Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/8/20230281

A comparative study of clinical risk prediction using limited patient electronic health records

Ziyu Li

University Of New South Wales

liziyu0410@gmail.com

Abstract. Predictive modeling of clinical risk using patient electronic health records (EHRs) has the potential to enhance healthcare outcomes by enabling early detection and intervention for high-risk patients. However, dealing with sparse, irregular, and temporal EHR data presents significant challenges. This paper presents a comparative study of clinical risk prediction with limited patient electronic medical records. The related literature is categorized and compared based on research objectives, methods and experimental analysis. Additionally, potential research opportunities for future work in this area are discussed. Meta-learning-based algorithms have the ability to overcome data scarcity challenge by learning shared feature representations. Nevertheless, further research is necessary to address limitations such as the interpretability and generalizability of the model across different patient populations.

Keywords: meta-learning, machine learning, clinical risk prediction, electronic health records.

1. Introduction

The integration of machine learning (ML) in healthcare has garnered significant attention in recent years. One of the most auspicious applications of ML in healthcare is clinical risk prediction, which entails forecasting the probability of a patient developing a specific disease or condition based on the patient's medical history and other pertinent factors. Nonetheless, developing precise clinical risk prediction models poses challenges due to the intricate nature of medical data, the vast amount of data necessary to train the model, and the intricacy of interpreting the model output.

This paper provides a review of a meta-learning-based approach to clinical risk prediction that can accurately predict patient risk levels even in instances where electronic medical records are limited. In Section 2, we categorize research pertaining to risk prediction into two categories: clinical medicine-related prediction and non-clinical medicine-related prediction. We also classify research objectives into two separate directions: accuracy improvement and performance improvement, which includes reducing the required sample size and increasing speed, among others. This classification facilitates a deeper understanding of the scope and nature of risk prediction research.

Section 3 of this paper partitions research methods pertaining to risk prediction into two distinct categories: general machine learning methods and dedicated machine learning methods. Within the realm of machine learning, models are further classified into discriminative and generative models. This classification enables comparison of different risk prediction modeling methods and facilitates the identification of common methods utilized in clinical risk prediction research.

DOI: 10.54254/2755-2721/8/20230281

Section 4 of this paper conducts a comparative experimental analysis of related literature on risk prediction modeling. Through comparison of experimental measures and system factors, we examine the approaches adopted by relevant papers towards clinical risk prediction and the insights that they have yielded.

In Section 5, we suggest that clinical risk prediction systems may not perform optimally on certain datasets. Furthermore, we discuss potential opportunities for future research in clinical risk prediction modeling and highlight areas that warrant further exploration and may serve as possible future research directions.

Lastly, this paper summarizes the meta-learning-based clinical risk prediction model discussed within, as well as the research findings of related papers on clinical risk prediction models. The key points of analysis are emphasized to provide a comprehensive understanding of the efficacy of the proposed model and the insights gained from prior research in this field.

The remainder of the paper is organized as follows: Section 2 presents the classification of research objects pertaining to clinical risk prediction. Section 3 introduces the classification of research methods. Section 4 provides a comparative analysis of experimental studies in related literature. Section 5 discusses potential opportunities for future research, and Section 6 offers a conclusion to the paper.

2. Classification of research objects

Table 1. Different research objects.

Research Scope	**Research Directions**	
	Efficiency	**Accuracy**
Medical	I. [6][7][8][14][22][23]	II.[3][34][42]
Non-medical	III.[9][10][13][15][16][17][25][26][27][30][32][33][35][40][41]	IV. [20]

2.1. Criteria

Given that the focus of the studied paper is the application of neural networks in the medical field, it is appropriate to classify research objects into two distinct types based on their medical relevance. In this section, two independent criteria will be employed to differentiate research objects:

1)Research Scope: This criterion divides research objects into two categories - medical and non-medical - based on their application scenarios in machine learning.

