

A Convolution Neural Network Based Framework for Similarity Learning in Healthcare

Krishan Kumar Saraswat, Sandhya Tarar, Sandeep Gupta

Abstract: *Foreseeing patients' danger of building up specific ailments is a vital research theme in social insurance. Precisely distinguishing and positioning the similitude among patients in view of their chronicled records is a key advance in customized social insurance. The medical data for two different diseases are unpredictably examined and have fluctuated tolerant visit lengths, can't be specifically used to quantify quiet similitude because of the absence of a suitable portrayal. Also, there needs a powerful way to deal with measure quiet similitude on electronic health records. In this paper, we propose a novel significant equivalence learning frameworks which in the meantime get the hang of understanding depictions and measure pairwise similarity. We utilize convolutional neural system (CNN) to catch nearby vital data in electronic health data and after that feed the scholarly portrayal into triplet metric similarity learning. In the wake of preparing, we can get pairwise separations and likeness scores. Using the closeness data, we at that point perform patient grouping. Test results demonstrate that CNN can more readily speak to the longitudinal medical data groupings, and our proposed systems can efficiently cluster patients into different disease groups.*

Index Terms: *Convolutional Neural Network ; Patient Clustering; Personalized Healthcare; Triplet Loss Metric Learning*

I. INTRODUCTION

The pervasiveness and developing the volume of Electronic Health Records (EHRs) gives remarkable chances to enhance clinical choice help. The EHR information, which is a longitudinal electronic record of patient wellbeing data, is a profitable hotspot for prescient demonstrating which can help clinical and therapeutic research [1]. Mining EHRs is particularly testing contrasted with standard information mining assignments, because of its uproarious, unpredictable and heterogeneous nature.

Customized social insurance has gotten expanding enthusiasm from specialists [2]. A conventional system for customized prediction contains two phases: estimating the similitude among patients and gathering patients into accomplices, and examining the companion to per frame malady conclusion, treatment medicine, and so forth [5]-[8].

Revised Manuscript Received on December 22, 2018.

Krishan Kumar Saraswat, Research scholar, Department of Computer Science and Engineering, Shri Venkateshwara University, Gajraulla, India

Sandhya Tarar, Department of CSE, School of ICT, Gautam Buddha University, Greater Noida, India

Sandeep Gupta, Department of CSE, JIMS Engineering Management Technical Campus, Greater Noida, India

This structure is conceptualized by the functioning procedure of human specialists, particularly in the wake of checking on or reviewing the determined patients to have comparative illnesses or indications, the specialists at that point precisely settle on a choice [4]. In the event that specialists can discover comparative patients, the likelihood of effectively restoring a patient possibly enhance a ton. Accordingly, the process of evaluating and measurement of understanding similitude is a necessary as well as a testing issue.

Numerous comparability acquirements strategies have been suggested on medicinal services [9]–[13]. Nonetheless, all the previous structures are produced for high-quality vector portrayals, for example, socioeconomic or normal numerical qualities, neglecting the fleeting data from various visits. For the longitudinal health records data, the number of visits of patients differ to a great extent, because of patients' unpredictable visits and inadequate accounts. The previously mentioned learning measurements can't be straightforwardly apply to the data which is longitudinal in nature as the verifiable data belonging to a patient don't normally frame a com-illustration vector. In this manner, one of the key difficulties in estimating understanding similitude is to determine a successful portrayal for every patient. A customary portrayal based on vector is to outline information measurements (for example, entirety, normal, maximum, and so forth) of relating occasions inside a day and age and figure likeness separate over those patient vectors. In any case, this evacuates transient relations crosswise over neighboring visits. As of late, deep learning strategies have been broadly embraced and quickly created in quiet understanding of portrayal [14]–[21], for example, convolution neural systems (CNNs) repetitive neural systems (RNNs) and auto encoders etc. In this paper, we suggest a profound system of metric learning on Electronic health record to gauge patient likeness and similar patterns. The model is divided into two parts : portrayal learning and comparability learning. In portrayal learning, the ability of CNN to be applied on longitudinal data is utilized efficiently in order to get a vector portrayal consisting of nearby critical first information data. In the second part which is similarity learning , we utilize two different methods to take in the comparability across different persistent sets. The first method depends on triplet loss work, which takes in an edge to isolate the separation of negative and positive examples.



