Bayesian Network Modeling Applied on Railway Level Crossing Safety

Ci LIANG, 15/11/2017

Co-authors: Mohamed GHAZEL (IFSTTAR-COSYS/ESTAS)
Olivier CAZIER (SNCF Réseau)
Laurent BOUILLAUT (IFSTTAR-COSYS/GRETTIA)
El-Miloudi EL-KOURSI (IFSTTAR-COSYS/ESTAS)

CONTENTS

Introduction

- More than **118,000** LXs in the 28 countries of the E.U.;
- More than **1,200** accidents in E.U. every year;
- More than **300** deaths in E.U. per year;
- Over **15,000** LXs in France;
- Around **13,000** show heavy roads and railway traffic.

Present Works

BNI-RR framework: Bayesian Network (BN) based Inference for Risk Reasoning (BNI-RR)

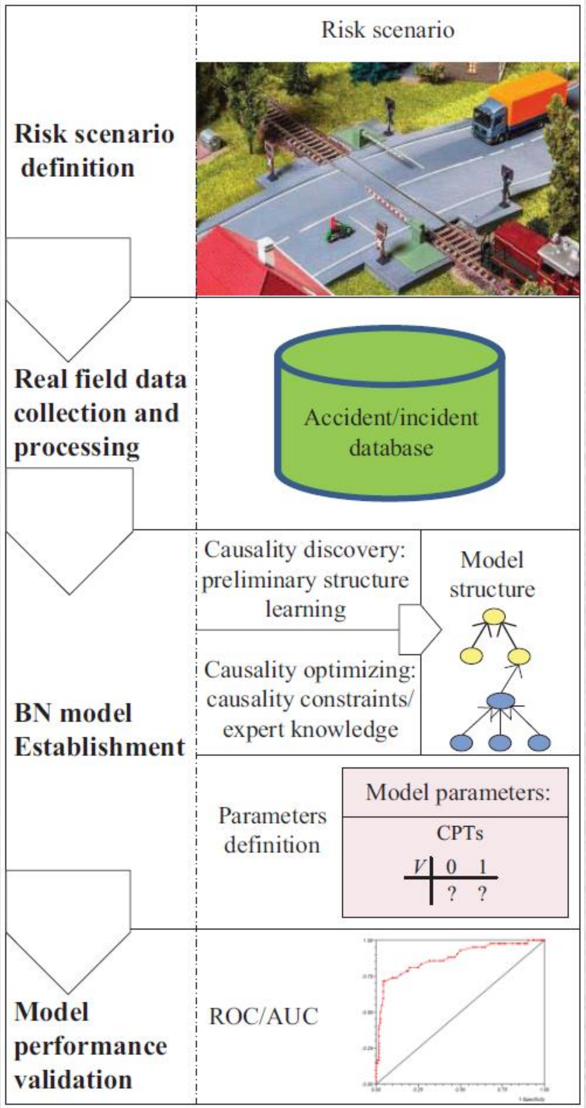


Fig. 1. BNI-RR framework

- Risk scenario definition
- Real field data collection and processing
- BN model establishment
  - Causality discovery
  - Causality optimizing
  - Parameter definition
- Model prediction performance validation
  - The Receiver Operating Characteristic (ROC) curve
  - The Area Under the ROC Curve (AUC)

Present Works - Risk scenario definition

1) Automated LX with 4 half barriers and lights (**SAL4**);
2) Automated LX with 2 half barriers and lights (**SAL2**);
3) Automated LX with lights but without barriers (**SAL0**);
4) Crossbuck LX.

SAL: Signalisation Automatique Lumineuse

The main transport mode causing accidents at SAL2: motorized vehicle

“Motorized vehicles cross SAL2 LXs when trains are approaching”

Table 1. Accidents at French LXs during the last 40 years

Type of LX	Number	# Accident
SAL4	> 600	> 600
SAL2	> 10,000	> 4,000
SAL0	> 60	> 50
Crossbuck LX	> 3,000	> 700

(a) SAL4 (b) SAL2 (c) SAL0 (d) Crossbuck LX

Present Works - Real field data collection and processing

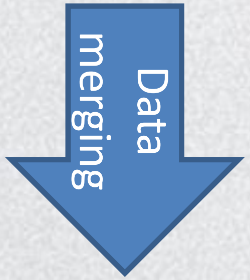
Database 1 (D1): accident 2004-2013

- Get all static parameters we want to analyze;
- Fatalities and injuries;
- No accident causes and no relationship between human factors and accidents;

Database 2 (D2): accident 2010 – 2013

- Including accident causes: zigzag, alignment, etc.
- Lack of static parameters for each LX :

Using LX ID, line ID and Date to make data merging

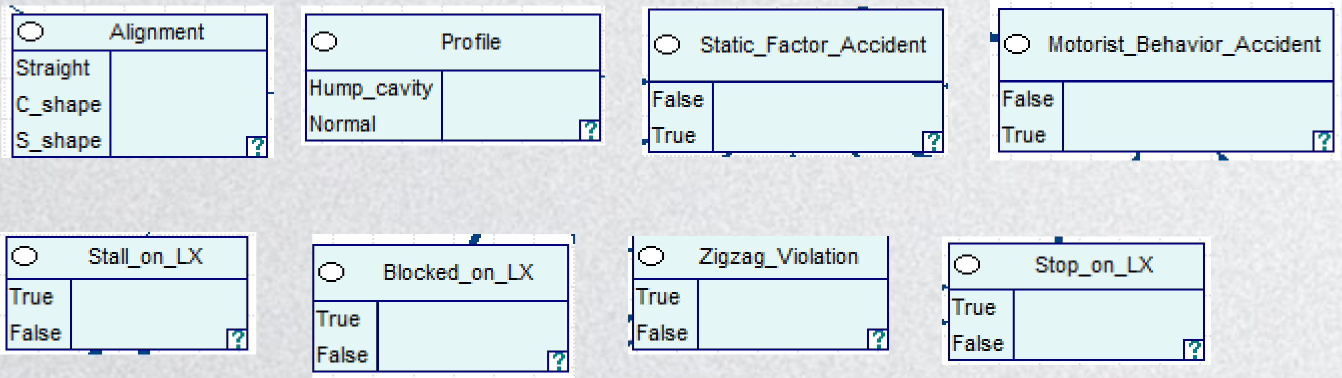


A new database, named ND, 2010 - 2013: 1) the LX accident information, 2) static railway, roadway and LX characteristics, 3) the number of fatalities and injuries, and accident causes related to static factors and motorist behavior.

