$. Without any observation made and based on the data employed, in Level 2, the probabilities of the average speed are low, medium, and high are 25.05%, 34.00%, and 40.95%, respectively. The states of vehicle count being low, medium, and high are 50.01%, 26.17%, and 23.81%, respectively.

In Level 3, the V/C levels, A (free flow), B, C, D, E, F (breakdown) indicates traffic states from the free flow to the breakdown state occurred with the probability of 40.17%, 12.20%, 12.01%, 11.89%, 11.83%, and 11.89%. The expectations of very smooth, smooth, mild, and heavy SPI levels are 38.55%, 24.68%, 18.96%, and 17.81%. The found probabilities of final congestion states are 52.35% smooth, 24.11% mild, and 23.54% heavy. This indicates that in the current situation, the congestion state is most likely to be smooth.

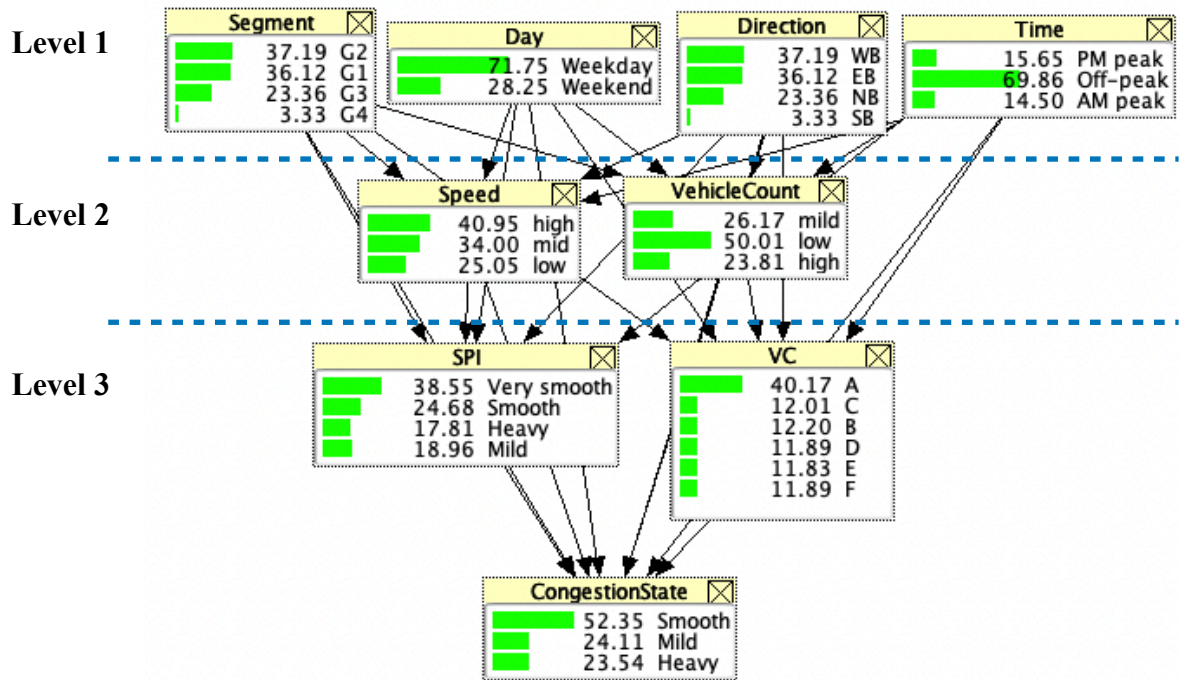


Fig. 4. Graph representation of the proposed recurring congestion BN model when there was no observation made during Covid-19 period.

For evaluating the proposed BN model, ten scenarios were generated by changing the probabilities of different variables' states to 100%. The scenario descriptions, along with the changes in expectations of states of V/C, SPI, and the final congestion states, could be found in Table 5, Table 6, and Table 7.