The second method performs arrangement on the educated conventions giving a positive name to comparable combines while a negative name to divergent sets. Eventually, closeness likelihood across a couple of patients shows the probability of them building up a similar infection, we can use it as a measure of the comparability across patients. In the wake of acquiring the likeness data, we perform two assignments: illness expectation and patient bunching which are application regions of customizing medicinal services, keeping in mind the end goal to approve the scholarly measurements. In outline, our commitments are to conceptualize two systems to mutually learn understanding the health portrayals and patient likeness, neglecting the handmade element totals. In our suggested systems, CNN makes utilization of sequential structure at the same time learning the neighborhood imperative data, triplet metric guarantees huge edge to isolate the examples in a similar class and tests in various classes. Our exploratory outcomes demonstrate that our similitude learning system can understand better portrayal vectors for historical data and enhance the likeness learning precision contrasted with other best in class gauge methods.

II. PROCEDURE FOR PAPER SUBMISSION

In this segment, we audit previous methods of assessing persistent closeness and designing of customized models. For another patient, recognizing the authentic data of similar patients could help recover comparable reference cases for anticipating the clinical results of the new patient. [1] consolidated patient similitude, medication likeness investigation and suggested a composite name proliferation technique to recognize whichever tranquilize is probably going to be viable for a particular patient. By and by, various doctors have distinctive understandings of patient likeness in view of the particular cases. Utilizing doctor supervision, [9] displayed a privately managed learning of metric(LSML) calculation that takes in a summed up Mahalanobis separate, provided that getting doctors' information is troublesome and costly in all actuality, further Wang et al. [22] suggested a pitifully managed quiet comparability learning technique which just uses a little measure of supervision data given by the doctors. Because of the way that patient comparability is exceedingly setting delicate.[23] utilized a number of measurable along with wavelet dependent highlights to catch the attributes of patients, also displayed a patient resemblance learning strategy which uses restricted regulated metric learning. Thinking about the high spatial nature and excess of therapeutic information, Zhan et al. [10] suggested to include choice and patient comparability learning simultaneously.

As of late, customized forecast in medicinal services receives expanding enthusiasm from scientists. It plans to discover the novel attributes for singular patients, and perform focused on, understanding particular forecasts, proposals, and analysis [24]- [25]. The greater proposed part is customized forecast by coordinating clinical comparative patients. [6] played out a relative investigation of worldwide, neighborhood, and customized demonstrating, and found that customized models can accomplish better execution crosswise over various bioinformatics characterization

undertakings. [26] utilized a privately administered metric learning for comparability estimation and calculated relapse as the prescient diabetes model beginning expectation [2]used cosine separation to get patient's separation and constructed refiners for mortality forecast. [7] propose the basis model of learning of the populace and a customized model of every patient by means of a scanty multiple assignment learning technique.

The previously mentioned strategies need the contribution of every patient in the form of a vector. A conventional route is to physically get highlight vectors by utilizing patients static records, for example, statistic, and information insights (e.g. aggregate, normal, and so forth) inside a specific time go, as the patient portrayal. In any case, these high-quality element vectors totally disregard the worldly relations crosswise over visit successions. To represent the fleeting data, [12] utilized a dynamic programming calculation to discover ideal nearby arrangements of patient groupings,[27] created two answers for understanding similitude learning, administered and unsupervised, utilizing a CNN-dependent likeness coordinating system, and [5] built up a 2D-Recurrent neural network for dynamic worldly coordinating of patient successions to acquire the patient closeness positioning [37]-[38].

III. METHODOLOGY

Within this section, we introduce the subtle elements of our metric learning models the acquired datasets. Initially, we demonstrate a viable portrayal for the longitudinal EHR information, and after that acquaint two techniques with measure the likeness between understanding sets. With the educated likeness data, we at that point perform the patient grouping.

A. Representation Learning

a. Basic Notations

The data of each patient contains an arrangement of visit data, and during each visit, restorative codes are noted demonstrating the illness or analysis the patient endured or received. We signify all the one of a kind medicinal symptoms as $c_1; c_2 \dots c_j$. Expecting there are N patients and each person P_i has made V_i number of visits. Each visit V_i is meant by a high dimensional two fold vector. Figure 1 demonstrates a case of the obtained health record information, which can be seen as a grid. The level hub relates to visits, and the vertical hub is medicinal symptoms.

However, the quantity of visits of various patients differs, we cushion 0 to the visit measurement, making every patient to have a settled visits length $t = \max(\sum_{i=1}^{i=N})$ for Convolution Neural Network activities.