Data processing

- **Continuous variables -> Discretization: Railway Speed Limit, Corrected Moment, Railway Traffic Density, Road Traffic Density, Width, Length, Region risk factor;**
  Divided into 3 groups, each group having the similar number of samples;

- **Finite discrete variables: Alignment, Profile, Motorist Inappropriate Behavior, Stall on LX, Blocked on LX, Stop on LX;**
  Each value corresponds to a state;



Definition of states of nodes

Table 2. States of nodes in the BN risk model

Node name	Node property	Node state
TC nodes		
Average Daily Railway Traffic (ADRT)	Chance node	ADRT_below_9 ($0 \leq$ ADRT < 9), ADRT_9_25 ($9 \leq$ ADRT < 25), ADRT_25_up ($25 \leq$ ADRT);
Average Daily Road Vehicle (ADRV)	Chance node	ADRV_below_72 ($0 \leq$ ADRV < 72), ADRV_72_403 ($72 \leq$ ADRV < 403), ADRV_403_up ($403 \leq$ ADRV);
Blocked on LX (B)	Chance node	True, False;
Stop on LX (Stop)	Chance node	True, False;
SC nodes		
Corrected Moment (CM)	Chance node	CM_below_19 ($0 \leq$ CM < 19), CM_19_49 ($19 \leq$ CM < 49), CM_49_up ($49 \leq$ CM);
Railway Speed Limit (RLS)	Chance node	RLS_below_70 (0 km/h $\leq$ RLS $<$ 70 km/h), RLS_70_110 (70 km/h $\leq$ RLS $<$ 110 km/h), RLS_110_up (110 km/h $\leq$ RLS);
Alignment (A)	Chance node	Straight, C_shape, S_shape;
Profile (P)	Chance node	Normal, Hump_cavity;
Width (W)	Chance node	W_below_5 (0 m $\leq$ W $<$ 5 m), W_5_6 (5 m $\leq$ W $<$ 6 m), W_6_up (6 m $\leq$ W);
Length (L)	Chance node	L_below_7 (0 m $\leq$ L $<$ 7 m), L_7_11 (7 m $\leq$ L $<$ 11 m), L_11_up (11 m $\leq$ L);
Region Risk (R)	Chance node	R_low (region with low risk level), R_medial (region with medial risk level), R_high (region with high risk level);
Stall on LX (Stall)	Chance node	True, False;
Zigzag Violation (ZV)	Chance node	True, False;
PriC nodes		
Motorist Behavior Accident (MB)	Chance node	True, False;
Static Factor Accident (SF)	Chance node	True, False;
Consequence nodes		
SAL2 MV Accident (SA)	Chance node	True, False;
Fatalities (F)	Chance node	F_0 (F = 0), F_0_up (0 < F);
Severe Injuries (S)	Chance node	S_0_2 ($0 \leq$ S < 2), S_2_up ($2 \leq$ S);
Minor Injuries (M)	Chance node	M_0_3 ($0 \leq$ M < 3), M_3_up ($3 \leq$ M);
Consequence Severity (CS)	Deterministic node	Level_1, Level_2, Level_3, Level_4, Level_5;

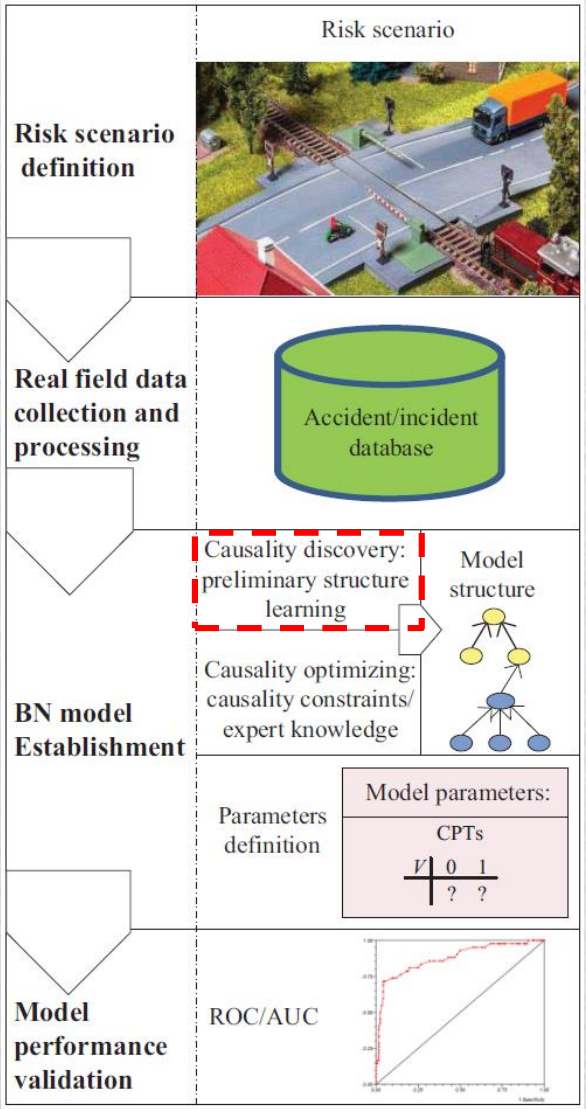
- **PriC nodes: Primary causes;**
- **SC nodes: Secondary causes;**
- **TC nodes: Third-level causes;**
- **Consequence nodes;**

Table 3. Consequence severity definition

Consequence severity	Level 1	Level 2	Level 3	Level 4	Level 5
0 = fatalities, $0 \leq$ severe injuries < 2, $0 \leq$ minor injuries < 3;	×	–	–	–	–
0 = fatalities, $0 \leq$ severe injuries < 2, $3 \leq$ minor injuries;	–	×	–	–	–
0 = fatalities, $2 \leq$ severe injuries, $0 \leq$ minor injuries < 3;	–	–	×	–	–
0 = fatalities, $2 \leq$ severe injuries, $3 \leq$ minor injuries;	–	–	–	×	–
0 < fatalities;	–	–	–	–	×