Scenarios 1 – 4 were generated for comparing the congestion states during weekdays and weekends. Scenario 1 was observed for setting both weekends and off-peak to be true (100%), which shows level A (free flow) of V/C with 42.43% and a very smooth level of SPI with 40.75%, which comprises to 54.15% probability for a smooth congestion state. Although the states of V/C, SPI, and the congestion states remain the same in Scenarios 2, 3, and 4, the probabilities vary. From the probability distribution, all the scenarios are observed the smooth congestion, but is slightly more to happen during the weekday and off-peak hours of weekdays, which is a reasonable results. Please refer to Table 5.

Table 5. Probability distribution in Scenario 1–4 of recurring congestion BN model during Covid.

Scenario	Description	V/C	SPI	Congestion state
1	Weekend, off-peak	A (42.43%)	Very smooth (40.75%)	Smooth (54.15%)
2	Weekday, off-peak	A (43.00%)	Very smooth (39.18%)	Smooth (54.40%)
3	Weekday, AM peak	A (39.81%)	Very smooth (41.54%)	Smooth (52.90%)
4	Weekday, PM peak	A (41.76%)	Very smooth (39.36%)	Smooth (53.93%)

Scenarios 5, 6, and 7 were generated for observing low speed and varying vehicle count. Scenario 5 (low speed, high volume) indicates a 16.75% probability of V/C level F, 25.17% mild SPI, and 33.45% heavy congestion state. However, a probability of 33.29% of the mild congestion state, 33.27% of the smooth congestion state were also observed for this combination. The states in Scenario 6 (low speed and medium volume) has 17.60% V/C level A and 26.15% mild SPI comprising 24.78% heavy congestion state. In Scenario 7 (low speed, low volume), V/C of level A (free flow) with 29.62% and SPI with 34.39% mild result 23.39% heavy congestion state, example image can refer to Fig. 5. These three scenarios demonstrate that low speed with high and medium volume is associated with a heavy congestion state, but others with smooth congestion state. Please refer to Table 6.

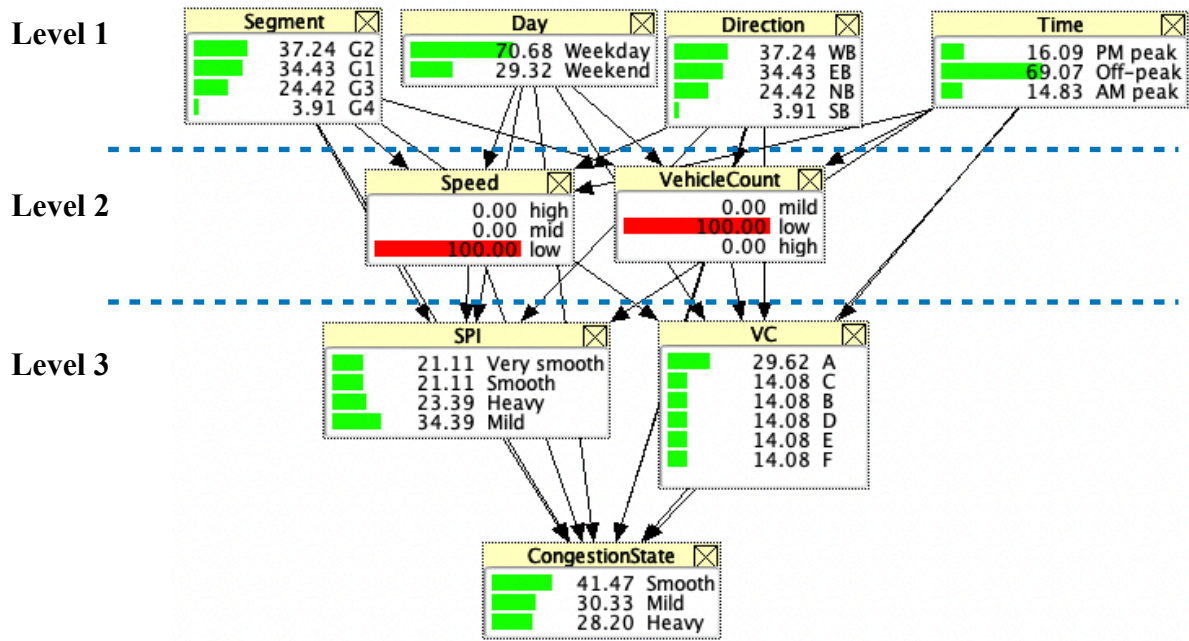


Fig. 5. Variation of probability distributions of variables given that low speed and low vehicle count was observed (Scenario 7).