2)Research Direction: This criterion divides research objects into two categories - efficiency and accuracy - based on the direction of research. Efficiency includes factors such as generalization ability, learning rate, sample requirements, and interpretability.

2.2. The classification

Based on the appeal classification standard, we give the classification in Table 1. The meaning of each class is as follows:

2.2.1. Type I: efficiency & medical. This category focuses on research that aims to improve the efficiency of machine learning methods in the medical field. References ([6] [7] [8] [14] [22] [23]) belong to this type. Reference [6] proposes the development of a more explanatory machine learning model using electronic health record (EHR) data. Reference [7] proposes a graph-based attention model for healthcare representation learning. The proposed model utilizes the graph structure of electronic health records to capture the relationships between different medical concepts and learn a low-dimensional representation of patients. Reference [8] proposes the development of a RNN model for the early detection of heart failure episodes. Reference [14] proposes a DNN approach to classifying skin lesions into three categories: benign, malignant, and nonthreatening. Reference [22] aimed to predict a model for frequent COPD exacerbates using variables such as prior COPD exacerbation history, smoking status, and medication use. Reference [23] proposes an image-based deep learning

DOI: 10.54254/2755-2721/8/20230281

approach for identifying medical diagnoses and treatable diseases. The proposed approach utilizes a deep neural network to analyze medical images.

2.2.2. Type II: accuracy & medical. This type of research focuses on improving the accuracy of machine learning methods related to the medical field. References ([3] [34] [42]) belong to this type. Reference [3] proposes a deep learning framework based on longitudinal electronic health record data for subtyping patients. The framework is time-aware. The results demonstrate that the method outperforms other methods in terms of accuracy. Reference [34] proposes to use EHR data to develop predictive models for predicting the likelihood of readmission within 30 days of patient discharge based on pre-admission data such as demographics, past medical history, and medication use. The results showed that the prediction model achieved high accuracy in identifying patients at risk of readmission. Reference [42] used a dataset of EHRs from patients with heart disease and applied a convolutional neural network (CNN) to discover temporal patterns involving multiple types of clinical events, such as diagnoses, procedures, and medications. The research found that the CNN method effectively improved accuracy.

2.2.3. Type III: efficiency & non-medical. This type of research on improving the efficiency of machine learning methods is related to non-medical fields. References ([9] [10] [13] [15] [16] [17] [25] [26] [27] [30] [32] [33] [35] [40] [41]) belong to this type. Reference [9] proposes an approach called "RL2", which combines reinforcement learning and meta-learning. The approach is designed to enable agents to learn, so they can quickly adapt to new tasks and environments. Reference [10] proposes a deep neural network architecture that can perform multiple natural language processing (NLP) tasks simultaneously through multitask learning. The research shows that a unified architecture with multitask learning can improve the performance of NLP tasks. Reference [13], a new language representation model. The model performs better than the previous model on several benchmark datasets. Reference [15] proposes a method for quickly adapting deep neural networks to new tasks. Reference [16] proposes a meta-learning approach for one-shot visual imitation learning, which allows a robot to learn how to perform a new task from just one demonstration. Reference [17]'s proposed method utilizes a meta-learner to learn how to quickly adapt to a new low-resource language pair by leveraging knowledge learned from other language pairs. Reference [25] describes a new optimization algorithm called Adam, which is designed to be computationally efficient and has been shown to converge quickly and robustly on a wide range of optimization problems, including those involving deep neural networks. Reference [26] proposes a method for image recognition using Siamese neural networks, which are deep learning models composed of two or more identical sub-networks that share the same weights. The experiments show Siamese neural networks outperform traditional approaches for one-shot learning. Reference [27] proposes a method to learn concepts using probabilistic program induction (PPI) and demonstrates the effectiveness of the approach on a number of tasks. Reference [30] proposes a pre-trained language representation model specifically designed for biomedical text mining. Reference [32] and [35] propose an approach to address the problem of few-shot learning, which is the ability to recognize new objects or classes with limited training examples. Reference [33] proposes a new approach to meta-learning called the Memory-Augmented Neural Network (MANN). The research shows that the proposed Memory-Augmented Neural Network outperforms existing methods. Reference [40] proposes a neural network architecture called "Transformer," which relies on recurrent neural networks but suffers from computational inefficiencies and difficulty in parallelization. Reference [41] proposes a novel approach for one-shot learning, where a model can recognize a new object from just one example. A comparison with several baselines shows that the performance of the matched network is significantly better than them.