- Increasing progressively from level 1 to 5;

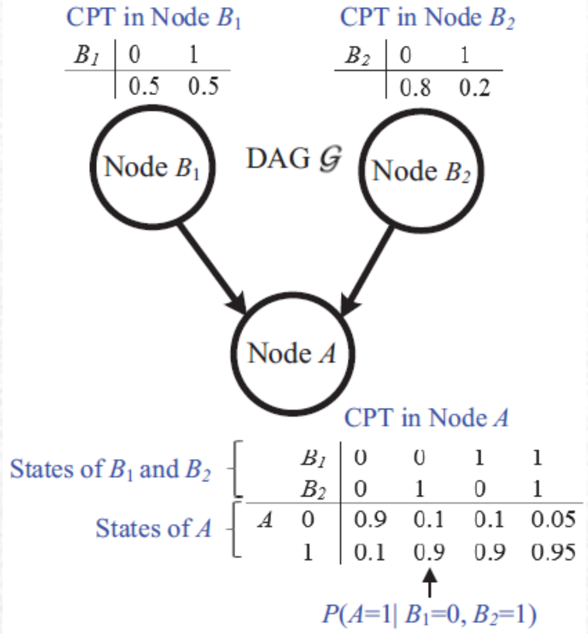
Fig. 1. BNI-RR framework



- **Causality discovery**
  - Causal BN

$BN = (P, \mathcal{G})$

$\mathcal{G} = (N, L)$



$\mathcal{G}_C$: a 3-tuple causal DAG (Directed Acyclic Graph)

$$\mathcal{G}_C = \{IF, THEN, CAK\}$$

IF: a set of causes, $IF = \{x_1, x_2, \dots, x_n\}$

$THEN$: a set of consequences, $THEN = \{y_1, y_2, \dots, y_m\}$

CAK (CAusal Knowledge): a set of directed pairs of the cause x_i and consequence y_j

$$CAK = \{(x_i, y_j) | x_i \in IF, y_j \in THEN, IF \neq \emptyset, THEN \neq \emptyset\}$$

(x_i, y_j): a directed variable pair that describes the structure of $\mathcal{G}_C$: $x_i \rightarrow y_j$

e.g., $\mathcal{G}_C = \{IF = \{B_1, B_2\}, THEN = \{A\}, CAK = \{(B_1, A), (B_2, A)\}\}$

Present Works - BN model establishment

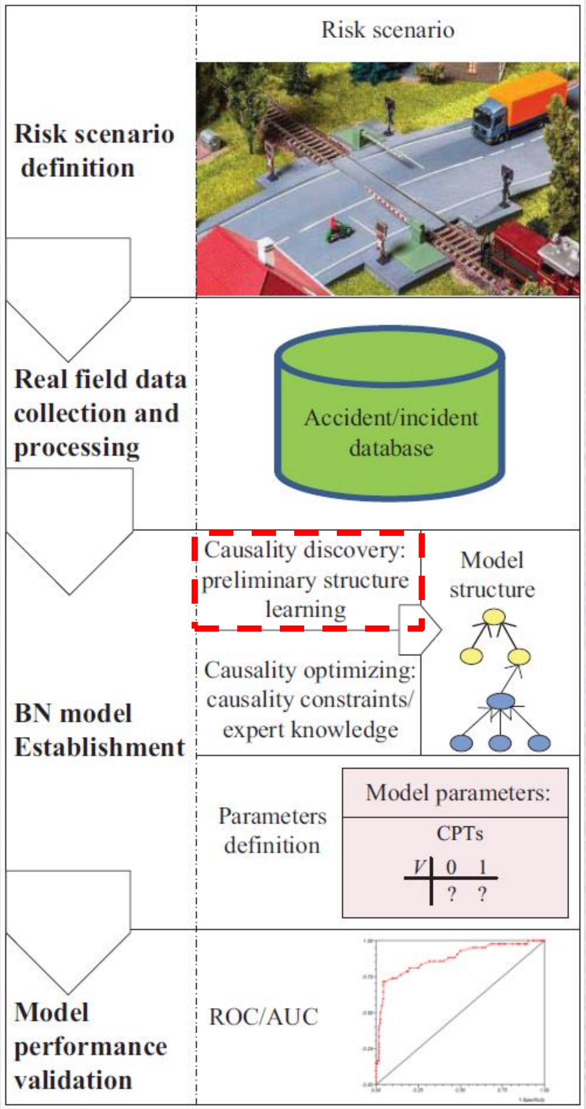


Fig. 1. BNI-RR framework

- **Preliminary causality discovery: automatic structure learning**
  - The Bayesian Search (BS) algorithm
  - The Essential Graph Search (EGS) algorithm
  - The Greedy Thick Thinning (GTT) algorithm
  - The Naive Bayes approach
  - The Augmented Naive Bayes (ANB) algorithm
  - The Tree Augmented Naive Bayes (TAN) algorithm

Shortcomings:

- **inconsistent with the causal relationships in reality;**
- **more likely correlations rather than causalities in reality;**
- **impede identification of important causes;**

