HHS Public Access
Author manuscript
Am Surg. Author manuscript; available in PMC 2024 January 01.

Published in final edited form as:
Am Surg. 2023 January ; 89(1): 25–30. doi:10.1177/00031348221101488.

Artificial Intelligence and Machine Learning in Prediction of Surgical Complications: Current State, Applications, and Implications

Abbas M. Hassan, MD[1], **Aashish Rajesh, MBBS**[2], **Malke Asaad, MD**[3], **Nelson A. Jonas, MD**[4], **J. Henk. Coert, MD, PhD**[5], **Babak J. Mehrara, MD**[4], **Charles E. Butler, MD**[1]

[1]Department of Plastic and Reconstructive Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[2]Department of Surgery, University of Texas Health Science Center, San Antonio, TX, USA

[3]Department of Plastic Surgery, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

[4]Department of Plastic and Reconstructive Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[5]Department of Plastic and Reconstructive Surgery, University Medical Center Utrecht, Utrecht, Netherlands

Abstract

Surgical complications pose significant challenges for surgeons, patients, and health care systems as they may result in patient distress, suboptimal outcomes, and higher health care costs. Artificial intelligence (AI)-driven models have revolutionized the field of surgery by accurately identifying patients at high risk of developing surgical complications and by overcoming several limitations associated with traditional statistics-based risk calculators. This article aims to provide an overview of AI in predicting surgical complications using common machine learning and deep learning algorithms and illustrates how this can be utilized to risk stratify patients preoperatively. This can form the basis for discussions on informed consent based on individualized patient factors in the future.

Keywords

artificial intelligence; machine learning; deep learning; surgical complications; risk assessment; calculator

---

Corresponding Author: Aashish Rajesh, MBBS, Department of Surgery, University of Texas Health Science Center at San Antonio, 1400 Pressler St., Unit 1488, Houston, TX 77030, USA., rajesha@uthscsa.edu.

Author Contributions

AH: draft of preliminary manuscript, revision, and approval of final version. AR and MA: manuscript revision, critical review of manuscript with revision for incorporating intellectual content, and approval of final version. NJ, JC, and BM: critical review of manuscript with revision for incorporating intellectual content and approval of final version. CB: senior author, critical review of manuscript with revision for incorporating intellectual content, and approval of final version.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Dr. Butler is a consultant for Allergan Inc. The remaining authors do not have any conflicts of interest to report in regards to the contents of this article.

Introduction

The number of surgical procedures performed each year in the United States has steadily increased over the past few decades, rising from 13.4 million in 1995 to 19.2 million in 2018.[1] The rate of complications resulting from surgical procedures varies depending on patient comorbidities, disease or treatment factors, and the circumstances of the surgical procedure (eg, trauma, emergency surgery, and contaminated wounds).[2,3] Complications can range from minor events that can be resolved rapidly without the need for intervention to more serious problems that can be life threatening, necessitate a return to operating room, or prolong hospital stay.[4] Therefore, surgical complications pose significant challenges for surgeons, patients, and health care systems as they may result in patient distress, suboptimal outcomes, and higher health care costs.[5–7]

Artificial intelligence algorithms have advanced surgery by helping clinicians quantify the risk of surgical procedures for a given patient based on their individual risk factors and circumstances.[8–11] AI-driven models can learn relationships and patterns between complicated variables and determine which complex combinations of patient and treatment features indicate higher or lower risk of a given complication. In this article, we review the use of AI models in predicting surgical complications. We provide surgeons with an overview of AI in predicting surgical complications using common machine learning (ML) and deep learning algorithms, and the current applications and limitations of AI in the surgical field.