2.2.4. Type IV: accuracy & non-medical. This type of research focuses on improving the accuracy of machine learning methods related to non-medical fields. References ([20]) belong to this type. Reference [20] proposes a technique called batch normalization, which normalizes the inputs to a layer by subtracting the mean and dividing by the standard deviation of the inputs within a mini-batch. The

DOI: 10.54254/2755-2721/8/20230281

experimental results provided in this paper show that batch normalization can significantly improve the accuracy of deep neural networks in the task.

3. Classification of research methods

Table 2. Different research methods.

Algorithm Types	Model Types	
	Discriminative Model	**Generative Model**
Specialized Machine Learning Algorithms	I.[14][30]	II. [23]
General-purpose Machine Learning Algorithms	III. [8][34][38]	IV.[13][15][16][17][27][33][41]

3.1. Criteria

Machine learning can be applied in many fields, such as image processing, recommendation systems, medical prediction, etc. Using algorithms that meet the characteristics of different application scenarios can improve the performance of models and save costs. Research can be done to generate a generative model by preprocessing the data or to face the predictions directly. In this section, two independent and different criteria will be used to divide research objects into different types:

1)Algorithm types: There are two types here: Specialized Machine Learning Algorithms and General-purpose Machine Learning Algorithms.

2)Model types: There are two kinds of models here: Discriminative Model and Generative Model. The research can generate the generative model by learning the conditional probability distribution from the joint probability distribution of the data, and it can also directly learn the decision function or conditional probability distribution from the data to generate the discriminative model.

3.2. The classification

Based on the appeal classification standard, we give the classification in Table 2. The meaning of each class is as follows:

3.2.1. Type I: discriminative model & specialized machine learning algorithms. This type of algorithm, when applied to specific domains, generates discriminant models. References ([14] [30]) belong to Type I. Reference [14] proposed in the paper is a discriminative model that uses a deep convolutional neural network (CNN) to classify skin lesions as benign or malignant. Reference [30] created a new language model called BioBERT, which can be used to identify disease characteristics.

3.2.2. Type II: generative model & specialized machine learning algorithms. This type of algorithm is applied to specific domains to generate generative models. References ([23]) belong to Type II. Reference [23] re-processes large datasets of medical images to generate generative models to identify treatable diseases by analyzing medical images.

3.2.3. Type III: discriminative model & general-purpose machine learning algorithms. This type of algorithm is applied to general-purpose domains and generates discriminant models. References ([8] [34] [38]) belong to Type III. Reference [8] uses RNN models to analyze EHR data, the researchers are able to identify patterns that could indicate the early stages of heart failure. Reference [34] describes using machine learning algorithms to analyze EHR data to predict which patients are at higher risk for readmission. Reference [38] compares the performance of several machine learning algorithms, including logistic regression, decision trees, and random forests, in predicting three outcomes: patient discharge, hospital admission, and transfer to another healthcare facility.

3.2.4. Type IV: generative model & general-purpose machine learning algorithms. This type of algorithm is applied to general-purpose domains and generates generative models. References ([13] [15] [16] [17] [27] [33] [41]) belong to Type IV. Reference [13] suggests pre-training it on a large corpus of text data to enable it to learn general language understanding and capture the semantic relationships between words. In the method of reference [15] [16] [17], the initial parameters of the model are optimized to adapt to the new task quickly, so as to train the model to adapt to the new task quickly. Reference [27] proposes a method for learning probabilistic programs from input-output examples. Reference [33] proposed a new method of meta-learning. The model is trained on sequences of tasks, which combine a neural network with an external memory module. The memory module stores past experiences, and the neural network learns to use these experiences to adapt to new tasks. Reference [41] proposes a new framework called "Matching Networks," which can learn to classify new objects from just one or a few examples.