Present Works - BN model establishment

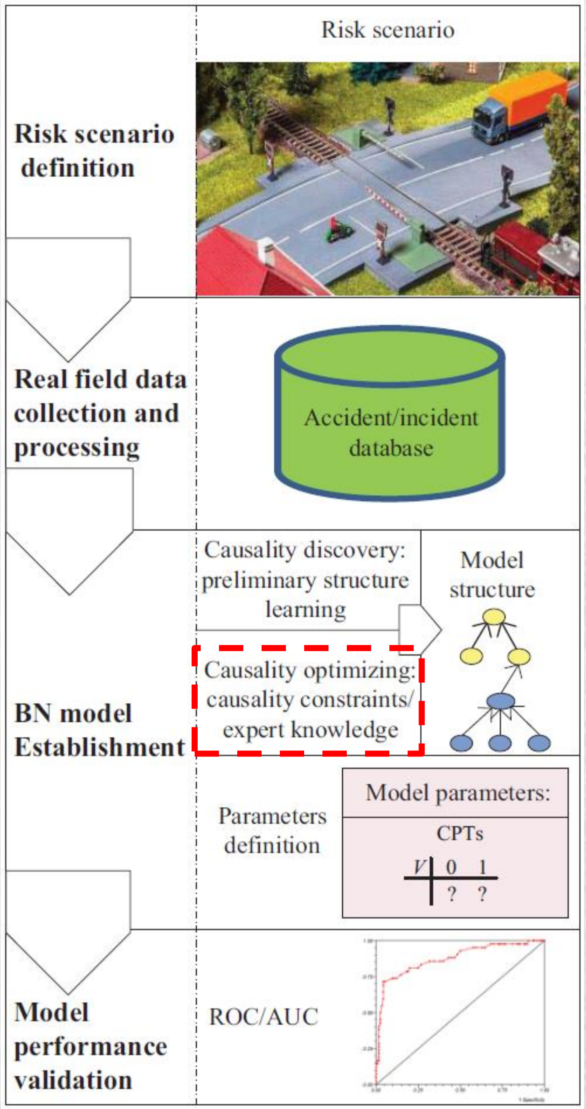


Fig. 1. BNI-RR framework

- **Causality optimizing**
  - **Causal structural constraints (CSCs) : expert knowledge**
  - Definition of 3 types of directed CSCs: $x \in IF, y \in THEN$
    - ✓ Existence Constraint (EC), $(x, y)_e$

      Must be a direct connection from x to y;
    - ✓ Forbidden Constraint (FC), $(x, y)_f$

      Must not be a direct connection from x to y;
    - ✓ Potential Directed Constraint (PDC), $(x, y)_p$

      If there exists a direct connection between x and y,

      it should be from x to y;

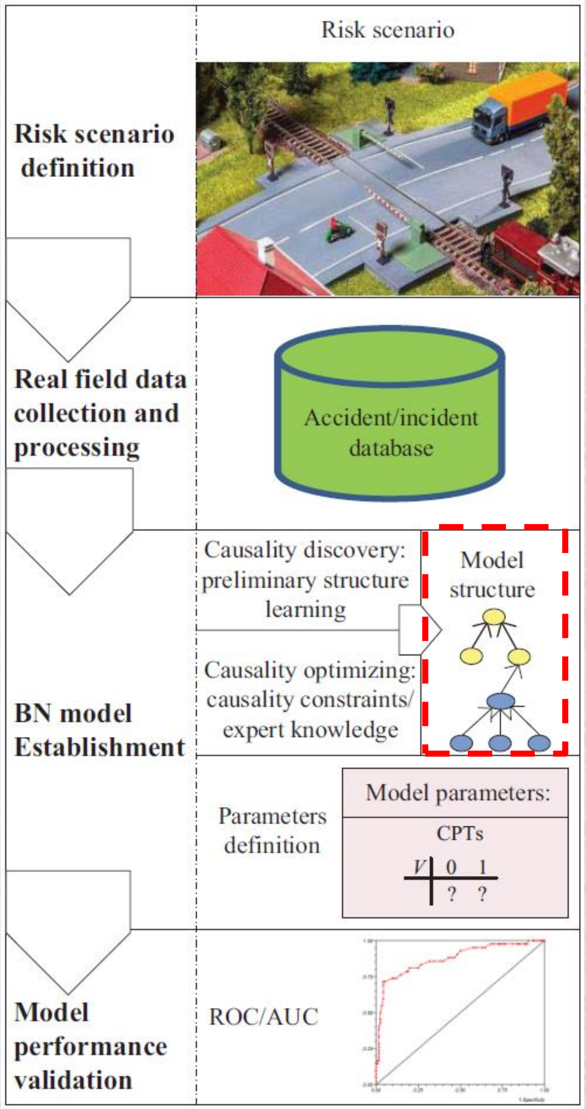
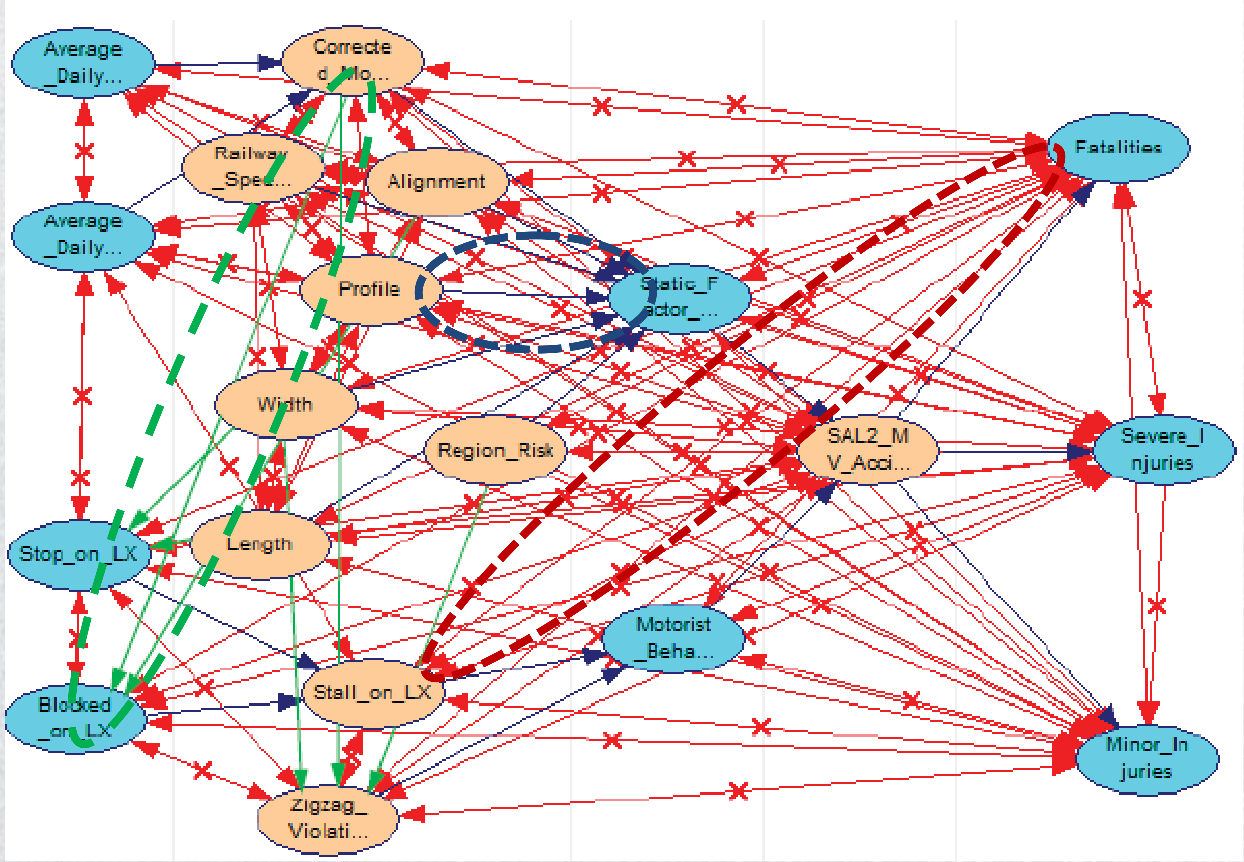