Artificial Intelligence vs Traditional Risk Calculators

Predicting the likelihood of a postoperative complication accurately is critical for optimizing patient selection before surgery, directing perioperative decision-making, gauging the threshold for concern in the postoperative period, and guiding early intervention.[12] To stratify patients' postoperative morbidity and mortality risk, various risk stratification and predictive models have been developed, including the American College of Surgeons Surgical Risk Calculator (ACS-SRC), the American Society of Anesthesiologists (ASA) score, and the Physiologic and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM).[13–15]

Although the development of traditional risk assessment instruments such as the ACS-SRC, ASA score, and POSSUM has provided surgeons and patients with valuable tools for assessing the risk of complications following surgical procedures, these tools have limitations that can decrease their predictive power. Traditional risk assessment tools use Cox proportional hazards regression analysis or logistic regression models that may overestimate or underestimate risks because in many cases these models rely on variables that have statistically significant effects on outcomes. Such approaches may, therefore, exclude subtle, but important, factors that may also influence outcomes. In addition, traditional models often assume a linear correlation between variables and outcomes and may not account for more complex interactions, particularly at the extremes of variable ranges (Figure 1).[16] Furthermore, the overfitting and multicollinearity limitations of regression analysis preclude the examination of a large number of variables. Therefore, current prediction models are often limited to a relatively small number of variables and may exclude other important modulators of outcome.[17,18]

AI techniques often have better predictive performance than do traditional statistical models limited to logistic regression.[17,19–21] ML algorithms allow for the evaluation of a higher number of clinical variables than do traditional modeling approaches and may help identify weak predictors or interactions between variables that may improve prediction accuracy.[17,19] By developing nonlinear models that use multiple data sources, such as diagnoses, treatments, and laboratory values, ML has outperformed logistic regression for predicting postoperative outcomes.[22–25] Recent work from our group, for example, showed that, compared with multivariable logistic regression, ML demonstrated higher predictive discriminatory performance and identified more predictors of complications in both abdominal wall reconstruction and reconstruction following mastectomy.[20,21] And in contrast to conventional statistical approaches, incremental learning enables ML to improve continuously as new data are added.[26,27] Thus, unlike traditional risk calculators, which are static, ML models are dynamic and continuously improve over time.

Applications in Surgery

Hepatobiliary and Colorectal Surgery

Merath et al[12] developed a model to predict complications of patients undergoing hepatic, pancreatic, and colorectal surgery using the National Surgical Quality Improvement Program (NSQIP) database. The model was trained using 15,657 patients and achieved AUCs ranging from .76 for prediction of surgical site infections to .98 for prediction of stroke. The researchers also demonstrated that the ML models outperformed the ASA and ACS-SRC. Other researchers have used deep learning models to predicted complications for patients with locally advanced or recurrent colorectal cancer who undergo pelvic exenteration.[28] In that study, an ANN model was developed using 1,147 patients and achieved AUCs ranging from .61 to .79. The authors concluded that their deep learning model was better than logistic regression at predicting an outcome from a complex combination of patient- and procedure-related variables.

Cardiothoracic Surgery

Lapp et al[29] developed ML models to predict severe postoperative complications in patients who underwent coronary artery bypass graft and/or valve surgery. Using data from 3,700 patients, their random forest model, with an AUC of .72, achieved better performance than other models. Subsequently, Salati et al[30] developed an ML model to predict cardiopulmonary complications in patients undergoing lung resection. Their extreme gradient boosting model, trained on 1,360 patients, achieved an AUC of .75 and an accuracy of 70%, outperforming other models. The authors concluded that the ML models provide personalized predictions by analyzing the characteristics that were available for each patient and support surgical decision-making by suggesting individualized postoperative care, selecting preoperative regimens for patients at high risk of complications, and evaluating the quality of care.

Breast Cancer Surgery

Recently, our team developed ML models to predict skin flap necrosis in patients undergoing mastectomy for breast cancer.[20] Using data from 694 patients, we evaluated the predictive performance of nine ML algorithms. The algorithms, which were trained on readily available perioperative data, predicted individual patients' risk of mastectomy skin flap necrosis with an AUC of .70 and 89% accuracy. Additionally, by performing PFI and ALE analysis, we identified risk factors associated with this complication. In another recent study,[31] our team used ML models to predict sentinel lymph node status in elderly patients with primary breast cancer. Using data from 1,706 consecutive patients, a support vector machine was developed to preoperatively predict node positivity using patients' demographics, tumor stage, genetic profile, and imaging data. The model identified these patients' sentinel lymph node status with high accuracy (accuracy, 84%; 95% CI: 80–88%) and predictive performance (AUC, .70; 95% CI: .62-.77). Additionally, analysis of the model helped reveal factors associated with lymph node positivity, such as disease stage, younger age, family history of breast cancer, margin status, and estrogen and progesterone receptor positivity. This model can help patients understand their risk for node-positive disease, which may influence recommendations regarding surgery and adjuvant therapy.