4. Review of experimental analysis

In this section, we will classify the metrics of evaluation and system factors, as shown in Table 3. In Table 3, all experimental analysis is also classified according to the metric and factors. It can be seen from Table 3 that most of the references compare precision, recall, and F1-score.

Table 3. Experiments with different metric and factors.

Metric	System Factors			
	Algorithms	Dataset	Epoch	Others
Precision/Accuracy	[3][8][10][13][14][15][18][23][32][38][41][42]	[2][3]		
Recall	[3][15][38][41][42]			
F1-score	[3][7][10][13][15][30][35][38][41][42]			
AUC-ROC	[3][6][7][8][13][14][18][23]][38][42]			[11][15]
Sensitivity/Specificity	[14][18][23][38]			
Others	[7][13][16][18][23][27][30][38][40]	[17][23]	[25][23]	[22]

4.1. Metric of evaluation

Precision means the ratio of true positives (TP) to the total number of samples that are predicted as positive by a classifier. It measures how accurately the classifier identifies positive samples. The formula is as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{false positive(FP)}}$$

Recall means the ratio of true positives (TP) to the total number of actual positive samples. It measures how well the classifier can identify all positive samples. The formula is as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{false negative(FN)}}$$

F1-score means the harmonic mean of precision and recall, which combines both metrics into a single value. It provides a balanced evaluation of the classifier's performance by taking both precision and recall into account. The higher the F1-score, the better the performance of the classifier. The formula is as follows:

$$\mathrm{F1-score} = \frac{2 * (\mathrm{Precision} * \mathrm{Recall})}{(\mathrm{Precision} + \mathrm{Recall})}$$

AUC-ROC is a performance metric used to evaluate binary classification models. The AUC is the area under the ROC curve and ranges between 0 and 1. The higher the AUC, the better the performance of the model.

Sensitivity means the proportion of people with a particular disease who test positive for it. It measures the ability of a test to correctly identify patients who have the disease. Specificity means the proportion of people without a particular disease who test negative for it. It measures the ability of a test to correctly exclude healthy individuals.

Other metrics include loss function, reward function, BLEU and PPV, and NPV.*System factors* Algorithms represent the research using models and algorithms. The dataset represents the research that used a training set, a validation set, and a test set. Epoch represents a complete pass through the entire training dataset during model training. Other factors include environment, learning rate, and neural network layers.

4.3. Experimental comparison

In reference [3][13][15], the authors used precision, recall,AUC-ROC, accuracy and F1-score metrics to evaluate the performance. In reference [6], the authors use receiver ROC curves to measure the tradeoff between the true positive rate and the false positive rate for different classification thresholds. They also report the area under the AUC-ROC scores as a summary metric of model performance. In reference [7], the model is evaluated with using standard metrics such as the AUC-ROC and F1 score. In reference [8], the performance of the model is evaluated using standard metrics such as AUC-ROC and accuracy. Hyperparameters such as the number of LSTM cells and learning rate are tuned during training to optimize performance. In reference [10], the performance of the model is evaluated using standard metrics such as accuracy and F1 score. Hyperparameters such as the number of hidden units and learning rate are tuned during training to optimize performance. In reference [14], evaluation metrics such as accuracy, sensitivity, specificity, and AUC-ROC are used to assess model performance. Hyperparameters such as learning rate and dropout rate are tuned during training to optimize performance. In reference [18], evaluation metrics such as AUC-ROC, sensitivity, specificity, PPV, NPV, and accuracy are used to assess model performance. Several hyperparameters are tuned during training, including learning rate, batch size, weight decay, and dropout rate. In reference [23], evaluation metrics such as AUC-ROC, sensitivity, specificity, PPV, NPV, and accuracy are used to assess model performance. Several hyperparameters were tuned during training, including learning rate, batch size, weight decay, and dropout rate. In reference [30], evaluation metrics such as the F1 score and EM are used to assess model performance on specific tasks such as NER, relation extraction, and question answering. Several hyperparameters were tuned during training, including learning rate, batch size, number of training epochs, and maximum sequence length. In reference [32][38][42], evaluation metrics such as accuracy, precision, recall, and the F1 score are used to assess model performance. Several hyperparameters are tuned during training, including the number of iterations used to train the meta-learner, the number of examples used for each task, and the learning rate used for both the meta-learner and base learner. In reference [41], evaluation metrics such as accuracy, precision, recall, and F1 score are used to assess model performance. Several hyperparameters are tuned during training, including learning rate, batch size, number of training epochs, and embedding dimension.