Fig. 1. BNI-RR framework

- **Model structuring**
  - **Causal structural constraints (CSCs) : expert knowledge**

Fig. 2. CSCs identified for our BN risk model



- **ECs: Blue,**
- **FCs: red,**
- **PDCs: green;**

- Conditional probability parameters: generated directly from our field data

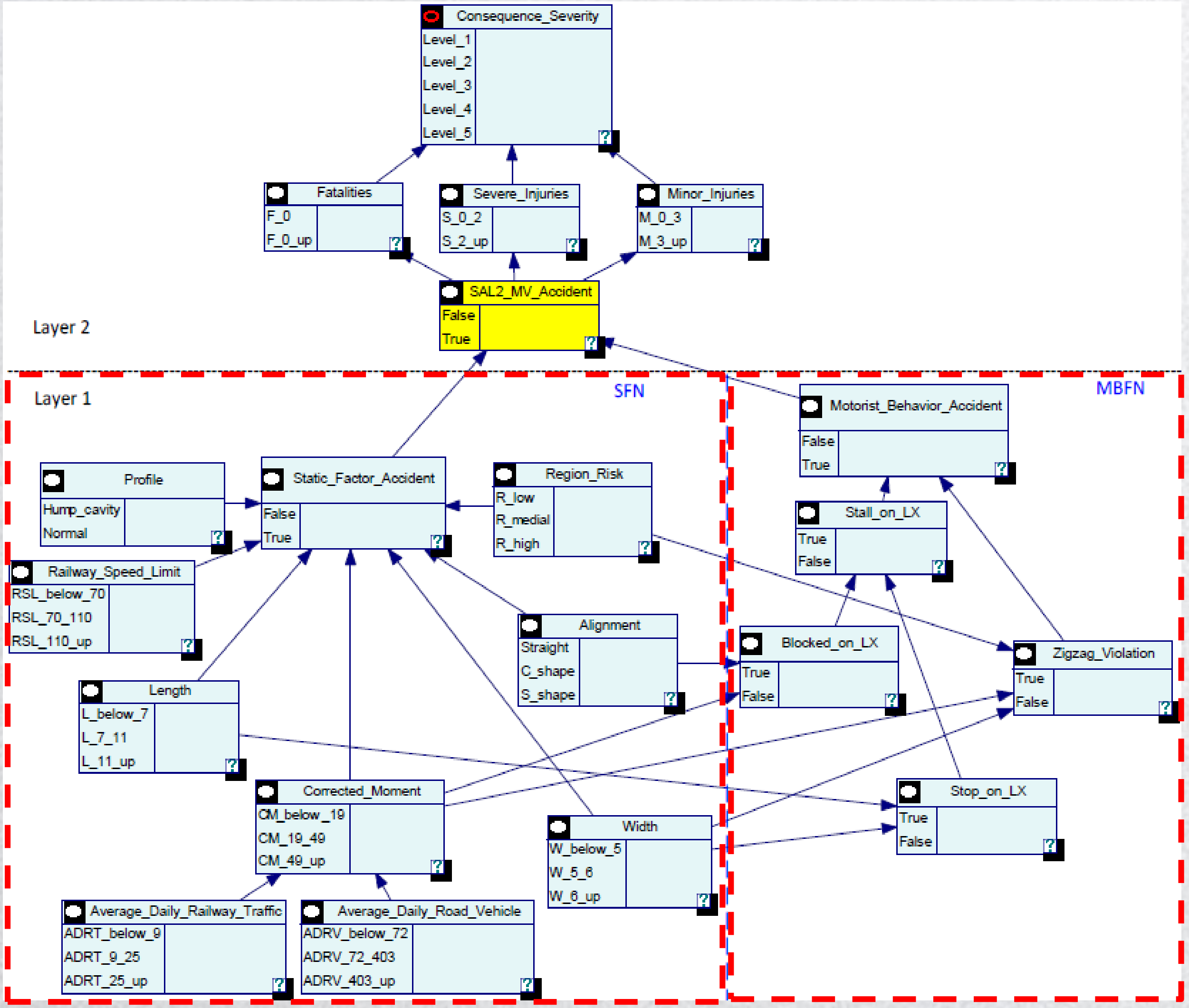


Fig. 3. Final BN risk model

- **Layer 1: cause diagnosis**
  - Part1: static factor network (SFN);
  - Part2: motorist behavior factor network (MBFN);
- **Layer 2: evaluating consequences**
  - Fatalities, Severe Injuries, Minor Injuries;
  - Consequence severity;