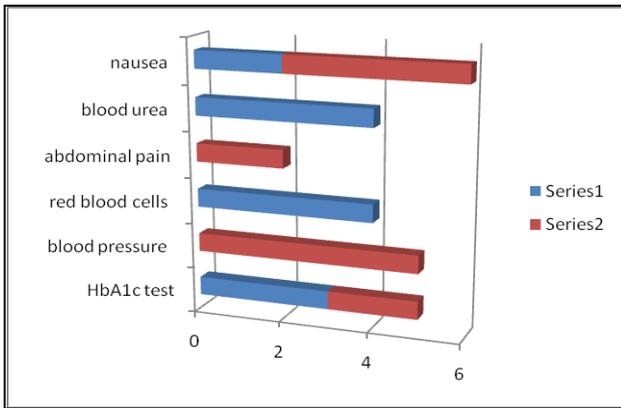


Fig. 1: Figure to demonstrate patient symptoms vs. visits where the y-axis represents the parameters recorded and the x-axis denotes the number of visits. Different colors represent two different patients

b. Visit Enclosing

The first one-hot portrayal expressed in previous section overlooks relations between codes and renders the data lattice inadequate in nature. To lessen include measurements and understand connections across codes, we utilize a completely associated arrange a pattern to implant a coded vector space. Accordingly, a visit (v_i) to vector (x_i) mapping is performed by utilizing the following equation

$$x_i = \text{ReLU}(W_i V_i + b_i) \quad (1)$$

Where ReLU is the Rectified Linear Unit activation function b and W are respective bias and Weight metric.

c. Convolutional Neural Network

A 2D Pixel framework can be generated from the embedded patient matrix and a convolution activity can be connected to catch the consecutive connection crosswise over nearby visits. Be that as it may, not quite the same as pictures with dimensionally related crosswise over pixels in two measurements, the places of restorative codes includes no dimensional significance, which makes the convolution activity crosswise over component spatially absurd [29]. In this way, a one-side convolution task over the time measurement is connected to catch the successive connection crosswise over contiguous visits as opposed to utilizing a standard 2D CNN.

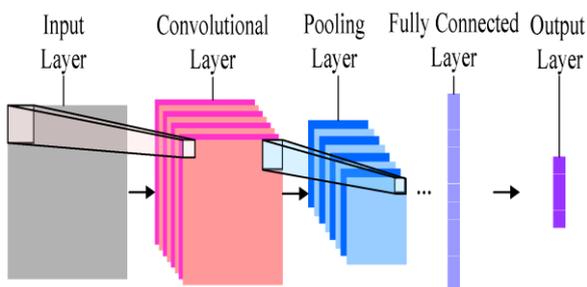


Fig 2: Convolutional Neural Network

The CNN layer includes x diverse filters and the quantity of channels per measure is y , so the aggregate quantity of channels becomes $m = xy$. Every channel is characterized as a network $W_c \in \mathbb{R}^{h \times d}$, where h is an assumed visit length size,

implying that our convolution task is connected onto h successive timeframes. Now, a scalar value c can be calculated by applying convolution between two visits

$$c_i = \text{ReLU}(W_{c \times x_i + h-1} + b_c) \quad (2)$$

where b_c is bias term, x_i and x_{i+h-1} are the two visit vectors where the concatenation is applied and \cdot is the convolution activity. A feature map containing values $c = \{c_1, c_2, \dots, c_{i+h-1}\}$ can be obtained by applying convolution between successive visits from x_i to x_{i+h-1} taking a stride of 1 unit. A pooling layer receives the output from convolution layer. A maximum pooling is connected onto c as $C_m = \text{Max}[c]$, where C_m contains the most extreme esteem relating to a specific channel. The key thought here is to catch the most essential data for each element outline. It can normally manage visits having variable lengths as the cushioned visits carry zero commitment as in the pooled yields.

The pooled yields from every one of the channels are connected to shape a vector portrayal $h \in \mathbb{R}^m$. The scholarly vector h is the portrayal of the first installing framework X and includes visit data as well as the relationship crosswise over neighboring time focuses.