Table 6. Probability distribution in Scenario 5–7 of recurring congestion BN model during Covid.

Scenario	Description	V/C	SPI	Congestion state
5	Low speed, high volume	F (16.75%)	Mild (25.17%)	Heavy (33.45%)
6	Low speed, medium volume	A (17.60%)	Mild (26.15%)	Smooth (33.77%)
7	Low speed, low volume	A (29.62%)	Mild (34.39%)	Smooth (41.47%)

The rest three Scenarios 8, 9, and 10 were generated for observing high speed and varying vehicle count. For scenario 8 (high speed, high volume), V/C level F (breakdown) with 17.49% and SPI with 27.10% very smooth, and 34.81% mild congestion state was observed. For scenario 9 (high speed, medium volume), V/C level A (free flow) with 25.73% and SPI with 38.08% very smooth, and 44.29% very smooth congestion state was observed. Finally, for scenario 10 (high speed, low volume), V/C level A (free flow) with 72.33% and SPI with 75.01% very smooth, and 77.279% very smooth congestion state was observed. These three scenarios demonstrate that high speed with low and medium volume is most likely representing a smooth congestion state with high percentage probability. Please refer to Table 7.

Table 7. Probability distribution in Scenario 8–10 of recurring congestion BN model during Covid.

Scenario	Description	V/C	SPI	Congestion state
8	High speed, high volume	F (17.49%)	Very smooth (27.10%)	Mild (34.81%)
9	High speed, medium volume	A (25.73%)	Very smooth (38.08%)	Smooth (44.29%)
10	High speed, low volume	A (72.33%)	Very smooth (75.01%)	Smooth (77.79%)

For testing dataset, I hid the columns of congestion state on testing set, training the model on training set and applying it on the data of testing set to predict, result can refer to Fig. 6. My goal is to predict the heavy congestion state via all variables including segment, direction, day, time, speed, vehicle count, V/C and SPI.

#	Segment	Direction	Day	Time	Speed	VehicleC...	VC	SPI	CongestionStateIgnored	P(CongestionState=Heavy)
0	G2	WB	Weekday	PM peak	high	mild	A	Very smo...	Smooth	0
1	G2	WB	Weekday	PM peak	high	mild	A	Very smo...	Smooth	0
2	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0
3	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0
4	G2	WB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
5	G2	WB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
6	G3	NB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
7	G3	NB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
8	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0
9	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0
10	G3	NB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
11	G3	NB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
12	G2	WB	Weekday	PM peak	high	mild	A	Very smo...	Smooth	0
13	G2	WB	Weekday	PM peak	high	mild	A	Very smo...	Smooth	0
14	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0
15	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0
16	G3	NB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
17	G3	NB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
18	G2	WB	Weekday	PM peak	high	mild	A	Very smo...	Smooth	0
19	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0
20	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0
21	G2	WB	Weekday	PM peak	high	mild	A	Very smo...	Smooth	0
22	G3	NB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
23	G2	WB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
24	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0
25	G3	NB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
26	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0
27	G2	WB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
28	G3	NB	Weekday	PM peak	high	low	A	Very smo...	Smooth	0
29	G1	EB	Weekday	PM peak	mid	low	A	Smooth	Smooth	0

Fig. 6. Partial prediction results for heavy congestion state on testing dataset during Covid.

The prediction results for this dataset achieved for 100% accuracy on predicting heavy congestion state. Plotted an ROC curve (receiver operating characteristic curve), getting the AUC (area under ROC curve) = 1. An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: True Positive Rate, False Positive Rate, refer to Fig. 7.