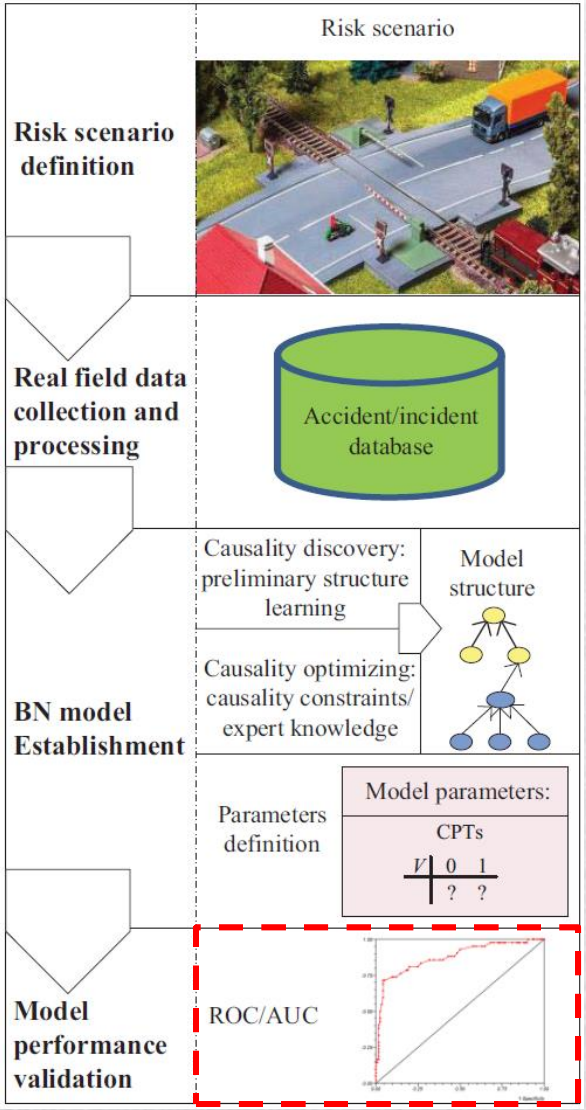


Fig. 1. BNI-RR framework

- **Receiver Operating Characteristic (ROC) curve and the Area Under the ROC Curve (AUC)**

- 1) If AUC = 1, a perfect prediction model.
- 2) If 0.5 < AUC < 1, better than random guessing and has relatively sound predictive value;
- 3) If AUC = 0.5, the same as random guessing, for example, throwing coin, having no predictive value;
- 4) Otherwise, AUC < 0.5, worse than random guessing and valueless;

- The ideal perfect ROC curve is a point (0, 1);
- The closer the AUC to 1, the better the performance of a prediction model.

Model performance validation

Table 4. Comparison of entire prediction performance

Approach	SA ACCU, AUC_T, AUC_F	F ACCU, AUC_0, AUC_{0_up}	S ACCU, AUC_{0_2}, AUC_{2_up}	M ACCU, AUC_{0_3}, AUC_{3_up}
Our model	0.9963, 0.9846, 0.9846	0.9801, 0.9964, 0.9964	0.9982, 0.9929, 0.9929	0.9913, 0.9963, 0.9963
BS	0.8751, 0.9187, 0.9187	0.8101, 0.5708, 0.5708	0.8638, 0.9541, 0.9541	0.8657, 0.9683, 0.9683
EGS	0.9134, 0.8857, 0.8857	0.9203, 0.8306, 0.8306	0.8509, 0.7790, 0.8509	0.8917, 0.8157, 0.8157
GTT	0.8706, 0.8216, 0.8216	0.8610, 0.8213, 0.8213	0.8704, 0.7126, 0.7126	0.8713, 0.8315, 0.8315
NB	0.6356, 0.5163, 0.5163	0.7704, 0.5856, 0.5856	0.8333, 0.6012, 0.6012	0.6181, 0.2015, 0.2015
ANB	0.9287, 0.9015, 0.9015	0.9516, 0.9340, 0.9340	0.9287, 0.9111, 0.9111	0.9414, 0.9202, 0.9202
TAN	0.9539, 0.9431, 0.9431	0.9636, 0.9616, 0.9616	0.9891, 0.9680, 0.9680	0.9847, 0.9794, 0.9794

ACCU: accuracy;

The closer to 1, the better;

Table 5. Comparison of prediction performance for accident/consequence occurrence

Approach	SA $ACCU_{False}$,	$ACCU_{True}$	F $ACCU_0$,	$ACCU_{0\_up}$	S $ACCU_{0\_2}$,	$ACCU_{2\_up}$	M $ACCU_{0\_3}$,	$ACCU_{3\_up}$
Our model	1	0.9622	1	0.9020	1	0.6	1	0.75
BS	0.8815	0.7181	0.9101	0.6728	0.9168	0	0.9674	0.1250
EGS	0.9242	0.6875	0.9233	0.7326	0.9615	0	0.9510	0.1250
GTT	0.9562	0.7162	0.9410	0.7813	0.9704	0	0.9613	0.1250
NB	0.6905	0.5637	0.7514	0.5576	0.7334	0	0.7182	0
ANB	1	0.8011	0.9616	0.8341	0.9617	0.2000	0.9202	0.2500
TAN	0.9693	0.8437	0.9561	0.6614	0.9238	0	0.9742	0

ACCU: accuracy;

The closer to 1, the better;

*Note that: the sample size of single accident related to "severe injuries more than 2" and "minor injuries more than 3" is small in reality, which lead to the lower accuracy.

Forward inference and reverse inference

Fig. 4. General prediction (forward inference)

