5_25_2024---5.3.1403---AI_Webinar IUMS

World Neurosurgery: X 23 (2024) 100280

Contents lists available at ScienceDirect

World Neurosurgery: X

journal homepage: www.journals.elsevier.com/world-neurosurgery-x

Development of a regional-based predictive model of incidence of traumatic spinal cord injury using machine learning algorithms

Seyed Behnam Jazayeri [a,b,1], Seyed Farzad Maroufi [b,c,1], Shaya Akbarinejad [d], Zahra Ghodsi [e,f], Vafa Rahimi-Movaghar [a,e,f,*]

[a] Sina Trauma and Surgery Research Center, Tehran University of Medical Sciences, Tehran, Iran
[b] Students' Scientific Research Center (SSRC), Tehran University of Medical Sciences (TUMS), Tehran, Iran
[c] Faculty of Medicine, Tehran University of Medical Sciences (TUMS), Tehran, Iran
[d] TRON - Translational Oncology at the University Medical Center of the Johannes Gutenberg University Mainz gGmbH, Mainz, Germany

Traumatic Spinal Cord Injury (TSCI)

- Incidence: 10.4-59.0 injured/million/year

Singh 2014; Wyndaele 2006

- Data: 49 countries

Jazayeri 2023; Golestani 2022

Machine learning models (ML)

- Identify hidden relationships: covariate/outcome
- Traditional methods: relied on predefined rules.
- Automatically weight predictors: observed data pattern

Cruz 2007; Panesar 2019

- ML s: support vector machines (SVMs), & artificial neural networks (ANNs): extract complex non-linear relationships - better predictions compared to linear models.

Kuhle 2018

Methods. Data collection and preprocessing

- Incidence data: Systematic reviews

Rahimi-Movaghar 2013; Jazayeri 2015

- Covariates extracted from IRF for 2000–2015
- 207 features: road infrastructure/safety 205 countries

Methods. Data collection and preprocessing

To train our model:

- (1) Country profiles: country name, income group and region,
- (2) Road network: Length of the road network in kilometer (Km)
- (3) Road traffic: Traffic volume in million (Mio) Vehicle-Km unit.
- (4) Multimodal traffic comparisons in passenger-km unit.
- (5) Vehicles in use: total number of passenger cars, buses, motor coaches
- (6) Road accidents: No. crashes with at least one person killed or injured
- (7) Road expenditures: Expenditure on constructions/investments
- (8) CO2 emission as a measure of transport activities

Jazayeri 2024

Before applying any model to the collected data

- Clean data
- Deal with the missing values
- Select important features to prevent over-fitting
- Normalize: Convergence the model

Clean data

- All datapoints: in a single table
- 168 rows (the number of instances), 70 columns
- 68 IRF indicators, one year, country
- Another column: Incidence rates to predict
- National and sub-national incidences
- Excluded: mean values across several years

Missing data -Define a parameter

- Ranged from 0 to 1
- Max. undefined percentage still remain included
- Missing-value threshold: 0.4, best performance
- 12 out of 68 indicators were removed
- Other missing values: imputed: Python
- Mice: Multivariate Imputation by Chained Equations in R

Over-fitting: So powerful, captures noise/random fluctuations

- No. training datapoints: ~5–10x features
- 168 datapoints: 56 features: over-fitting
- Removed highly correlated features
- Computed the Pearson correlation matrix
- Hyper-parameter: empirically chosen: 0.7
- Feature selection phase: 23 remained

Normalize some features

- Total highways (Km), magnitude larger than others
- National currency per dollar
- Standard normalization: Feature: SD=1/mean=0

Models – Regression – Predict incidence

- Simple linear regression: Simple, interpretable/linear
- Support vector regression (SVR): Non-linear/less interpretable

  ϵ-SVR: more robust against noisy input

- Multi-layer perceptron (MLP): Powerful to learn;

  discard irrelevant features/black box: Why?