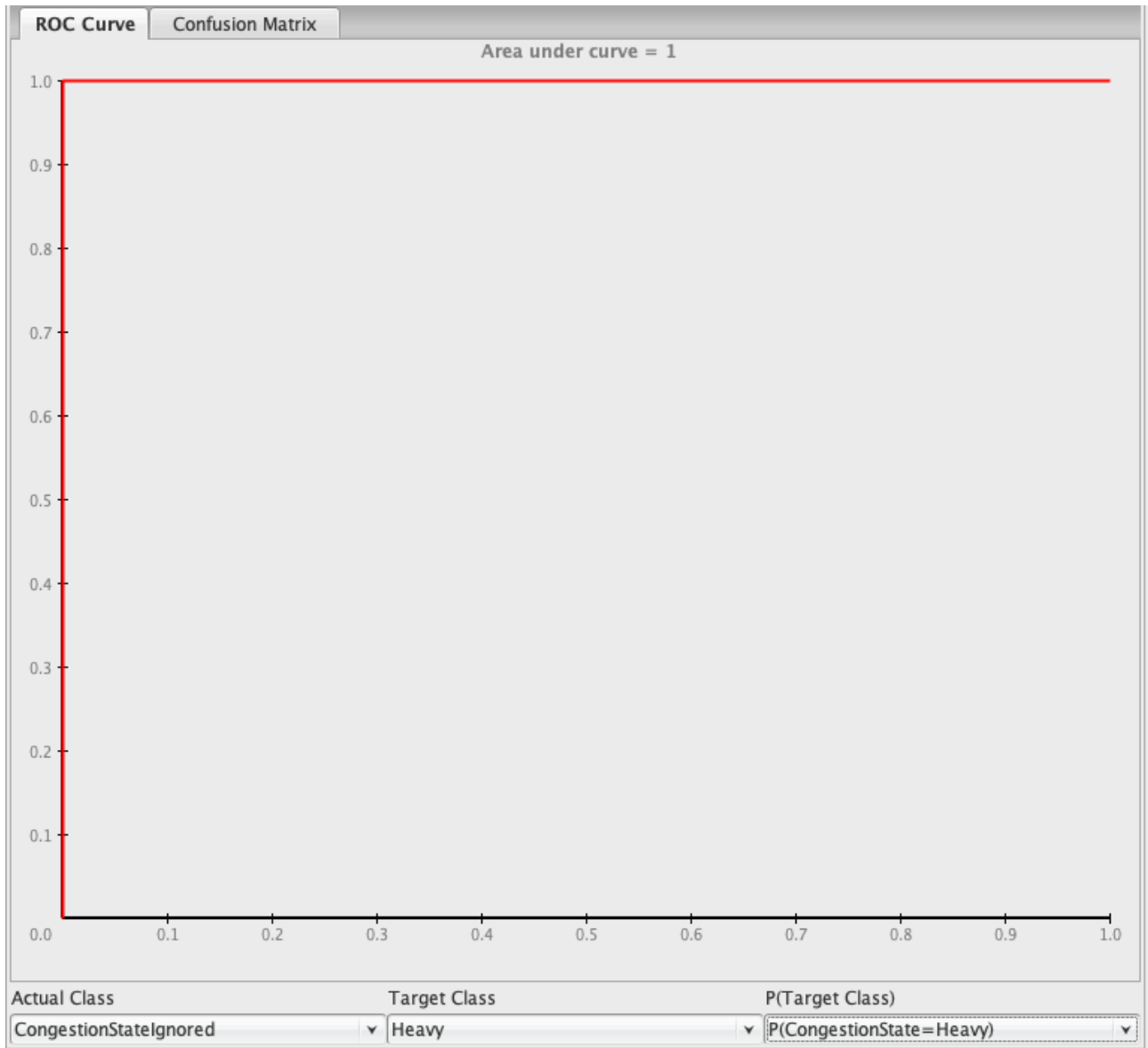


Fig. 7. ROC curve on Covid-19 period testing dataset. (Area Under Curve = 1)

For confusion matrix please refer to Table 8, it is a table that is used to define the performance of a classification algorithm. A confusion matrix visualizes and summarizes the performance of a classification algorithm.