Plastic and Reconstructive Surgery

ML models have been developed to predict complications after implant-based breast reconstruction.[32] Using perioperative data from 481 patients, we trained ML models to predict periprosthetic infection and the need for device explantation. We found that the ML models demonstrated high performance in predicting infection (AUC, .73; accuracy, 83%) and the need for device explantation (AUC, .78; accuracy, 84%). Furthermore, ML models out-performed traditional multivariable logistic regression in identifying contributing risk factors such device placement plane, type of acellular dermal matrix used, and adjuvant therapy. Using the same data, for example, multivariable logistic regression found two predictors of infection, whereas ML identified nine. These algorithms can help surgeons make informed decisions and provide an objective and accurate metric to use when counseling patients about prospective reconstructive alternatives and consequences. The models can also assist a patient in the informed consent process by predicting the risks and benefits of a particular procedure and identifying modifiable variables that can be addressed before reconstruction to optimize the patient's suitability for the procedure.

Neurological Surgery

Niftrik et al[33] developed an extreme gradient boosting model to predict complications for patients undergoing intracranial tumor surgery. Using data from 668 patients, the model had an accuracy of 70% and AUC of .74, and it overperformed a conventional statistical model in predictive power. Subsequent work by Farrokhi et al used ML algorithms to predict complications after deep brain stimulation surgery. The supervised models demonstrated high discriminatory performance in predicting any complication (AUC, .86), a complication within 12 months (AUC, .91), need to perform a second surgical procedure (AUC, .88), and infection (AUC, .97). The authors concluded that ML can be utilized to improve risk assessment, preoperative informed consent, and treatment planning for neurosurgery patients.[34]

Orthopedic Surgery

Kim et al[35] leveraged the NSQIP database to develop an ML model to predict complications following posterior lumbar spine fusion. They developed an ANN model using 22,629 patients to predict cardiac complications, wound complications, venous thromboembolism, and death. The ANN model achieved an AUC of .71 and outperformed benchmark ASA scores. Subsequently, Devana et al[36] used data from 156,750 patients to develop an ensemble of ML models to predict complications after total knee arthroplasty. The models demonstrated good discriminative performance with an AUC of .68. The authors concluded that the ensemble was useful for pre-operative counseling, shared decision-making, informed consent, and risk adjustment reimbursement programs and for managing postoperative expectations for both patients and surgeons.

General Surgery

Recently, our team developed ML models to predict outcomes of abdominal wall reconstruction.[21] Using data from 725 patients, we developed an ensemble (using multiple ML algorithms to obtain better predictive performance) of 9 supervised ML models that used the majority rule to predict hernia recurrence, surgical site occurrences (SSOs), and readmissions within 30 days. The ML models achieved high predictive performance in identifying over a long follow-up period (mean, 3 years) complications such hernia recurrence (accuracy, 85%; AUC, .71) as well as short-term complications such as SSOs (accuracy, 72%; AUC, .75) and 30-day readmission (accuracy, 84%; AUC, .73). Furthermore, model analysis assisted in identifying factors associated with poor outcomes that were masked in traditional statistical approaches such as logistic regression. These factors included surgical techniques, prior abdominal surgeries, and wound contamination level. Using the same database, multivariable logistic regression identified five predictors of SSOs, whereas ML identified 12 predictors. The information provided by ML models can therefore optimize surgical planning, preoperative optimization, and shared decision-making.

