Supplementary Material

Explainable Artificial Intelligence Models Using Real-World Electronic Health Record Data: a Systematic Scoping Review

Seyedeh Neelufar Payrovnaziri[1], Zhaoyi Chen[2], Pablo Rengifo-Moreno[3,4], Tim Miller[5], Jiang Bian[2], Jonathan H. Chen[6,7], Xiuwen Liu[8], and Zhe He[1,*]

[1]School of Information, Florida State University, Tallahassee, Florida, USA

[2]Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, Florida, USA

[3]College of Medicine, Florida State University, Tallahassee, Florida, USA

[4]Tallahassee Memorial Hospital, Tallahassee, Florida, USA

[5]School of Computing and Information Systems, The University of Melbourne, Melbourne, Victoria, Australia

[6] Center for Biomedical Informatics Research, Department of Medicine, Stanford University, Stanford, California, USA

[7] Division of Hospital Medicine, Department of Medicine, Stanford University, Stanford, California, USA

[8]Department of Computer Science, Florida State University, Tallahassee, Florida, USA

Appendix Table 1. Inclusion and exclusion criteria for the systematic review

Type	Criteria	Rationale
Inclusion	XAI in medicine and healthcare	XAI can be applied to many different applications, but the focus of our review is in medicine and healthcare, since this area is one of the most critical applications of AI and XAI.
	The AI model is built based on EHR data	In this review, we focus on predictive modeling using EHR data.
	They study is not older than 2009	We are interested in XAI in medicine in the past decade. Also, emergence of new generation of complex AI models like deep learning, makes the XAI challenge more significant.
	The study's focus is on XAI, thus a relevant term is used in the title and/or abstract	The goal of the study is on XAI and the relevant terms include: explainable, explainability, interpretable, interpretability, understandable, understandability, comprehensible, comprehensibility, intelligible, machine learning, artificial intelligence, prediction model, predictive model, deep learning, AI, neural network
Exclusion	Details of the paper is not available	Abstract papers, or the papers that could not be accessed through the university library or the interlibrary loan, and system demonstrations are not included.
	Unpublished	We excluded papers that were uploaded on arXiv or other archiving systems but have not been published in a peer-reviewed venue.
	Opinion or other review papers	This systematic review is not a review of reviews, also opinion papers do not fulfill the requirement of delivering an XAI method.
	Duplicated	Usually querying multiple databases returns similar papers, thus we removed the duplicates.

Appendix Table 2. Reproducibility analysis of the articles. The source codes are either available and explicitly referred to in the article (marked as "✓") or not (marked as "-"). The datasets are either proprietary (marked as "-") or not (marked as "✓"). If the provided link to the dataset and/or the source code was broken or there was no dataset and/or source code available in the referred link to the host website, we also marked it as not available "–".

Article	Source code is accessible	Source code repository	Dataset is accessible	Dataset
Das et.al, 2019 [1]	✓	GitHub	✓	ADNI dataset
Knijnenburg et al., 2016 [2]	✓	GitHub	✓	Data provided in the supplementary material of the article, the validation dataset: drug sensitivity dataset from the Cancer Therapeutics Response Portal
Xiao et al., 2018 [3]	✓	GitHub	-	A real-world EHR repository of Congestive Heart Failure cohort and simulated EHR data
Choi et al., 2016 [4]	✓	GitHub	-	Sutter Health EHR data
Hu et al., 2019 [5]	✓	GitHub	✓	Retrovirus Integration Database
Hao et al., 2018 [6]	✓	GitHub	✓	The Cancer Genome Atlas (TCGA)
Bernardini et al., 2019 [7]	✓	Code Ocean	✓	Federazione Italiana Medici di Medicina Generale (FIMMG) dataset[1]
Zhang et al., 2015 [8]	✓	Downloadable from the Journal's website	✓	SIDER database
Aditya and Pande, 2017 [9]	✓	Downloadable from the Journal's website	✓	Open Access Series of Imaging Studies (OASIS)
Fejza et al.,2018 [10]	-	N/A	✓	EHR data from the Premier healthcare database
Yoon et al., 2018 [11]	-	Journal website[2]	✓	United Network for Organ Sharing (UNOS) database