Table 8. Confusion matrix for testing set during Covid.

	Actual values		
		Heavy	!Heavy
	Predicted values		
	Heavy	20 (True Positive)	0 (False Positive)
	!Heavy	0 (False Negative)	6434 (True Negative)

3.2.2 Recent data

For recent data, using the same variable values was taken for every 10 minutes for seven consecutive days from February 27, 2023, to May 28, 2023 as a training set, totally 21206 data; May 29, 2023 to June 4, 2023 as testing set, totally 1552 data. The estimated probability results are shown in Fig. 8. Using the same scenarios on recent dataset, all the details probability please refer to Table 9, Table 10, and Table 11.

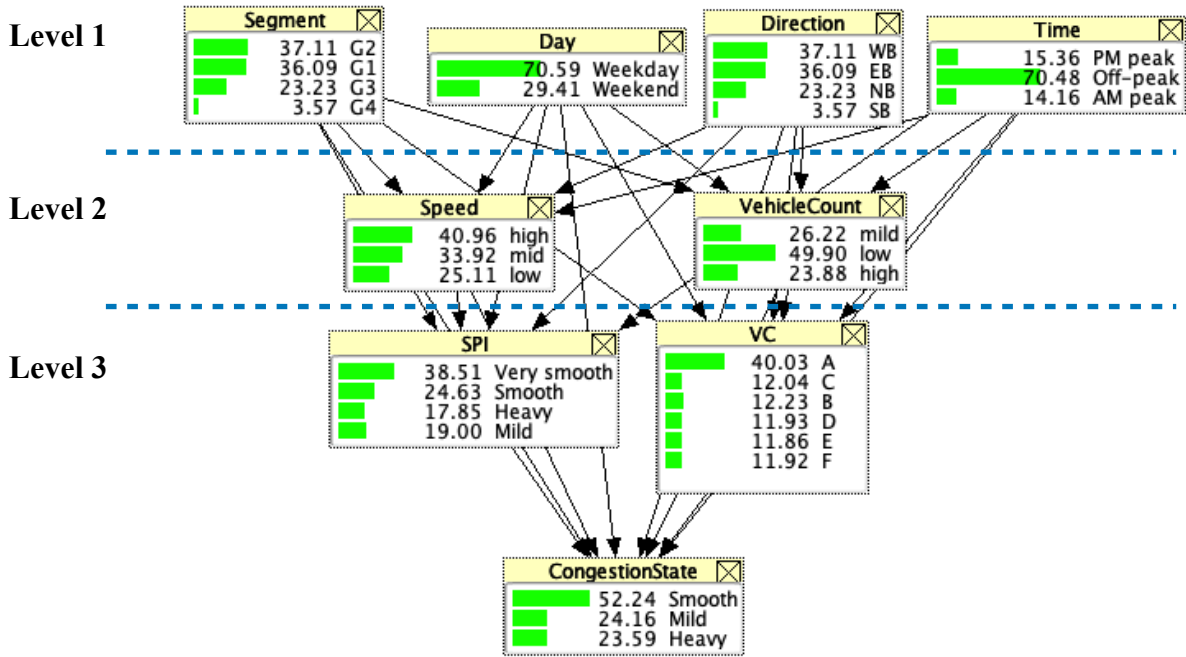


Fig. 8. Graph representation of the proposed recurring congestion BN model when there was no observation made during recent period.

Table 9. Probability distribution in Scenario 1–4 of recurring congestion BN model on recent data.

Scenario	Description	V/C	SPI	Congestion state
1	Weekend, off-peak	A (42.25%)	Very smooth (40.70%)	Smooth (54.01%)
2	Weekday, off-peak	A (42.88%)	Very smooth (39.18%)	Smooth (54.32%)
3	Weekday, AM peak	A (41.39%)	Very smooth (39.79%)	Smooth (52.88%)
4	Weekday, PM peak	A (41.67%)	Very smooth (39.48%)	Smooth (53.88%)

Table 10. Probability distribution in Scenario 5–8 of recurring congestion BN model on recent data.

