Deep-insight visible neural network (DI-VNN) as a responsible framework of human and machine learning:

Implementation for prognosticating prelabor rupture of membranes

Herdiantri Sufriyana;[a,b] Yu-Wei Wu;[a,c] Emily Chia-Yu Su[a,c,*]

[a] Taipei Medical University, Taiwan
[b] Universitas Nahdlatul Ulama Surabaya, Indonesia
[c] Taipei Medical University Hospital, Taiwan

*Corresponding author: emilysu@tmu.edu.tw

1. Introduction

- ✓ Many problems in medicine, e.g., premature rupture of membranes (PROM), needs prognostication and causal reasoning to develop a prevention strategy, warranted by growing implementation of insurance-based healthcare worldwide.
- ✓ While prognostication is achievable, causality cannot be inferred yet by machine learning.[1]
- ✓ These issues urge human involvement to mitigate harmful machine learning (ML) prediction with causal reasoning, i.e., estimating what may happen if the conditionals are different to what a machine learns from the previous data.
- ✓ To provide such framework, we developed deep-insight visible neural network (DI-VNN) pipeline based on recent studies.[2,3]

Objective: To develop, validate, and deploy a prognostic prediction model by DI-VNN for PROM using a nationwide health insurance database.

2. Methods

This study has been fully described elsewhere for clinician audience with addition of the time of delivery estimation.[4]

- ✓ **Study design:** Retrospectively selected visits (*n*=170,730)
- ✓ **Outcome:** PROM (*n*=23,791)
- ✓ **Candidate predictors:** Medical history (ICD-10 codes)
- ✓ **Predictive modeling:**
  Deep learning, convolutional neural network for non-image data with hierarchical architecture derived from ontology of the predictors (DI-VNN)
  Statistical ML, ridge regression based on systematic human learning and causal inference (causal RR)
  Computational ML, state-of-the-art algorithms for pregnancy outcomes[5] (PC-ENR, PC-RF, PC-GBM)
  Other models from the previous studies,[6,7] selected by conducting a systematic review in this study
- ✓ **Recalibration:** A general additive model by locally weighted scatterplot smoothing (GAM-LOESS) based on different data partition
- ✓ **Evaluation:** Four external validation sets

3. Results

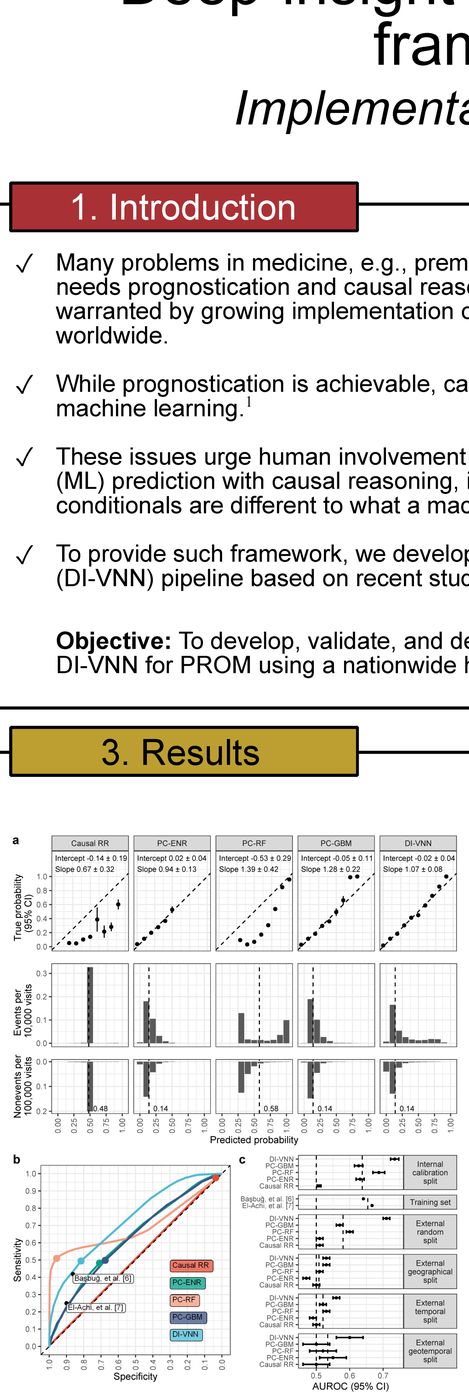

Figure 1. Model evaluation: (a) calibration; (b) receiver operating characteristics (ROC); (c) areas under ROC (AUROCs). Showing thresholds (a, b) and average AUROCs per set (c).