Jalali and Pfeifer, 2016 [12]	✓	GitHub	✓	The Cancer Genome Atlas (TCGA)
Kaji et al., 2019 [13]	✓	GitHub	✓	MIMIC-III
Brisimi et al., 2018 [14]	-	N/A	-	Made available by the Boston Medical Center
Eck et al., 2017 [15]	✓	Sourceforge	✓	Data generated by IS-pro technique, made available on SourceForge
Settouti et al., 2012 [16]	-	N/A	✓	Pima Indian Dataset in UCI repository
Davoodi and Moradi, 2018 [17]	-	N/A	✓	MIMIC-III, Diabetes (PID), Heart Disease (HD) and Liver datasets from UCI repository
Kwon et al., 2018 [18]	-	GitHub	✓	Health Insurance Review and Assessment Service (HIRA) National Patients Sample
Shickel et al., 2019 [19]	-	N/A	✓	MIMIC-III, University of Florida Health Hospital EHR data
Pan et al., 2019 [20]	-	N/A	-	Endocrinology Department at Guangzhou Women and Children's Medical Center EHR data
Huang, 2009 [21]	-	N/A	-	Research data not available
Van den Bulcke et al., 2011 [22]	-	N/A	-	Data from a systematic screening for newborns by the PCMA screening center (Belgium)
Bouktif et al., 2014 [23]	✓	Institutional website	✓	Three heart disease datasets in UCI repository

Ponce and Martinez-Villaseñor, 2017 [24]	-	N/A	✓	Wisconsin Breast Cancer Dataset in UCI repository
Crielaard and Papapetrou, 2018 [25]	-	N/A	-	Stockholm electronic patient record (EPR) corpus (HealthBank)
Ming et al., 2019 [26]	✓	GitHub	✓	Wisconsin Breast Cancer Dataset and Pima Indian Diabetes Dataset in UCI repository
Hajiloo et al., 2013 [27]	-	N/A	✓	Prostate cancer dataset and colon cancer dataset in other publications
Luo, 2016 [28]	-	N/A	-	Practice Fusion Diabetes Classification
Zhang et al., 2018 [29]	✓	GitHub	-	De-identified EHR data from the University of Virginia Health System
Hartono, 2018 [30]	-	N/A	-	Cancer microarray data sets obtained from Gene Expression Model Selector[3]
Zhao and Bolouri, 2016 [31]	-	N/A	✓	The Cancer Genome Atlas (TCGA)
Kim et al., 2016 [32]	-	N/A	✓	NIH Epigenome Roadmap, NHGRI, ENCODE
Valdes et al., 2016 [33]	✓	Dedicated website (https://www.mediboostml.com/)	✓	13 data sets corresponding to all binary classification problems in the field of life sciences within the UCI Repository
Stiglic et al., 2012 [34]	✓	Open source Weka package	✓	40 datasets in the UCI Repository, eSol database, 9 Gene Expression Machine Learning Repository (GEMLeR) datasets

Zuallaert et al., 2018 [35]	✓	Project website	-	Datasets in other publications
Ghafouri-Fard et al., 2019 [36]	-	N/A	-	Proprietary genotyping data
Paredes et al., 2018 [37]	-	N/A	-	Portuguese dataset of ACS patients
Barakat et al., 2010 [38]	-	N/A	-	A diabetes dataset in Oman
Park et al., 2018 [39]	-	N/A	-	Seoul National University Bundang Hospital EHR data
Ge et al., 2018 [40]	-	N/A	-	Asan Medical Center EHR data
Lakkaraju et al., 2016 [41]	-	N/A	-	Medical diagnosis records of about 150K patients collected by a web-based EHR company and two other non-medical datasets
Che et al., 2017 [42]	-	N/A	-	Pediatric ICU dataset collected at the Children's Hospital Los Angeles.