B. Similarity Learning

Taking in the corresponding comparability/separate across every match of patients generates an important advance for customized healthcare. We introduce two techniques to quantify the comparability across persistent vectors according to Section A, and triplet-loss based system.

a. Triplet-Loss Metric Learning

Metric learning expects to take in an appropriate separation metric for a specific errand, which is essential to the execution of numerous calculations. We use the possibility of metric figuring out how to take in the relative separation of patients. In customary metric taking in, a straight change L is utilized to delineate crude information into another space. The new metric can all the more likely measure the relative separation of info occurrences. The separation between occasions x_i and x_j can be gotten by ascertaining the Euclidean separation in the contemporary space. We utilize triplet-loss as the goal work [31]. This includes an arrangement where every triplet consist a grapple, a positive and a negative illustration. A positive example has a similar classmark as the grapple, one the other hand the negative example has the distinctive class name. Amid the preparation, the positive ought to be drawn nearer to the grapple and the negative ought to be separated, i.e. the separation between stay p_i and positive example p_{-i} ought to be nearer as compared to separation amongst p_i and negative example p_i with some settled edge



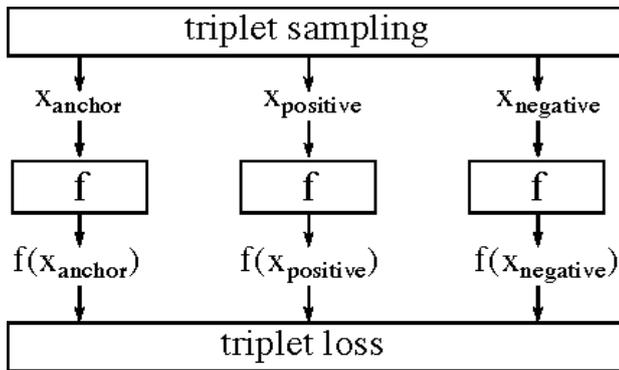


Fig. 3: Triplet loss metric learning

The scholarly comparability can be utilized for customized expectation. The closeness score can be utilized to quantify the similitude degree between a couple of patients. We utilize the K-Nearest Neighbor (KNN) classifier to anticipate patients' danger of building up specific illnesses later on. For every test tolerant, we include the nearest k patients from the preparation consisting of primary k littlest separation, furthermore, the trivial most mark showing up across the k preparing tests regarded as anticipated name.

Naturally, because patient already possesses comparative wellbeing records/side effects, it is very conceivable that there are chances of them building up a similar ailment. The grouping errand can dole out marks to tests, however, it doesn't give the data of how close or how far the separations are. Subsequently, to all the more likely assess the scholarly separation measurements/similitude, we additionally perform grouping strategy on the mapped spaces. Thusly can give specialists a natural visualization of the separation appropriation of patients

IV. EXPERIMENT AND RESULTS

A. Data Characterization

We lead investigates a certifiable dataset, which comprises of therapeutic cases from in excess of 10000 patients more than 110 symptoms. To perform illness forecast, we extricate two patient associates from the obtained datasets: diabetes and CKD (Chronic Kidney Disease). Following the ailment determination benchmark in [32], we recognize the sick patients possessing disease enabling parameters for a particular illness in the experience records or prescription requests, and no less than four clinical experiences with disease enabling codes happen inside a year. We divide patient groupings at the filing date across two sections, and utilize just the section previous to the record date which contains early side effects and complexities for comparability learning and sickness forecast. To empower unmistakable accomplices, we expel covered patients with the goal which every person just experiences one sickness.

B. Experimental Setup

a. Model Implementation

We first prepare the comparability model portrayed in the part 3.1 to get the improved criterion of Convolution Neural Network's and coordinating metric. At that point, utilizing the systems, we figure and mark the distance/similitude of every

single testing occurrence. In the wake of getting the closeness data, we perform two errands: ailment forecast and patient bunching. For the closeness preparing process we follow the criterion as two or more patients having a similar infection are assumed to be a positive example match on contrary, having diverse illnesses are a negative example combine. The similitude learning systems are executed with Tensorflow [31]. Adam [32] is utilized to upgrade demonstrate benchmarks. Not the same as a typical CNN demonstrate including contribution to be a scaled down cluster of patients, the likeness structure is prepared on a bunch of patient sets to guarantee the fact that every one of the patient sets has the ability to be estimated.