1 DI-VNN was the most well-calibrated (Figure 1a).

2 DI-VNN outperformed other models in this and previous studies (Figures 1b and 1c) by an external validation set (area under receiver operating characteristics curve [AUROC] 0.71, 95% CI 0.70 to 0.72), including one using a biomarker (AUROC 0.641; n=1,177).[6]

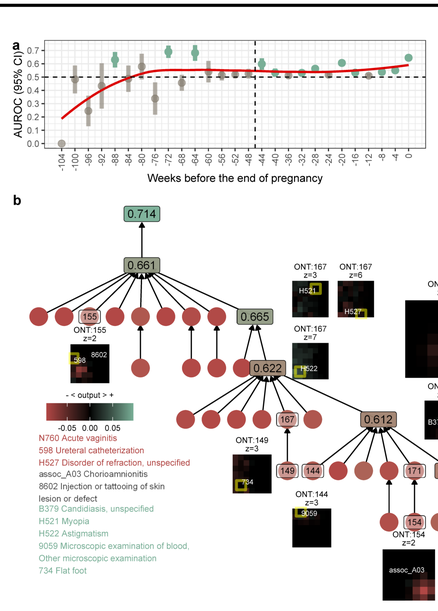

Figure 2. Exploratory data analysis of DI-VNN: (a) areas under receiver operating characteristics curve (AUROC) every 4 weeks; (b) ontology network and arrays of DI-VNN. Showing the best time window for the prediction by DI-VNN (a) and AUROCs >0.55 for prediction using parts of the network architecture up to each layer which a node resides (b). Each node is a ClixO term, prefixed by ONT. Only showing those with distinguished output arrays for a particular channel denoted by '2'. A feature in the array may tend to positive or negative output, color-coded based on the gradient as shown, including the feature description. Yellow square in an array refer to a feature if only its output is non-zero. A feature may not have this square, e.g., causal_A03 and 8602. ONT:154 is an example of a backpropagation effect from ONT:171. Methods to construct the network and array are described elsewhere.*

3 The prediction was robust from 44 ± 2 weeks before the end of pregnancy (Figure 2a).

4 A human can learn on 'subconscious mind' of the DI-VNN that distinguished signals from N760 (acute vaginitis) and causal_A03 (chorioamnionitis) at population level (Figure 2b).

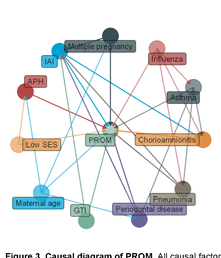

Figure 3. Causal diagram of PROM. All causal factors had been verified by inverse probability weighting (IPW) using our data. Causal inference was only conducted between PROM and each of the causal factors. Inter-causal factor relationship was not verified, but it demonstrates how a causal factor was included in the causal model of another causal factor in this figure.

5 Insights from DI-VNN were semantically matched with genital tract infection (GTI) and chorioamnionitis based on standard causal inference (Figure 3)