5. Discussion and suggestion

This paper discusses the research methods and research objects of various references and finds that although clinical risk prediction is able to perform clinical risk prediction with limited data, it should be noted that in some datasets, the clinical risk prediction algorithm still performs worse than other

DOI: 10.54254/2755-2721/8/20230281

algorithms. Therefore, this paper puts forward the following directions, which can provide directions for future clinical risk prediction research:

1) Further validation: Although the clinical risk prediction algorithm has achieved good prediction performance on multiple datasets, its effectiveness still needs to be further verified. Future research could consider using more and richer datasets to verify the performance of the algorithm.

2) Scalability: The scalability of the clinical risk prediction algorithm is an issue to be considered. Since the algorithm is based on meta-learning, more domain expert knowledge may be needed to extend the applicable scope of the algorithm.

3) Interpretability: In clinical risk prediction, the interpretability of the model is very important. Future research can consider increasing the interpretability of the algorithm in order to better understand the prediction results of the algorithm and improve its application value in clinical practice.

6. Conclusions

This research article introduces clinical risk prediction, a meta-learning-based algorithm for clinical risk prediction with limited electronic health records of patients. The algorithm overcomes the challenge of a lack of data by learning shared feature representations and performs well on multiple datasets. The main contribution of this paper is to propose a new method for clinical risk prediction with limited data. Clinical risk prediction learns a shared feature representation by meta-learning, so that it can be applied to multiple datasets and improve prediction performance. Experimental results show that the clinical risk prediction algorithm achieves better prediction performance than other algorithms on multiple data sets. Although the proposed algorithm performs well on several datasets, further verification of its effectiveness is needed. Future research could consider using more and richer datasets to verify the performance of the algorithm. In addition, the scalability and interpretability of the algorithm are also issues that need to be further explored. In general, the clinical risk prediction algorithm provides a new idea and method for clinical risk prediction with limited data, which has great research and application value.

References

[1] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. 2016. Learning to learn by gradient descent by gradient descent. In *NIPS.*

[2] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450* (2016).

[3] Inci M Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K Jain, and Jiayu Zhou. 2017.Patient subtyping via time-aware LSTM networks. In *KDD*.

[4] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. 2010. A theory of learning from different domains. *Machine learning* 79, 1-2 (2010).

[5] Rich Caruana. 1997. Multitask learning. *Machine learning* 28, 1 (1997).

[6] Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. 2015. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *KDD*.

[7] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun. 2017. GRAM: graph-based attention model for healthcare representation learning. In *KDD*. 787−795.

[8] Edward Choi, Andy Schuetz, Walter F Stewart, and Jimeng Sun. 2016. Using recurrent neural network models for early detection of heart failure onset. *JAMIA* 24, 2 (2016).

[9] Ignasi Clavera, Anusha Nagabandi, Simin Liu, Ronald S Fearing, Pieter Abbeel, Sergey Levine, and Chelsea Finn. 2018. Learning to Adapt in Dynamic, Real-World Environments through Meta-Reinforcement Learning. (2018).