b. Baseline Approaches

To approve the execution of the introduced profound patient comparability methods, we contrast them and the accompanying best in class gauge techniques. Euclidean and Cosine remove on crude sources of information are figured to gauge the likeness between test sets. The two techniques straightforwardly measure similitude on the first information space, with no parameter mapping to be educated. Separation of metric learning techniques. LMNN [34] is a traditional metric learning technique, working by bringing the K-closest neighbors having a place with a similar class nearer, and isolates cases from various classes by a substantial edge. ITML [34] takes in the Mahalanobis separate by limiting the differential relative entropy. GMML [36] details the understanding procedure as a curved improvement issue. SCML inclines an inadequate mix of locally separation measurements, which sees the Mahalanobis lattice as positive weighted aggregate of k-lowest-parameters premise. Component in the vector shows the recurrence of a comparing code

C. Results

To assess the execution of each and every patient comparability learning methods on illness pre-word usage, we ascertain the measures such as exactness, accuracy, review and F1 score. Furthermore the estimations for parallel arrangement can't be straightforwardly utilized as we have performed multiclass characterization. Persistent patient grouping can investigate ailment associate conveyance

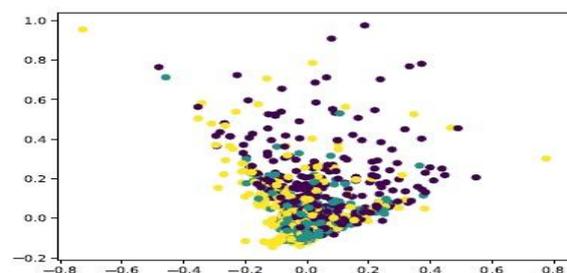


Fig. 4: Euclidean transform space result

Fig 4,5,6 and 7 represent a transformed space visualization of the samples taken for testing using Euclidean, ITML, LMML and proposed CNN-Triplet methods. The decision of different bunching calculation ought not impact the relative execution examination, and we receive k-implies here with k=3. Cosine, SCML and CNN softmax measure pair wise likeness data, yet don't acquire the mapped directions of tests. In this manner, we don't perform grouping on these strategies

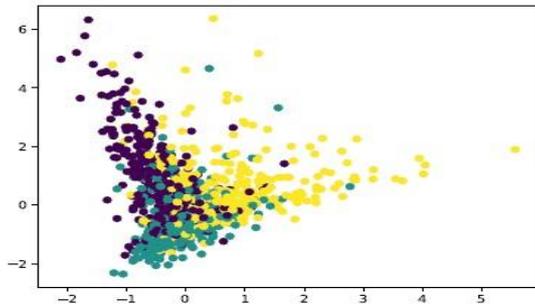


Fig. 5: LMMN based patient clustering result

LMNN, ITML and CNN triplet delineate information to another space and after that figure the Euclidean separation between sets, with the goal that we can perform grouping on the adapted new space.

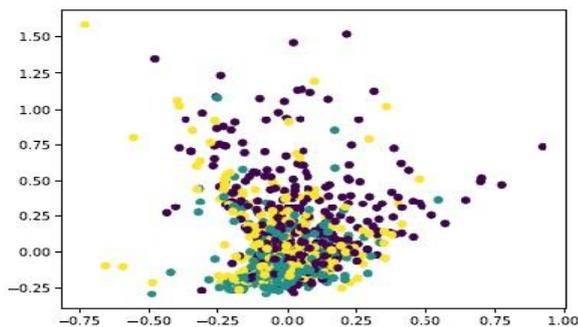


Fig. 6: ITML based clustering result

It is understandable that ITML and LMMN provide better clustering than Euclidean methods but the CNN triplet method, introduced herein performs better in terms of clustering and characterization of groups of patients.

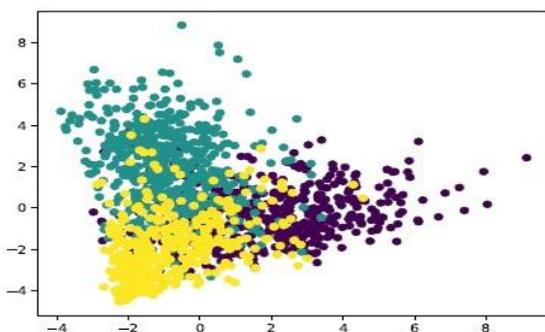


Fig. 7: A visualization of CNN-triplet based clustering result in the transformed space