  MLP (3 hidden layers): 10/20/10 nodes

  ReLU: Constant value/Sigmoid Scikit-Learn package: ML

Jazayeri, 2024

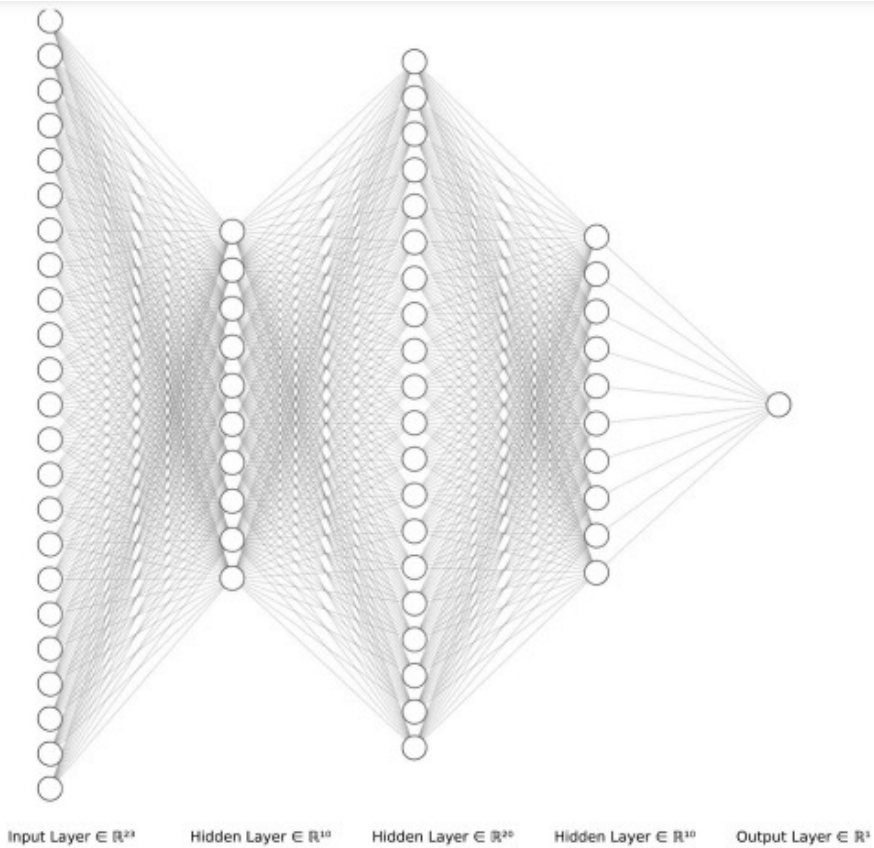

Fig. 1. Architecture of our neural network.

Three standard evaluation metrics

- Mean Absolute Error (MAE): average errors difference of predicted points/actual values

- Root Mean Square Error (RMSE): MAE similar more weight to higher errors
- R2 capture dataset variability proportion
  R2=1→ Explained data variability
  – predict actual values