[10] Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing:

Deep neural networks with multitask learning. In *ICML*.
[11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *CVPR*.
[12] Li Deng, Jinyu Li, Jui-Ting Huang, Kaisheng Yao, Dong Yu, Frank Seide, Michael L Seltzer, Geoffrey Zweig, Xiaodong He, Jason D Williams, et al. 2013. Recent advances in deep learning for speech research at Microsoft.. In *ICASSP*, Vol. 26.
[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
[14] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 7639 (2017).
[15] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic metalearning for fast adaptation of deep networks. In *ICML*.
[16] Chelsea Finn, Tianhe Yu, Tianhao Zhang, Pieter Abbeel, and Sergey Levine. 2017. One-Shot Visual Imitation Learning via Meta-Learning. In *Conference on Robot Learning*. 357–368.
[17] Jiatao Gu, Yong Wang, Yun Chen, Victor OK Li, and Kyunghyun Cho. 2018. Meta-Learning for Low-Resource Neural Machine Translation. In *EMNLP*.
[18] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 22 (2016).
[19] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997).
[20] Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *ICML*.
[21] Peter B Jensen, Lars J Jensen, and Søren Brunak. 2012. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics* 13, 6 (2012).
[22] Marjan Kerkhof, Daryl Freeman, Rupert Jones, Alison Chisholm, and David B Price. 2015. Predicting frequent COPD exacerbations using primary care data. *International journal of chronic obstructive pulmonary disease* 10 (2015).
[23] Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 172, 5 (2018).
[24] Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014).
[25] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
[26] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. 2015. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, Vol. 2.
[27] Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. 2015. Humanlevel concept learning through probabilistic program induction. *Science* 350, 6266 (2015).
[28] Neil D Lawrence and John C Platt. 2004. Learning to learn with the informative vector machine. In *ICML*.
[29] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015).
[30] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. BioBERT: pre-trained biomedical language representation model for biomedical text mining. *arXiv preprint arXiv:1901*.08746 (2019).
[31] Sinno Jialin Pan, Qiang Yang, et al. 2010. A survey on transfer learning. *IEEE Transactions on*

knowledge and data engineering 22, 10 (2010), 1345−1359.

[32] Sachin Ravi and Hugo Larochelle. 2016. Optimization as a model for few-shot learning. (2016).

[33] Samuel Ritter, Jane Wang, Zeb Kurth-Nelson, Siddhant Jayakumar, Charles Blundell, Razvan Pascanu, and Matthew Botvinick. 2018. Been There, Done That: Meta-Learning with Episodic Recall. In *ICML*.

[34] Efrat Shadmi, Natalie Flaks-Manov, Moshe Hoshen, Orit Goldman, Haim Bitterman, and Ran D Balicer. 2015. Predicting 30-day readmissions with preadmission electronic health record data. *Medical care* 53, 3 (2015).

[35] Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. In *NIPS*.

[36] Mengying Sun, Fengyi Tang, Jinfeng Yi, Fei Wang, and Jiayu Zhou. 2018. Identify Susceptible Locations in Medical Records via Adversarial Attacks on Deep Predictive Models. In *KDD*.

[37] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *CVPR*.

[38] Fengyi Tang, Cao Xiao, Fei Wang, and Jiayu Zhou. 2018. Predictive modeling in urgent care: a comparative study of machine learning approaches. *JAMIA Open* (2018).

[39] Sebastian Thrun and Lorien Pratt. 1998. Learning to learn: Introduction and overview. In *Learning to learn*.

[40] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*.

[41] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one shot learning. In *NIPS*.

[42] Fei Wang, Noah Lee, Jianying Hu, Jimeng Sun, and Shahram Ebadollahi. 2012. Towards heterogeneous temporal clinical event pattern discovery: a convolutional approach. In *KDD*.