6 By reasoning, one may find counterintuitive features (Figure 2b), extensively discussed elsewhere.[4]

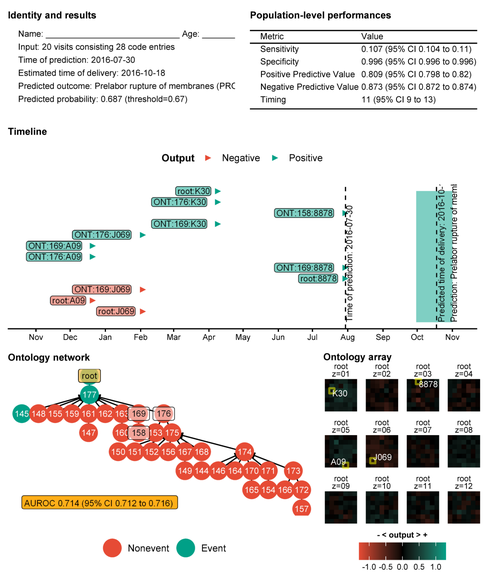

Figure 4. A report example of PROM prediction and estimation of the time of delivery. An ontology term in the timeline is prefixed by ONT, followed by the number and one of the feature members.

7 At individual level (Figure 4), we chose visits from a 19-years-old female as an example. The predicted outcome and probability were shown. We also included the time of delivery estimation by PC-RF.[4] A doctor can see the timeline of positive predictors and how these were connected in ontology network and array. Population-level performances computed from visits with the same predicted probability and estimated time of delivery were also shown and may serve as a second-line approach to mitigate optimistic bias. Local cohort can be made by a doctor in this web application to choose a local threshold. Our web application requires only diagnosis/procedure codes and dates (https://predme.app/promtime), making it possible for quick implementation in low-resource setting.

Acknowledgement

The social security administrator for health or badan penyelenggara jaminan sosial (BPJS) kesehatan in Indonesia gave permission to access the sample dataset in this study (dataset request approval no.: 5064/I.2/0421). This work was supported by the Ministry of Science and Technology (MOST) in Taiwan (grant number MOST109-2221-E-038-018) and the Higher Education Sprout Project from the Ministry of Education (MOE) in Taiwan (grant number DP2-110-21121-01-A-13) to Emily Chia-Yu Su. The sponsor had no role in the research design or contents of the manuscript for publication.

References

1. Wilkinson, J., et al. Time to reality check the promises of machine learning-powered precision medicine. Lancet Digit Health 2, e677-e680 (2020).
2. Sharma, A., Vans, E., Shigemizu, D., Boroevich, K.A. & Tsunoda, T. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. Sci Rep 9, 11399 (2019).
3. Ma, J., et al. Using deep learning to model the hierarchical structure and function of a cell. Nat Methods 15, 290-298 (2018).
4. Sufriyana, H., Wu, Y.W. & Su, E.C.Y. Human and machine learning of prognostic prediction for prelabor rupture of membranes and the time of delivery: a nationwide development, validation, and deployment using medical history. medRxiv, 2021.2006.2016.21258884 (2021).
5. Sufriyana, H., et al. Comparison of Multivariable Logistic Regression and Other Machine Learning Algorithms for Prognostic Prediction Studies in Pregnancy Care: Systematic Review and Meta-Analysis. JMIR Med Inform 8, e16503 (2020).
6. Başbuğ, D., Başbuğ, A. & Gülerman, C. Is unexplained elevated maternal serum alpha-fetoprotein still important predictor for adverse pregnancy outcome? Ginekol Pol 88, 325-330 (2017).
7. El-Achi, V., et al. First-Trimester Prediction of Preterm Prelabour Rupture of Membranes. Fetal Diagnosis and Therapy 47, 624-629 (2020).

Abbreviations

8878, diagnostic ultrasound of gravid uterus; A09, diarrhea and gastroenteritis of presumed infectious origin; APH, ante-partum hemorrhage; AUROC, area under receiver operating characteristics curve; CliXO, clique-extracted ontology; DI-VNN, deep-insight visible neural network; ENR, elastic net regression; GBM, gradient boosting machine; GTI, genital tract infection; IAI, intra-amniotic infection; J069, unspecified acute upper respiratory infection; K30, dyspepsia; PC, principal component; PROM, prelabor rupture of membranes; RF, random forest; RR, ridge regression; SES, socio-economic status.

4. Conclusion

DI-VNN allows a human to assess whether the prediction results can be safely taken into the decision case-by-case with moderate predictive performance at population level.