SAL2_MV_Accident

False	0.99390817
True	0.0060918269

Static_Factor_Accident

False	0.99894278
True	0.0010572159

Motorist_Behavior_Accident

False	0.99511565
True	0.0048843469

Fatalities

F_0	0.99931467
F_0_up	0.00068533053

Severe_Injuries

S_0_2	0.99987146
S_2_up	0.00012853755

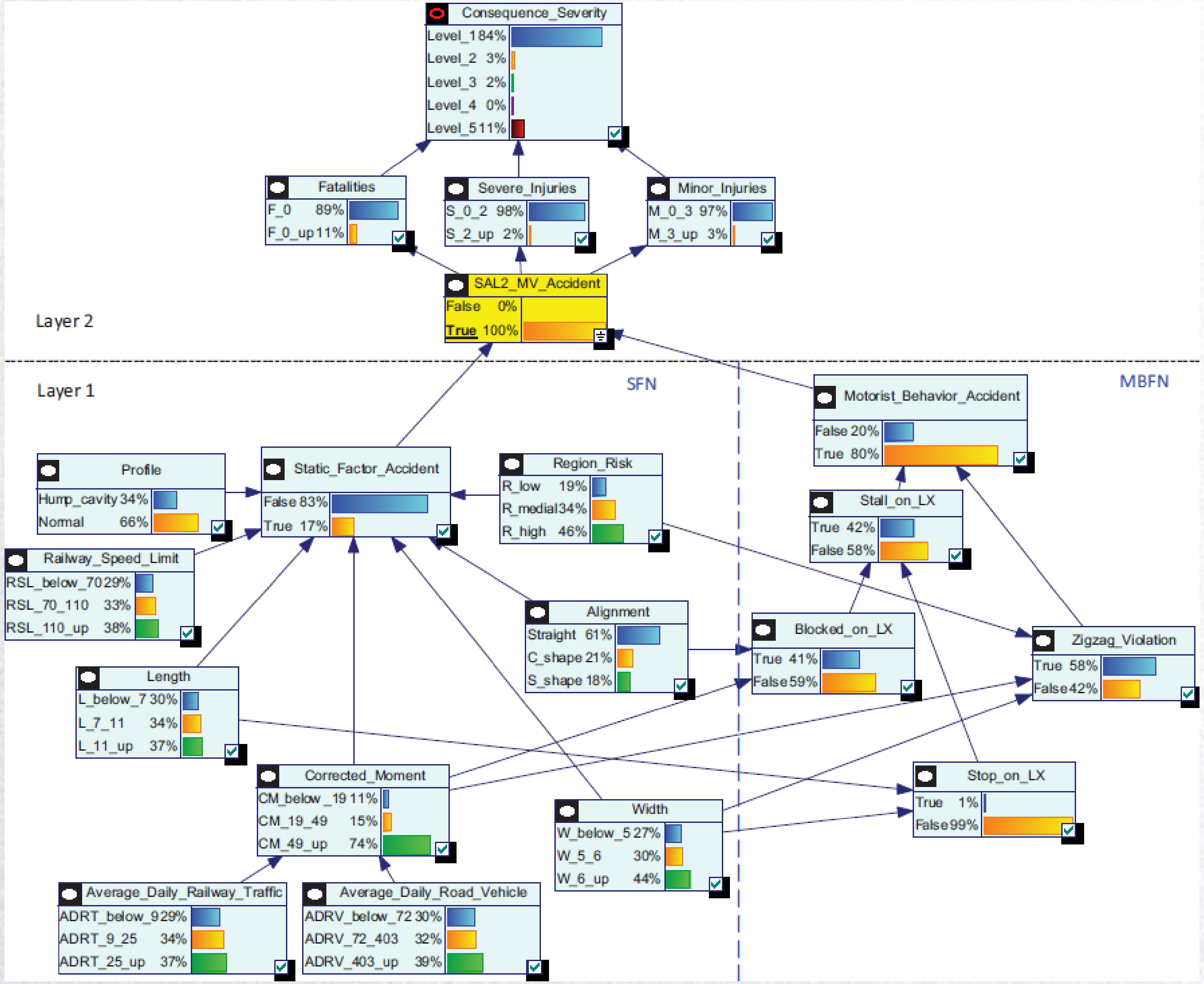
Minor_Injuries

M_0_3	0.99979531
M_3_up	0.00020468538

Consequence_Severity

Level_1	0.99902277
Level_2	0.00017782529
Level_3	0.00011024408
Level_4	3.8329897e-006
Level_5	0.00068533053

Fig. 5. Cause diagnosis when a train-MV accident occurs (reverse inference)



Influence strength analysis

The impact level of parent nodes on corresponding child nodes;

- **Consequence:**

Fatalities: strongest influence;

- **SAL2 accident occurrence:**

Inappropriate motorist behavior: strongest influence;

  - **Static factors:**

  Region risk factor: strongest influence;

  - **inappropriate motorist behavior:**

  Zigzag violation: strongest influence;