Results

- MLP best metrics, train/test sets

  Cross-validation error < training error

MLP Neural network MAE: 4.66 Excellent predicted incidence rate

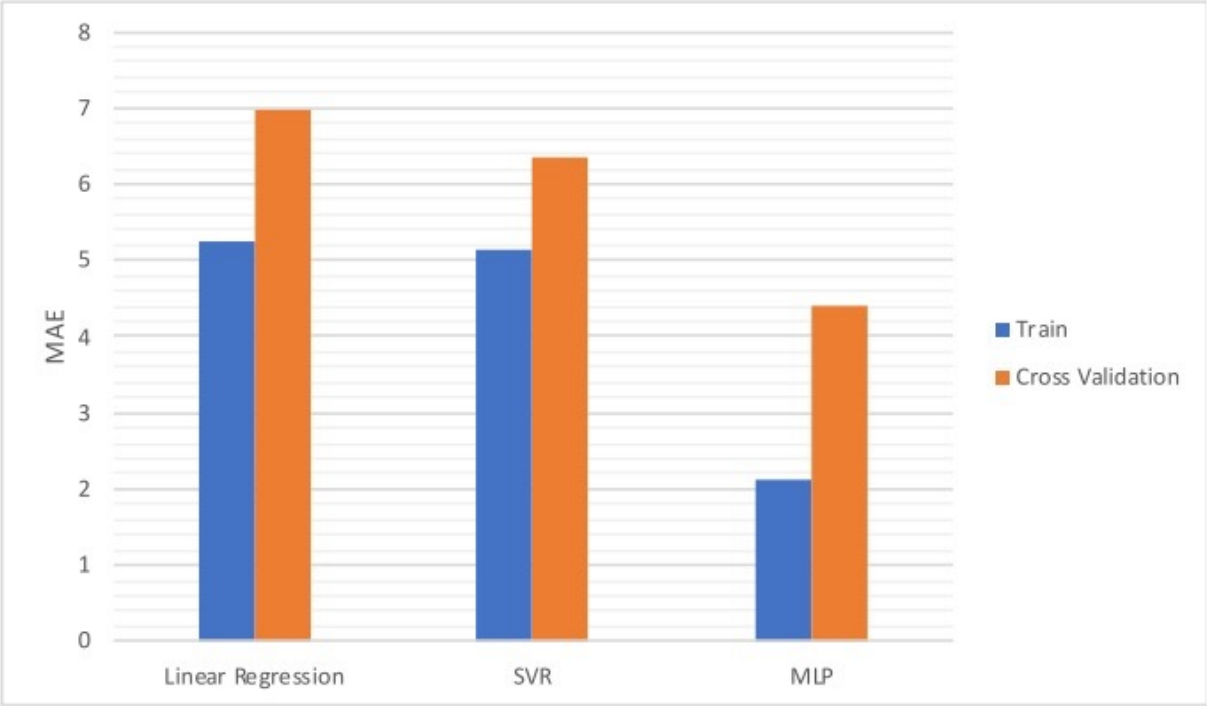

Fig. 2. Mean Absolute Error of different methods.

Jazayeri, 2024

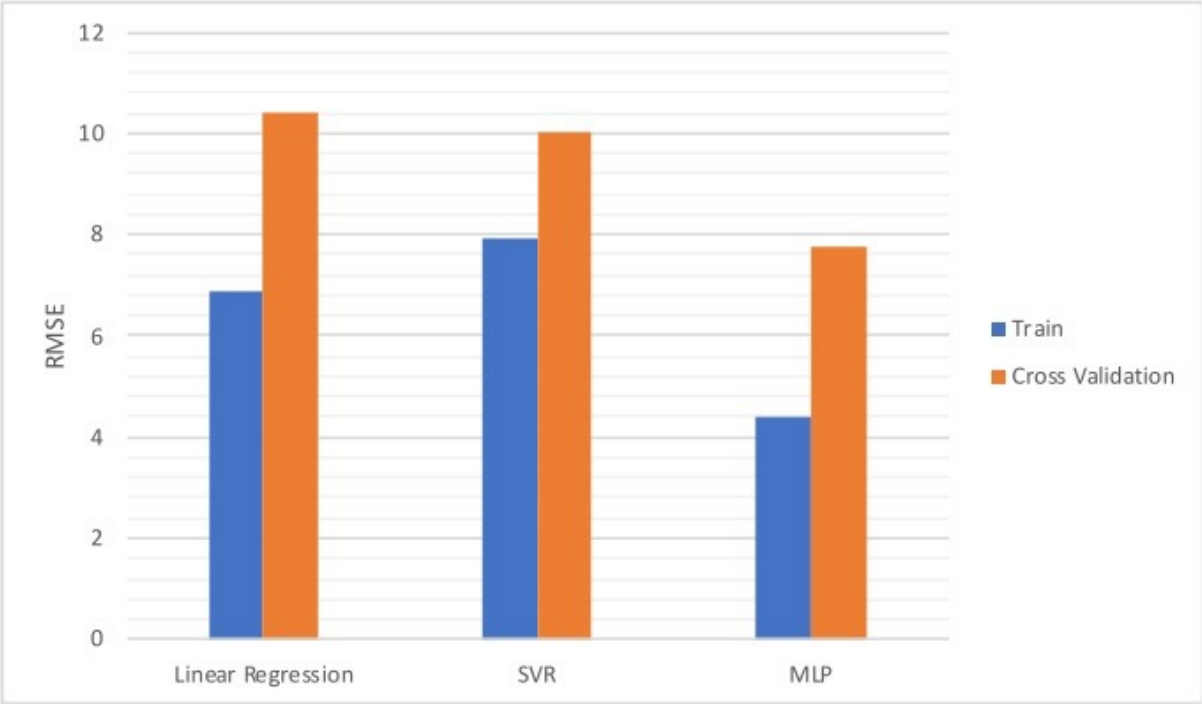

Fig. 3. Root Mean Square Error of different methods.