Risk Calculators

ML can produce updated risk assessments, enabling clinicians to assess in real time patients' risk and whether their condition has been optimized for surgery. Currently, two validated ML-driven risk calculators have been developed to predict major surgical complications.

MySurgeryRisk was developed and validated using data from 51,457 patients who had undergone major inpatient surgery to predict eight major postoperative complications (acute kidney injury, sepsis, venous thromboembolism, admission to intensive care after 48 h, mechanical ventilation after 48 h, wound, and neurologic and cardiovascular complications) as well as death within 24 months following surgery.[37] The model achieved an AUC ranging from .77 to .94. Subsequently, Brennan et al[38] compared the My-SurgeryRisk calculator's usability and accuracy with that of physicians' clinical judgment. They found that My-SurgeryRisk outperformed physicians' initial risk assessments for almost all postoperative complications. Additionally, the physicians' risk assessments improved significantly after they interacted with the ML model.

Another group of investigators developed the Predictive opTimal Trees in Emergency Surgery (POTTER) risk calculator.[39] The model was based on a decision tree ML algorithm developed using data from 382,960 patients in the NSQIP database. POTTER achieved high discriminatory performance in predicting morbidity (AUC, .84) and mortality (AUC, .92) and outperformed the ASA calculator, Emergency Surgery Score, and NSQIP risk calculator.

Limitations of AI Models

The power of AI prediction is dependent on the accuracy and comprehensiveness of the input data. Biases in clinical data collection can have an impact on the types of patterns AI recognizes and the predictions it makes.[40] Additionally, most AI models were developed using registry-based data such as NSQIP, posing several inherent limitations such as duration of follow-up, limiting the predictive accuracy to only short-term outcomes. Furthermore, while AI can reveal subtle patterns and risk factors that are masked in traditional risk modeling, failing to adhere to best practices in model development can result in poor outcomes and lower model performance. Finally, most studies using AI in surgery have been observational.

Despite the limitations of AI, implementation and analysis of ML algorithms are critical to understanding the various variables that predict real-life surgical outcomes. However, there remain significant challenges and risks associated with implementing AI in surgery, such as reliability, transparency, accountability, liability, data privacy and security, efficacy, structural inequality, workforce substitution, and ethical concerns. Therefore, an integrated governance framework is required for the development, implementation, and adoption of AI in surgery.

Conclusion

AI-driven predictive models trained using readily available clinical data can predict surgical complications with variable rates of precision and have been steadily improving. By providing patient-specific risk assessment and guiding perioperative shared decision-making, preoperative patient optimization, and surgical planning, these models can improve surgical outcomes.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. McDermott KW, Liang L. Overview of operating room procedures during inpatient stays in U.S. hospitals, 2018: Statistical brief #281. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs Rockville, MD: Agency for Healthcare Research and Quality (US); 2006.
2. Mayo NE, Feldman L, Scott S, et al. Impact of preoperative change in physical function on postoperative recovery: Argument supporting prehabilitation for colorectal surgery. Surgery 2011;150(3):505–514. [PubMed: 21878237]

3. Viste A, Haùgstvedt T, Eide GE, Søreide O. Postoperative complications and mortality after surgery for gastric cancer. Ann Surg 1988;207(1):7–13. [PubMed: 3337564]

4. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: A new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg 2004;240(2):205–213. [PubMed: 15273542]

5. Lemaine V, Schilz SR, Van Houten HK, Zhu L, Habermann EB, Boughey JC. Autologous breast reconstruction vs implant-based reconstruction: How do long-term costs and health care use compare? Plast Reconstr Surg 2020;145(2): 303–311. [PubMed: 31985608]

6. Aliu O, Zhong L, Chetta MD, et al. Comparing health care resource use between implant and autologous reconstruction of the irradiated breast: A national claims-based assessment. Plast Reconstr Surg 2017;139(6):1224e–1231e.

7. Pinto A, Faiz O, Davis R, Almoudaris A, Vincent C. Surgical complications and their impact on patients' psychosocial well-being: A systematic review and meta-analysis. BMJ Open 2016;6(2):e007224.