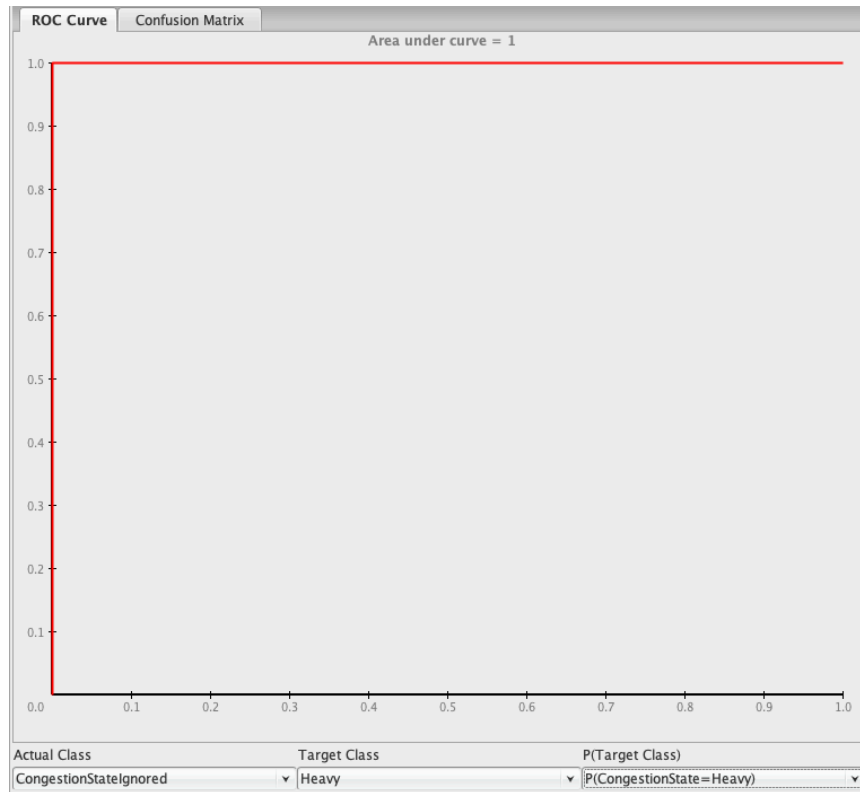
Scenario	Description	V/C	SPI	Congestion state
5	Low speed, high volume	F (16.75%)	Mild (25.18%)	Heavy (33.44%)
6	Low speed, medium volume	A (17.60%)	Mild (26.15%)	Smooth (33.86%)
7	Low speed, low volume	A (29.64%)	Mild (34.40%)	Smooth (41.48%)

Table 11. Probability distribution in Scenario 8–10 of recurring congestion BN model on recent data.

Scenario	Description	V/C	SPI	Congestion state
8	High speed, high volume	F (17.48%)	Very smooth (27.18%)	Mild (34.86%)
9	High speed, medium volume	A (25.85%)	Very smooth (38.13%)	Smooth (44.38%)
10	High speed, low volume	A (72.12%)	Very smooth (74.90%)	Smooth (77.69%)

Comparing with Table 5, Table 6, Table 7 with Table 9, Table 10, Table 11 respectively, there are very close to no difference, the V/C, SPI and congestion state categories are totally the same. It represent that the congestion state are similar during Covid-19 period and recent period in this report and result.

The prediction results for recent testing dataset achieved also 100% accuracy on predicting heavy congestion state, the AUC = 1, refer to Fig. 9. Confusion matrix refer to Table 12.

**Fig. 9.** ROC curve on recent testing dataset. (Area Under Curve = 1)**Table 12.** Confusion matrix for testing set on recent data.