[1] http://vrai.dii.univpm.it/content/fimmg-dataset
[2] Only the pseudo code is available
[3] No longer available

References

[1] D. Das, J. Ito, T. Kadowaki, and K. Tsuda, "An interpretable machine learning model for diagnosis of Alzheimer's disease," *PeerJ*, vol. 7, p. e6543, Mar. 2019, doi: 10.7717/peerj.6543.

[2] T. A. Knijnenburg *et al.*, "Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy," *Scientific Reports*, vol. 6, p. 36812, Nov. 2016, doi: 10.1038/srep36812.

[3] C. Xiao, T. Ma, A. B. Dieng, D. M. Blei, and F. Wang, "Readmission prediction via deep contextual embedding of clinical concepts," *PLOS ONE*, vol. 13, no. 4, p. e0195024, Apr. 2018, doi: 10.1371/journal.pone.0195024.

[4] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism,"

presented at the Advances in Neural Information Processing Systems, 2016, pp. 3504–3512.
[5] H. Hu *et al.*, "DeepHINT: understanding HIV-1 integration via deep learning with attention," *Bioinformatics*, vol. 35, no. 10, pp. 1660–1667, May 2019, doi: 10.1093/bioinformatics/bty842.
[6] J. Hao, Y. Kim, T.-K. Kim, and M. Kang, "PASNet: pathway-associated sparse deep neural network for prognosis prediction from high-throughput data," *BMC Bioinformatics*, vol. 19, no. 1, p. 510, Dec. 2018, doi: 10.1186/s12859-018-2500-z.
[7] M. Bernardini, L. Romeo, P. Misericordia, and E. Frontoni, "Discovering the Type 2 Diabetes in Electronic Health Records using the Sparse Balanced Support Vector Machine," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2019, doi: 10.1109/JBHI.2019.2899218.
[8] W. Zhang, F. Liu, L. Luo, and J. Zhang, "Predicting drug side effects by multi-label learning and ensemble learning," *BMC Bioinformatics*, vol. 16, no. 1, p. 365, Nov. 2015, doi: 10.1186/s12859-015-0774-y.
[9] C. R. Aditya and M. B. S. Pande, "Devising an interpretable calibrated scale to quantitatively assess the dementia stage of subjects with alzheimer's disease: A machine learning approach," *Informatics in Medicine Unlocked*, vol. 6, pp. 28–35, Jan. 2017, doi: 10.1016/j.imu.2016.12.004.
[10] A. Fejza, P. Genevès, N. Layaïda, and J. Bosson, "Scalable and Interpretable Predictive Models for Electronic Health Records," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 2018, pp. 341–350, doi: 10.1109/DSAA.2018.00045.
[11] J. Yoon, W. R. Zame, A. Banerjee, M. Cadeiras, A. M. Alaa, and M. van der Schaar, "Personalized survival predictions via Trees of Predictors: An application to cardiac transplantation," *PLOS ONE*, vol. 13, no. 3, p. e0194985, Mar. 2018, doi: 10.1371/journal.pone.0194985.
[12] A. Jalali and N. Pfeifer, "Interpretable per case weighted ensemble method for cancer associations," *BMC Genomics*, vol. 17, p. 501, 19 2016, doi: 10.1186/s12864-016-2647-9.
[13] D. A. Kaji *et al.*, "An attention based deep learning model of clinical events in the intensive care unit," *PLoS ONE*, vol. 14, no. 2, p. e0211057, 2019, doi: 10.1371/journal.pone.0211057.
[14] T. S. Brisimi, T. Xu, T. Wang, W. Dai, W. G. Adams, and I. C. Paschalidis, "Predicting Chronic Disease Hospitalizations from Electronic Health Records: An Interpretable Classification Approach," *Proceedings of the IEEE*, vol. 106, no. 4, pp. 690–707, Apr. 2018, doi: 10.1109/JPROC.2017.2789319.
[15] A. Eck *et al.*, "Interpretation of microbiota-based diagnostics by explaining individual classifier decisions," *BMC Bioinformatics*, vol. 18, no. 1, p. 441, Oct. 2017, doi: 10.1186/s12859-017-1843-1.
[16] N. Settouti, M. A. Chikh, and M. Saidi, "Generating fuzzy rules for constructing interpretable classifier of diabetes disease," *Australas Phys Eng Sci Med*, vol. 35, no. 3, pp. 257–270, Sep. 2012, doi: 10.1007/s13246-012-0155-z.
[17] R. Davoodi and M. H. Moradi, "Mortality prediction in intensive care units (ICUs) using a deep rule-based fuzzy classifier," *J Biomed Inform*, vol. 79, pp. 48–59, 2018, doi: 10.1016/j.jbi.2018.02.008.