8. Kuo PJ, Wu SC, Chien PC, et al. Artificial neural network approach to predict surgical site infection after free-flap reconstruction in patients receiving surgery for head and neck cancer. Oncotarget 2018;9(17): 13768–13782. [PubMed: 29568393]

9. Mantelakis A, Assael Y, Sorooshian P, Khajuria A. Machine learning demonstrates high accuracy for disease diagnosis and prognosis in plastic surgery. Plast Reconstr Surg Glob Open 2021;9(6):e3638. [PubMed: 34235035]

10. Buchlak QD, Esmaili N, Leveque JC, et al. Machine learning applications to clinical decision support in neurosurgery: An artificial intelligence augmented systematic review. Neurosurg Rev 2020;43(5):1235–1253. [PubMed: 31422572]

11. Cirillo MD, Mirdell R, Sjöberg F, Pham TD. Time-independent prediction of burn depth using deep convolutional neural networks. J Burn Care Res 2019;40(6): 857–863. [PubMed: 31187119]

12. Merath K, Hyer JM, Mehta R, et al. Use of machine learning for prediction of patient risk of postoperative complications after liver, pancreatic, and colorectal surgery. J Gastrointest Surg 2020;24(8):1843–1851. [PubMed: 31385172]

13. Copeland GP, Jones D, Walters M. POSSUM: A scoring system for surgical audit. Br J Surg 1991;78(3):355–360. [PubMed: 2021856]

14. Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ. An Apgar score for surgery. J Am Coll Surg 2007;204(2):201–208. [PubMed: 17254923]

15. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: A decision aid and informed consent tool for patients and surgeons. J Am Coll Surg 2013;217(5):833–842. [PubMed: 24055383]

16. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. N Engl J Med 2017;376(26):2507–2509. [PubMed: 28657867]

17. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inf 2007;2: 59–77.

18. Miller RA, Pople HE, Myers JD. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med 1982;307(8):468–476. [PubMed: 7048091]

19. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: Promises and perils. Ann Surg 2018;268(1):70–76. [PubMed: 29389679]

20. Hassan AM, Biaggi AP, Asaad M, et al. Development and assessment of machine learning models for individualized risk assessment of mastectomy skin flap necrosis. Ann Surg 2022; Published ahead of print. doi: 10.1097SLA.0000000000005386

21. Hassan A, Lu S, Asaad M, Offodile AC, Sidey-Gibbons C. Novel machine learning approach for prediction of hernia recurrence, surgical complications, and 30-day readmission following abdominal wall reconstruction. J Am Coll Surg 2022;234(5):918–927. [PubMed: 35426406]

22. Soguero-Ruiz C, Fei WM, Jenssen R, et al. Data-driven temporal prediction of surgical site infection. AMIA Annu Symp Proc 2015;2015:1164–1173. [PubMed: 26958256]

23. Lee CK, Hofer I, Gabel E, Baldi P, Cannesson M. Development and validation of a deep neural network model for prediction of postoperative in-hospital mortality. Anesthesiology 2018;129(4):649–662. [PubMed: 29664888]