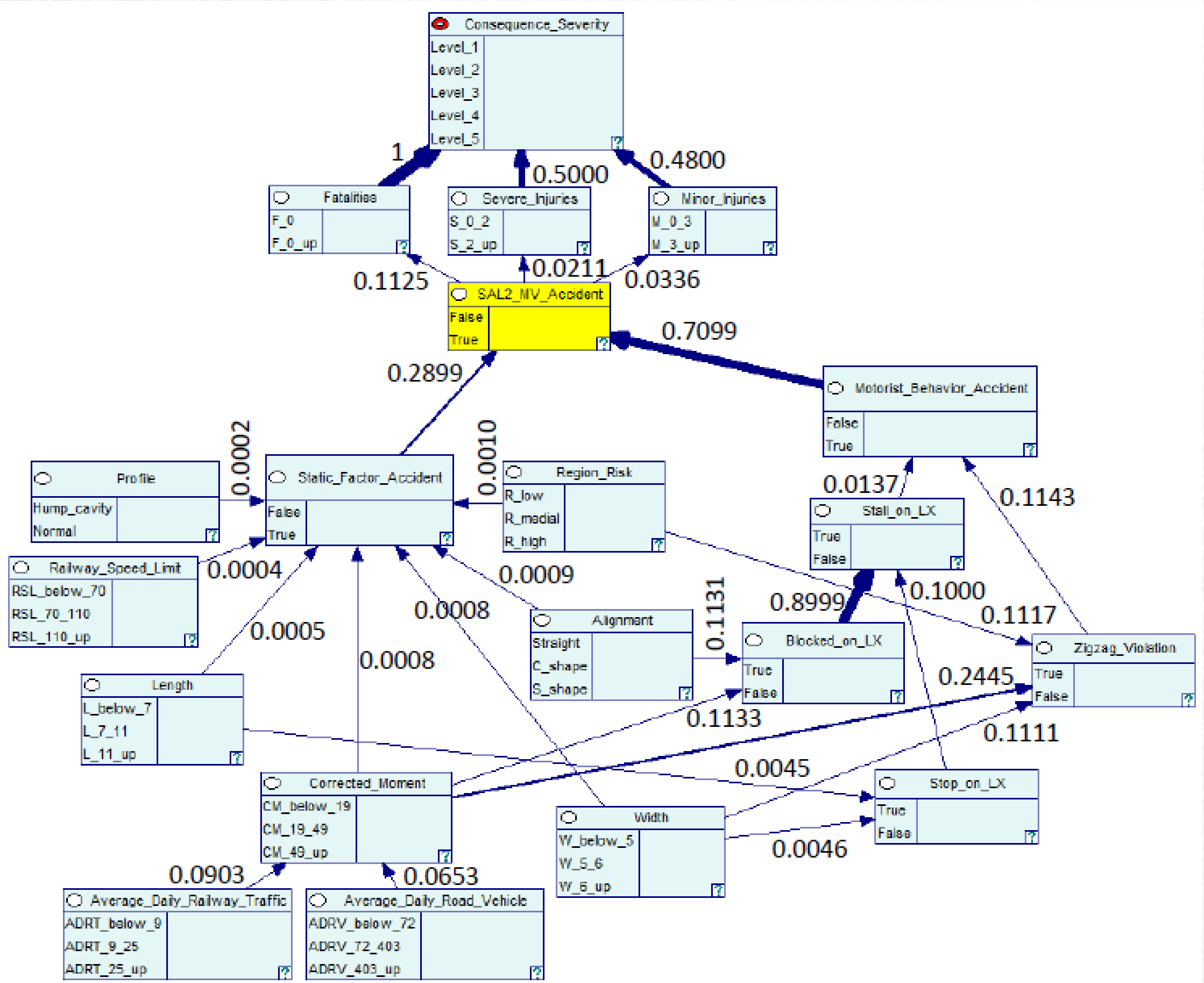
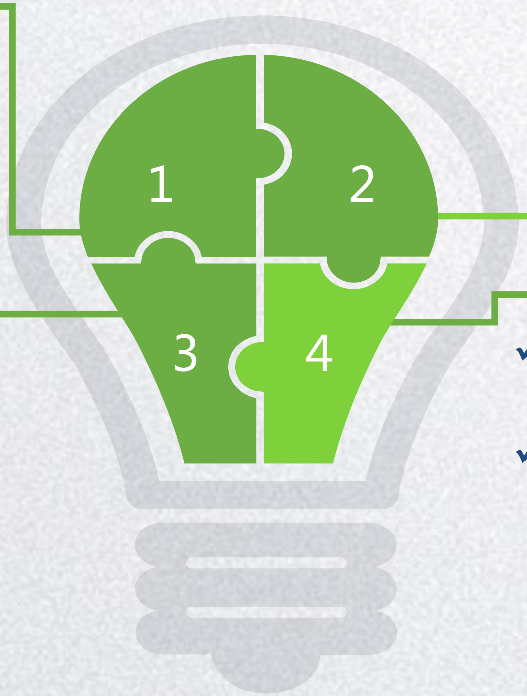


Fig. 6. Normalized influence strength analysis

Contributions & Perspective

- ✓ An effective and comprehensive modeling framework for risk reasoning: BNI-RR;
- ✓ A thorough performance validation process;

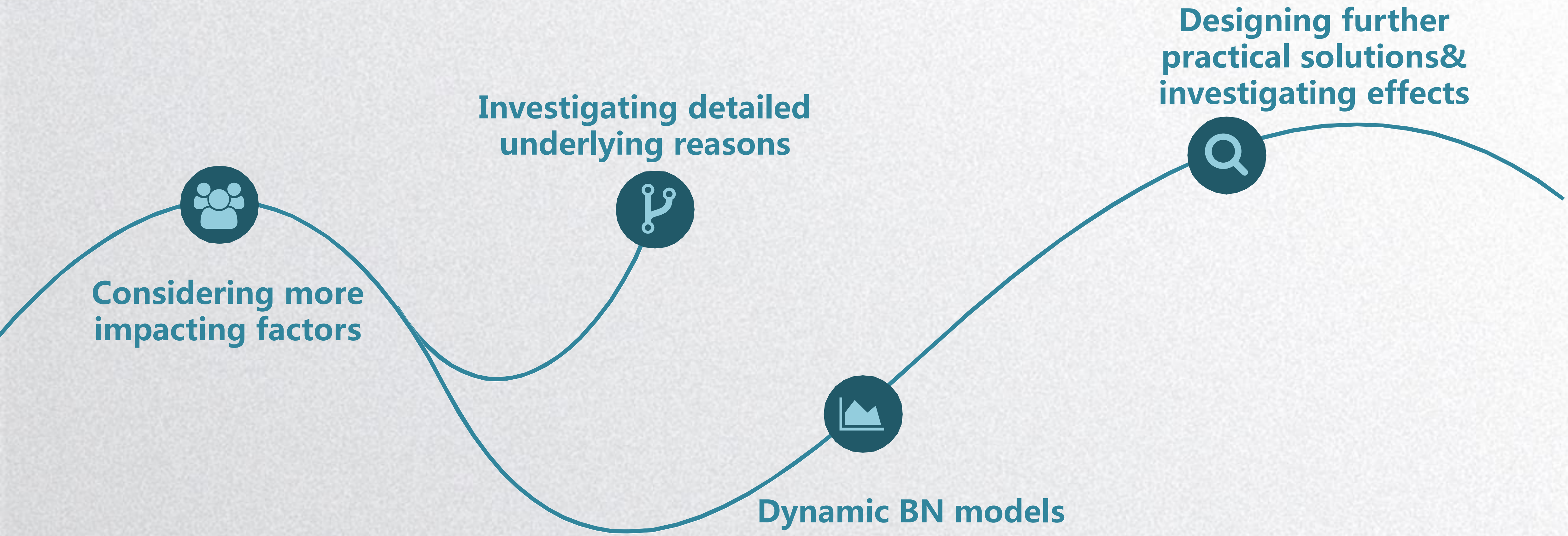


- ✓ CSCs based on the concept of CAK for empirical knowledge;
- ✓ Forward inference and reverse inference analysis;
- ✓ Influence strength analysis;

Valuable reference

Accident prediction &cause diagnosis

Identifying targeted practical design measures & improvement recommendations;


Designing further practical solutions& investigating effects
Investigating detailed underlying reasons
Considering more impacting factors
Dynamic BN models

THANK YOU FOR YOUR ATTENTION!

Bayesian Network Modeling Applied on Railway Level Crossing Safety

Project: MORIPAN: MOdèle de RIsque pour les PAssages à Niveau

Ci LIANG, 15/11/2017