		Actual values	
		Heavy	!Heavy
Predicted values	Heavy	1 (True Positive)	0 (False Positive)
	!Heavy	0 (False Negative)	1551 (True Negative)

4. Discussion

The Estimation of Recurring Traffic Congestion Using Bayesian Network has emerged as a promising method to predict traffic patterns. This report examined the performance of this approach using two different time slots: one during the COVID-19 pandemic and the other based on recent data. Surprisingly, the study found minimal differences between the two time slots, with both achieving 100% accuracy on the testing dataset. However, it may have happened because our heavy data were too little, the distribution of dataset are not balanced.

The COVID-19 pandemic brought about significant changes in commuting patterns and travel behavior worldwide. Lockdowns, remote work arrangements, and restricted movement led to a decline in overall traffic volume. Consequently, one might expect a noticeable impact on recurring traffic congestion patterns. The study in question aimed to assess the impact of the COVID-19 pandemic on recurring traffic congestion using a Bayesian Network model. Two distinct time slots were selected: one during the pandemic and another based on recent data when restrictions had been lifted. Surprisingly, the analysis revealed that there were almost no differences in traffic congestion patterns between the two time slots. The finding of minimal variation suggests that the Bayesian Network model remained robust during the pandemic, accurately predicting traffic congestion despite the altered travel patterns. Achieving 100% accuracy on the testing dataset further validates the model's reliability. The study's results also indicate that the pandemic-induced changes in travel behavior may not have significantly altered the underlying patterns of recurring traffic congestion. This insight challenges assumptions that disruptions like the pandemic would inevitably lead to fundamental shifts in traffic dynamics. However, it is important to note that this study focused on two specific time slots, and a broader investigation incorporating more diverse periods would provide a more comprehensive understanding of traffic patterns.

While the study's findings are promising, several limitations should be acknowledged. The analysis was restricted to two specific time slots, potentially overlooking variations that could arise during other periods. Additionally, the study primarily relied on data from a single geographical location, limiting the generalizability of the results. Future research should aim to incorporate data from diverse locations and examine longer time periods to validate the robustness of the Bayesian Network approach. Furthermore, exploring the model's performance on larger datasets and assessing its accuracy under different scenarios would strengthen its applicability in real-world traffic management.

5. Conclusion

In conclusion, the Estimation of Recurring Traffic Congestion Using Bayesian Network has proved to be a reliable and effective method for predicting traffic congestion patterns. By utilizing two different time slots, one during the COVID-19 pandemic and the other based on recent data, the study aimed to assess whether there were any notable variations in traffic congestion patterns.

Surprisingly, the findings revealed that there were almost no differences between the two time slots. This unexpected outcome suggests that the pandemic and subsequent changes in traffic patterns did not significantly impact the recurring traffic congestion. Despite the numerous changes in commuting patterns and travel behaviors during the pandemic, the Bayesian Network model exhibited remarkable consistency in its predictions.

The model's 100% accuracy on the testing dataset further reinforces its reliability. This high level of accuracy signifies the model's ability to capture and understand the complex relationships between various factors that contribute to traffic congestion, such as road conditions, weather, and time of day. By accurately predicting congestion patterns, transportation authorities and urban planners can make informed decisions to mitigate traffic-related issues, improve infrastructure, and optimize traffic flow.

The findings of this study have significant implications for transportation planning and management, as well as for policymakers. It highlights the robustness and stability of the Bayesian Network approach in estimating recurring traffic congestion, emphasizing its potential as a valuable tool in traffic management systems.

However, it is important to acknowledge the limitations of the study. The analysis focused solely on two specific time slots, and it would be beneficial to expand the investigation to include a more comprehensive range of time periods. Additionally, while the model achieved 100% accuracy on the testing dataset, it is essential to assess its performance on larger and more diverse datasets to ensure its generalizability.

Overall, the Estimation of Recurring Traffic Congestion Using Bayesian Network presents a promising avenue for accurately predicting and managing traffic congestion. The study's findings offer valuable insights for traffic planners, transportation authorities, and policymakers, paving the way for more efficient and effective traffic management strategies in the future.