V. CONCLUSION

Understanding comparability learning means to discover suitable separation metrics to quantify quiet combines for a particular undertaking. To catch the authentic data of patient record, a legitimate method to speak to longitudinal data records is essential. In addition, we require an approach to take in the similitude level or separation across each combination of patients. Herein, we introduced a patient comparability learning structures on the Health Record dataset. The crude patient data are sent as an input into a Convolution Neural Network demonstrate that catches the back to back successive data to take in a vector portrayal. At that point triplet loss depended separation, a metric learning strategy is utilized to take in the comparability of patient sets. Test results on the possibility of disease forecast and patient grouping demonstrate that Convolutional Neural Network can all the more likely to speak to the longitudinal medical data, and our proposed systems can efficiently cluster patients into different disease groups. In future, we can further include additional distance learning methods like GMML, ITML and a large broad dimensional dataset can be utilized. This model can be further refined and applied to broad dimensional datasets and can be used for disease forecast

REFERENCES

1. Zhang, P., Wang, F., Hu, J., & Sorrentino, R.(2014). Towards Personalized Medicine: Leveraging Patient Similarity and Drug Similarity Analytics. AMIA Summits on Translational Science Proceedings, 2014, 132–136.
2. Lee J, Maslove DM, Dubin JA (2015) Personalized Mortality Prediction Driven by Electronic Medical Data and a Patient Similarity Metric, PLoS ONE 10(5): e0127428.
3. A. Gottlieb, G. Y. Stein, E. Ruppim, R. B. Altman, and R. Sharan.(2013) "A method for inferring medical diagnoses from patient similarities," BMC Medicine,11(1).
4. Kasabov, N.K. (2007). Global, local and personalized modeling and pattern discovery in bioinformatics: An integrated approach. Pattern Recognition Letters, 28, 673-685.
5. C. Che, C. Xiao, J. Liang, B. Jin, J. Zho, and F. Wang (2017),A RNN architecture with dynamic temporal matching for personalized predictions of Parkinson's disease, Proceedings of SIAM International Conference on Data Mining. SIAM, 2017, 198-206.
6. Zhang, M. and Zhi-Hua Z. (2014), A review on multi-label learning algorithms. Knowledge and Data Engineering, IEEE Transactions 26(8): 1819-1837.
7. J. Xu, J. Zhou, P.N Tan,(2015).Formula: Factorized multi-task learning for task discovery in personalized medical models, SDM 2015,199-203.
8. Deng Cai, Xiaofei He, Kun Zhou, Jiawei Han, Hujun Bao(2007), Locality sensitive discriminate analysis, Proceedings of the 20th international joint conference on Artificial intelligence, p.708-713, January 06-12, 2007, Hyderabad, India
9. Yu Cheng, Fei Wang, Ping Zhang, and Jianying Hu,(2016). Risk Prediction with Electronic Health Records: A Deep Learning Approach Proceedings of the 2016 SIAM International Conference on Data Mining (SDM'16). 432-440.
10. Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun 2017. GRAM: Graph-based Attention Model for Healthcare Representation Learning, Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17). ACM
11. Peter B Jensen, Lars J Jensen, and Søren Brunak.(2012).Mining Electronic Health Records: Towards Better Research Applications and Clinical Care. Nature Reviews Genetics Vol. 13(6), 395-405.

A Convolution Neural Network Based Framework for Similarity Learning in Healthcare