[18] B. C. Kwon *et al.*, “RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, pp. 299–309, 2018, doi: 10.1109/TVCG.2018.2865027.
[19] B. Shickel, T. J. Loftus, L. Adhikari, T. Ozrazgat-Baslanti, A. Bihorac, and P. Rashidi, “DeepSOFA: A Continuous Acuity Score for Critically Ill Patients using Clinically Interpretable Deep Learning,” *Sci Rep*, vol. 9, no. 1, pp. 1–12, Feb. 2019, doi: 10.1038/s41598-019-38491-0.
[20] Q. Pan, K. Sun, J. Le, D. Chen, and H. Li, “An Intelligent Decision-Making Method for Graded Diagnosis and Treatment of Thyroid Disease Based on Deep Learning,” presented at the Proceedings of the 2018 International Conference on Machine Learning Technologies, 2018, pp. 6–11.
[21] L.-T. Huang, “An integrated method for cancer classification and rule extraction from microarray data,” *Journal of Biomedical Science*, vol. 16, no. 1, p. 25, Feb. 2009, doi: 10.1186/1423-0127-16-25.
[22] T. Van den Bulcke *et al.*, “Data mining methods for classification of Medium-Chain Acyl-CoA dehydrogenase deficiency (MCADD) using non-derivatized tandem MS neonatal screening data,” *J Biomed Inform*, vol. 44, no. 2, pp. 319–325, Apr. 2011, doi: 10.1016/j.jbi.2010.12.001.
[23] S. Bouktif, E. M. Hanna, N. Zaki, and E. A. Khousa, “Ant Colony Optimization Algorithm for Interpretable Bayesian Classifiers Combination: Application to Medical Predictions,” *PLOS ONE*, vol. 9, no. 2, p. e86456, Feb. 2014, doi: 10.1371/journal.pone.0086456.
[24] H. Ponce and M. de L. Martinez-Villaseñor, “Interpretability of artificial hydrocarbon networks for breast cancer classification,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 3535–3542, doi: 10.1109/IJCNN.2017.7966301.
[25] L. Crielaard and P. Papapetrou, “Explainable Predictions of Adverse Drug Events from Electronic Health Records Via Oracle Coaching,” in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, pp. 707–714, doi: 10.1109/ICDMW.2018.00108.
[26] Y. Ming, H. Qu, and E. Bertini, “RuleMatrix: Visualizing and Understanding Classifiers with Rules,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 342–352, Jan. 2019, doi: 10.1109/TVCG.2018.2864812.
[27] M. Hajiloo, H. R. Rabiee, and M. Anooshahpour, “Fuzzy support vector machine: an efficient rule-based classification technique for microarrays,” *BMC Bioinformatics*, vol. 14, no. 13, p. S4, Oct. 2013, doi: 10.1186/1471-2105-14-S13-S4.
[28] G. Luo, “Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction,” *Health Inf Sci Syst*, vol. 4, no. 1, p. 2, Dec. 2016, doi: 10.1186/s13755-016-0015-4.
[29] J. Zhang, K. Kowsari, J. H. Harrison, J. M. Lobo, and L. E. Barnes, “Patient2Vec: A Personalized Interpretable Deep Representation of the Longitudinal Electronic Health Record,” *IEEE Access*, vol. 6, pp. 65333–65346, 2018, doi: 10.1109/ACCESS.2018.2875677.
[30] P. Hartono, “A transparent cancer classifier,” *Health Informatics J*, p. 1460458218817800, Dec. 2018, doi: 10.1177/1460458218817800.