24. Fei Y, Hu J, Li WQ, Wang W, Zong GQ. Artificial neural networks predict the incidence of portosplenomesenteric venous thrombosis in patients with acute pancreatitis. J Thromb Haemostasis 2017;15(3):439–445. [PubMed: 27960048]
25. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: A local big data-driven, machine learning approach. Acad Emerg Med 2016;23(3):269–278. [PubMed: 26679719]
26. Silver DL. Machine lifelong learning: Challenges and benefits for artificial general intelligence Paper presented at: International conference on artificial general intelligence; 2011.
27. Gepperth A, Hammer B. Incremental learning algorithms and applications Paper presented at: European symposium on artificial neural networks (ESANN); 2016.
28. Collaborative P Predicting outcomes of pelvic exenteration using machine learning. Colorectal Dis 2020;22(12): 1933–1940. [PubMed: 32627312]
29. Lapp L, Young D, Kavanagh K, Bouamrane M-M, Schraag S. Using machine learning for predicting severe postoperative complications after cardiac surgery. J Cardiothorac Vasc Anesth 2018;32:S84–S85.
30. Salati M, Migliorelli L, Moccia S, et al. A machine learning approach for postoperative outcome prediction: Surgical data science application in a thoracic surgery setting. World J Surg 2021;45(5): 1585–1594. [PubMed: 33594578]
31. Hassan A, Tamirisa N, Singh P, Offodile AC, Butler CE. A novel support vector machine to predict sentinel lymph node status in elderly patients with breast cancer. J Clin Oncol 2022;34.
32. Hassan AM, Asaad M, Morris N, Liu J, Selber JC, Butler CE. Artificial intelligence modeling to predict periprosthetic infection and explantation following implant-based reconstruction. Plast Reconstr Surg 2022.
33. van Niftrik CHB, van der Wouden F, Staartjes VE, et al. Machine learning algorithm identifies patients at high risk for early complications after intracranial tumor surgery: Registry-based cohort study. Neurosurgery 2019;85(4): E756–E764. [PubMed: 31149726]
34. Farrokhi F, Buchlak QD, Sikora M, et al. Investigating risk factors and predicting complications in deep brain stimulation surgery with machine learning algorithms. World Neurosurg 2020;134:e325–e338. [PubMed: 31634625]
35. Kim JS, Merrill RK, Arvind V, et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. Spine 2018;43(12):853–860. [PubMed: 29016439]
36. Devana SK, Shah AA, Lee C, Roney AR, van der Schaar M, SooHoo NF. A novel, potentially universal machine learning algorithm to predict complications in total knee arthroplasty. Arthroplast Today 2021;10:135–143. [PubMed: 34401416]
37. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. My-surgeryrisk: Development and validation of a machine-learning risk algorithm for major complications and death after surgery. Ann Surg 2019;269(4):652–662. [PubMed: 29489489]
38. Brennan M, Puri S, Ozrazgat-Baslanti T, et al. Comparing clinical judgment with the Mysurgeryrisk algorithm for preoperative risk assessment: A pilot usability study. Surgery 2019;165(5):1035–1045. [PubMed: 30792011]
39. Bertsimas D, Dunn J, Velmahos GC, Kaafarani HMA. Surgical risk is not linear: derivation and validation of a novel, user-friendly, and machine-learning-based predictive optimal trees in emergency surgery risk (POTTER) calculator. Ann Surg 2018;268(4):574–583. [PubMed: 30124479]
40. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: Race-sex-and age-based disparities. JAMA 2004;291(22):2720–2726. [PubMed: 15187053]

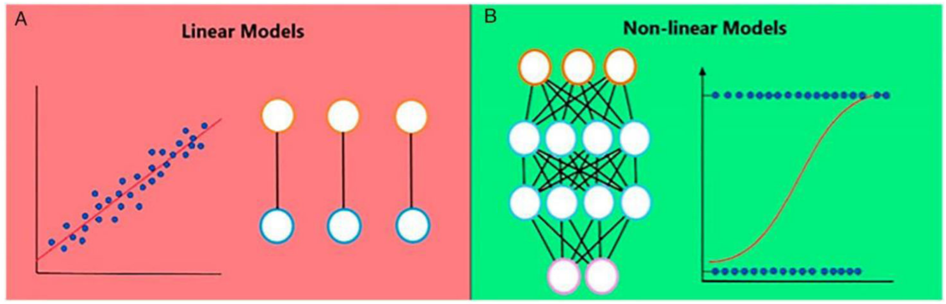

Figure 1.

Linear vs nonlinear models. (A) Linear models assume that variables interact in a linear and additive manner and therefore over- or underestimate risks at the extreme ranges of variables. (B) Nonlinear models assume that the interaction of patient demographics, comorbidities, and surgical factors is far from linear and that certain variables gain or lose relevance as a function of the presence or absence of other variables, better representing the interaction of risk factors in real life.