12. Phuoc Nguyen, Truyen Tran, Nilmini Wickramasinghe, and Svetha Venkatesh (2016). DeepR: A Convolutional Net for Medical Records. *IEEE Journal of Biomedical and Health Informatics*, Vol 12(9),13-21.
13. Sun, F. Wang, J. Hu, and S. Ebadollahi (2012) ,Supervised patient similarity measure of heterogeneous patient records, *ACM SIGKDD Explorations Newsletter*,10-15.
14. Q. Suo, F. Ma, Y. Yuan, M. Huai, W. Zhong, A. Zhang, and J. Gao,(2017), Personalized disease prediction using a CNN-based similarity learning method, *Bioinformatics and Biomedicine (BIBM)*, *IEEE International Conference* ,121-126.
15. M. Zhan, S. Cao, B. Qian, S. Chang, and J. Wei, (2016),Low-rank sparse feature selection for patient similarity learning, *Data Mining (ICDM)*, *2016 IEEE 16th International Conference* , *IEEE*, 2016.
16. A. Sharafoddini, J. A. Dubin, and J. Lee, (2017),Patient similarity in prediction models based on health data: a scoping review," *JMIR medical informatics*, 46-52.
17. Y. Sha, J. Venugopalan, and M. D. Wang,(2016),A novel temporal similarity measure for patients based on irregularly measured data in electronic health records, *Platelets*, 34(2) 12-19..
18. M. Huai, C. Miao, Q. Suo, Y. Li, J. Gao, and A. Zhang (2018),Uncorrelated patient similarity learning, *SIAM International Conference on Data Mining*, Vol(12) 191-195.
19. Q. Suo, H. Xue, J. Gao, and A. Zhang (2016) Risk factor analysis based on deep learning models, *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*,pp 394-403.
20. Fenglong Ma, Jing Gao, Qiuling Suo, Quanzeng You, Jing Zhou, and Aidong Zhang,(2018). Risk Prediction on Electronic Health Records with Prior Medical Knowledge. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. ACM, New York, NY, USA, 1910-1919. DOI: <https://doi.org/10.1145/3219819.3220020>.
21. Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. (2017). Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*. ACM, New York, NY, USA, 1903-1911.
22. Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning method for epileptic seizure detection using short-time Fourier transform, *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. ACM, 2017,pp-213-222
23. Fenglong Ma, Chuishi Meng, Houping Xiao, Qi Li, Jing Gao, Lu Su, and Aidong Zhang. 2017. Unsupervised Discovery of Drug Side-Effects from Heterogeneous Data Sources. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*. ACM, New York, NY, USA, 967-976
24. Y. Yuan, G. Xun, Q. Suo, K. Jia, and A. Zhang,(2017),Wave2vec: Learning deep representations for bio signals," in *Data Mining (ICDM)*,16th International Conference on *IEEE* , pp. 1159-1164..
25. Y. Yuan, G. Xun, K. Jia, and A. Zhang, (2017),A novel wavelet-based model for EEG epileptic seizure detection using multi-context learning, *Bioinformatics and Biomedicine (BIBM)*, 2017 *IEEE International Conference* ,pages:694-699
26. F. Wang and J. Sun,(2015),Psf: a unified patient similarity evaluation frame-work through metric learning with weak supervision," *IEEE Journal of biomedical and health informatics* , 19(3):1053-60.
27. F. Wang, J. Sun, and S. Ebadollahi (2011),Integrating distance metrics learned from multiple experts and its application in patient similarity assessment, *Proceedings of SIAM International Conference on Data Mining*. *SIAM*,pp:59-70.
28. R. Snyderman(2012),Personalized healthcare: from theory to practice," *Biotechnology Journal* .
29. C. L. Overby, J. Pathak, O. Gottesman, K. Haerian, A. Perotte, S. Murphy, K. Bruce, S. Johnson, J. Talwalkar, Y. Shen (2013) ,A collaborative approach to developing an electronic health record phenotyping algorithm for drug-induced liver injury, *Journal of the American Medical Informatics Association*.
30. K. Ng, J. Sun, J. Hu, and F. Wang(2015),Personalized predictive modelling and risk factor identification using patient similarity, *AMIA Summits on Translational Science Proceedings*, pp :132-136.
31. Z. Zhu, C. Yin, B. Qian, Y. Cheng, J. Wei, and F. Wang(2016), Measuring patient similarities via a deep architecture with medical concept embedding, *Data Mining (ICDM)*, *IEEE 16th International Conference*, pp. 749-758.
32. E. Choi, M. T. Bahadori, E. Searles, C. Coffey, M. Thompson, J. Bost, J. Tejedor-Sojo, and J. Sun, (2016),Multi-layer representation learning for medical concepts, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1495-1504 .
33. Y. Kim(2014), Convolutional neural networks for sentence classification, *arXiv preprint arXiv:1408.5882*, 2014.
34. H. O. Song, S. Jegelka, V. Rathod, and K. Murphy(2017),Deep metric learning via facility location, *Computer Vision and Pattern Recognition (CVPR)*, 2017. pp. 2206-2214.
35. E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart(2014), Retain: An interpretable predictive model for healthcare using reverse time attention mechanism,, *arXiv preprint arXiv:1608.05745*.
36. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin (2016), Tensor flow: Large-scale machine learning on heterogeneous distributed systems, *arXiv preprint arXiv:1603.04467*, 2016
37. D. Kingma and J. Ba,(2014), Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*.
38. K.Q. Weinberger and L. K. Saul (2009) ,Distance metric learning for large margin nearest neighbor classification, *Journal of Machine Learning Research*, Vol (10) pp. 207-244.