[31] L. P. Zhao and H. Bolouri, "Object-oriented regression for building predictive models with high dimensional omics data from translational studies," *J Biomed Inform*, vol. 60, pp. 431–445, Apr. 2016, doi: 10.1016/j.jbi.2016.03.001.

[32] S. G. Kim, N. Theera-Ampornpunt, C.-H. Fang, M. Harwani, A. Grama, and S. Chaterji, "Opening up the blackbox: an interpretable deep neural network-based classifier for cell-type specific enhancer predictions," *BMC Systems Biology*, vol. 10, no. 2, p. 54, Aug. 2016, doi: 10.1186/s12918-016-0302-3.

[33] G. Valdes, J. M. Luna, E. Eaton, C. B. Simone Ii, L. H. Ungar, and T. D. Solberg, "MediBoost: a Patient Stratification Tool for Interpretable Decision Making in the Era of Precision Medicine," *Scientific Reports*, vol. 6, p. 37854, Nov. 2016, doi: 10.1038/srep37854.

[34] G. Stiglic, S. Kocbek, I. Pernek, and P. Kokol, "Comprehensive Decision Tree Models in Bioinformatics," *PLOS ONE*, vol. 7, no. 3, p. e33812, Mar. 2012, doi: 10.1371/journal.pone.0033812.

[35] J. Zuallaert, F. Godin, M. Kim, A. Soete, Y. Saeys, and W. De Neve, "SpliceRover: interpretable convolutional neural networks for improved splice site prediction," *Bioinformatics*, vol. 34, no. 24, pp. 4180–4188, Dec. 2018, doi: 10.1093/bioinformatics/bty497.

[36] S. Ghafouri-Fard, M. Taheri, M. D. Omrani, A. Daaee, H. Mohammad-Rahimi, and H. Kazazi, "Application of Single-Nucleotide Polymorphisms in the Diagnosis of Autism Spectrum Disorders: A Preliminary Study with Artificial Neural Networks," *J Mol Neurosci*, vol. 68, no. 4, pp. 515–521, Aug. 2019, doi: 10.1007/s12031-019-01311-1.

[37] S. Paredes *et al.*, "A Clinical Interpretable Approach Applied to Cardiovascular Risk Assessment," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 3252–3255, doi: 10.1109/EMBC.2018.8512956.

[38] N. Barakat, A. P. Bradley, and M. N. H. Barakat, "Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 4, pp. 1114–1120, Jul. 2010, doi: 10.1109/TITB.2009.2039485.

[39] S. Park, Y. J. Kim, J. W. Kim, J. J. Park, B. Ryu, and J.-W. Ha, "[Regular Paper] Interpretable Prediction of Vascular Diseases from Electronic Health Records via Deep Attention Networks," in *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2018, pp. 110–117, doi: 10.1109/BIBE.2018.00028.

[40] W. Ge, J.-W. Huh, Y. R. Park, J.-H. Lee, Y.-H. Kim, and A. Turchin, "An Interpretable ICU Mortality Prediction Model Based on Logistic Regression and Recurrent Neural Networks with LSTM units.," *AMIA Annu Symp Proc*, vol. 2018, pp. 460–469, Dec. 2018.

[41] H. Lakkaraju, S. H. Bach, and J. Leskovec, "Interpretable Decision Sets: A Joint Framework for Description and Prediction," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, San Francisco, California, USA, 2016, pp. 1675–1684, doi: 10.1145/2939672.2939874.

[42] Z. Che, S. Purushotham, R. Khemani, and Y. Liu, "Interpretable Deep Models for ICU Outcome Prediction," *AMIA Annu Symp Proc*, vol. 2016, pp. 371–380, Feb. 2017.