MLP: 0.73

All models performed better on train - Necessity of having more clean data

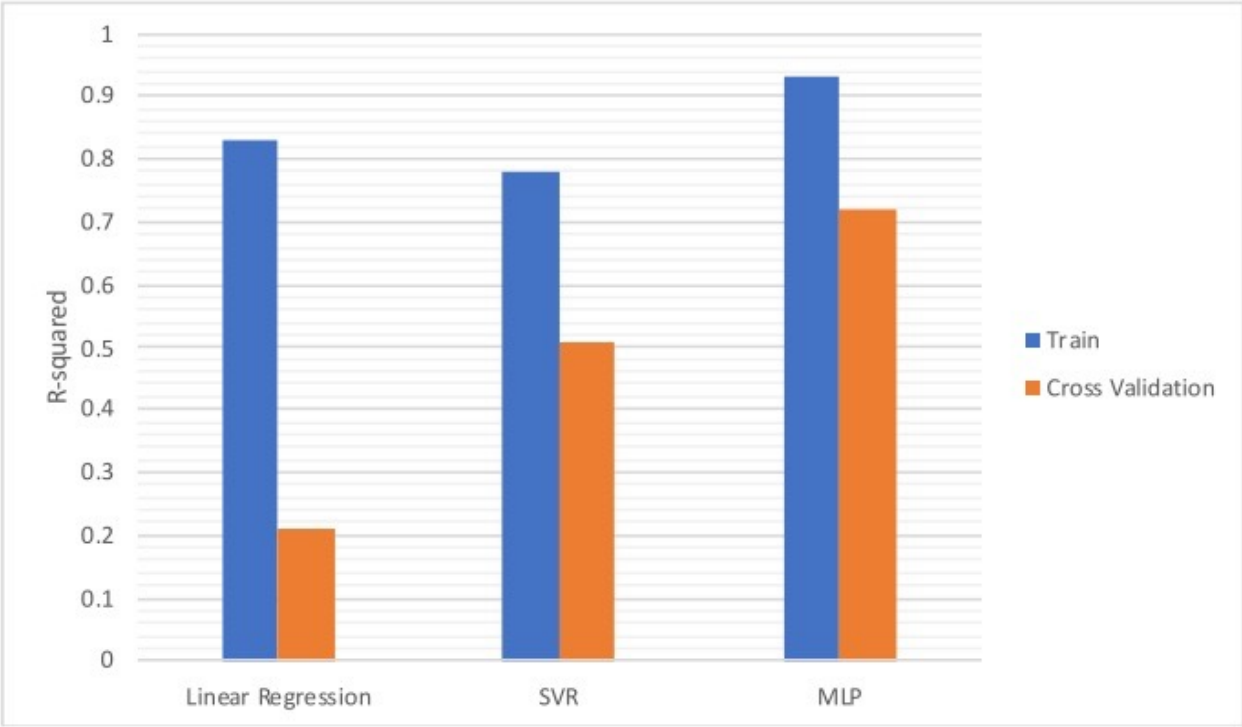

Fig. 4. R-squared of different methods.

Results evaluation

- E-SVR: 5 predictors:

  +Vehicles in use, Total vehicles/Km of roads

  +Injury accidents/100 Million Veh-Km

  +Vehicles in use, Vans, Pick-ups, Lorries, Road Tractors

  +Inland surface Passengers Transport (Mio Passenger-Km), Rail

  _ % paved

Results Outcome

cases per million (cpm) population per year)

- The highest incidence:

  Japan (114.18 cpm)

  Tajikistan (111.56 cpm)

  Bahamas (96.36 cmp) in 2000

  Monaco (172.05)

  Korea (154.79 cpm),

  Indonesia (139.45 cpm),

  The United States of America (USA) (138.62 cpm) in 2015

Results outcome

- The lowest incidence rate:

  Qatar (4.77 cpm),

  Netherlands (5.77 cpm) in 2000

  Botswana (1.09 cpm)

  Kenya (1.86 cpm)

  Saudi Arabia (2.22 cpm), in 2015

Jazayeri, 2024

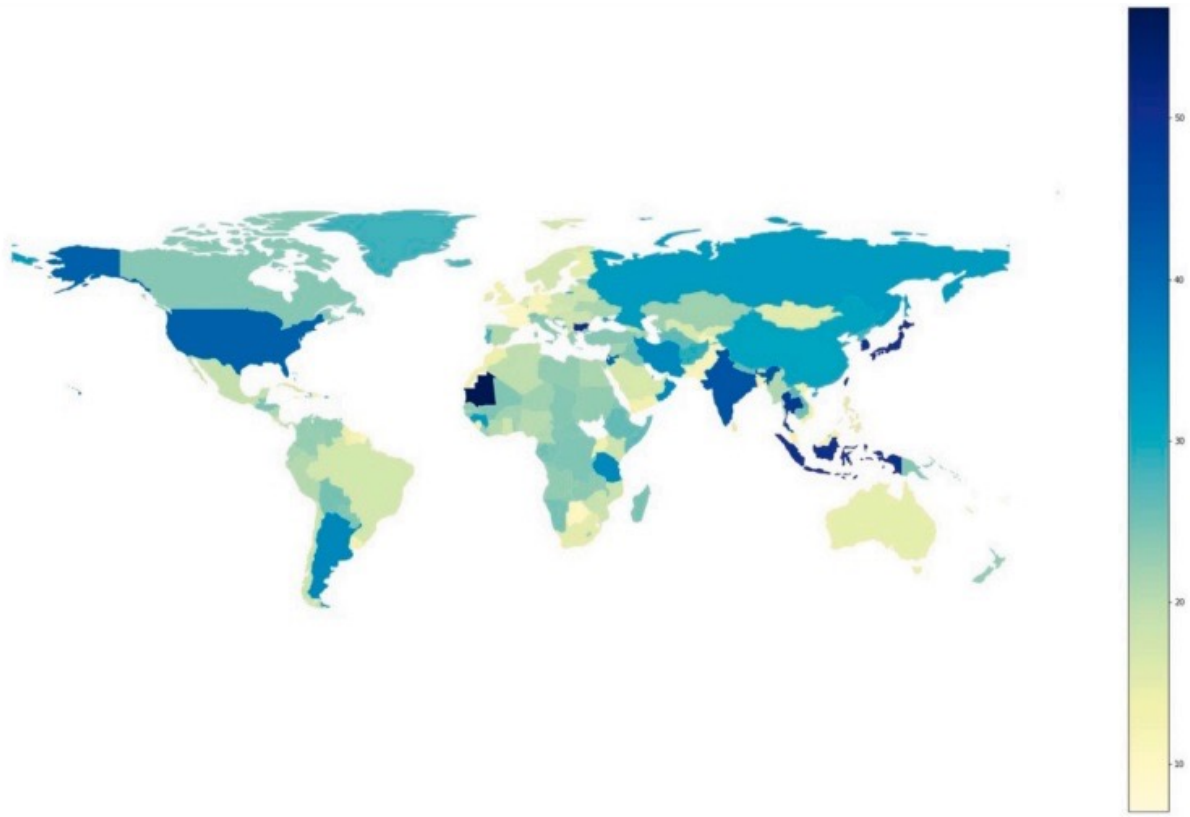

Fig. 5. World map view of average generated TSCI incidence by our model from 2000 to 2015.

Jazayeri, 2024

Table 1
List of used covariates to predict incidence.

'Motorways Km'
'Other Roads Km'
'Million Vehicle-Km, Annual, Passenger cars'
'Million Vehicle-Km, Annual, Buses & Motor coaches'
'Million Vehicle-Km, Annual, Total'
'Million Vehicle-Km, Annual, Motorcycles'
'Inland surface freight transport (Mio Tonne-Km), Rail'
'Inland surface freight transport (Mio Tonne-Km), Total'
'Inland surface Passengers Transport (Mio Passenger-Km), TOTAL'
'Inland surface Passengers Transport (Mio Passenger-Km), Rail'
'Vehicles in use, Vans, Pick-ups, Lorries, Road Tractors'
'Vehicles in use, Passengers cars/1'000 pop'
'Injury accidents'
'Vehicles in use, Motorcycles'
'Vehicles in use, Total vehicles/Km of roads'
'Passenger cars, Imports'
'Passenger cars, Exports'
'Total motor vehicles, Imports'
'Expenditures (Mio USD), Regional/Local Gov't'
'Diesel price (US $ cent/liter)'
'Diesel consumption, Kt'
'Road sector energy consumption in ktoe'
'CO2 Emissions from Inland Transport Sector (Mio Tonnes of C02), Rail'
Km: Kilometer, Kt: Kilotonnes, Mio: million, ktoe: kilotonnes of oil equivalent

Discussion - Incidence of TRTSCI

- Positive correlation:

  Number of vehicles,

  number of injuries,

  number of passengers

- Negative correlation: paved

Discussion

- Limitation

  Transport related

  Registry system –missed data: 40%

Conclusion

- Predict:

  Incidence rate of TSCI: MAE of 4.66