Journal of Advances in Mathematical & Computational Sciences
An International Pan-African Multidisciplinary Journal of the SMART Research Group
International Centre for IT & Development (ICITD) USA
© Creative Research Publishers
Available online at https://www.isteams.net/ mathematics-computationaljournal.info
DOI: dx.doi.org/10.22624/AIMS/MATHS/V10N1P6
CrossREF Member Listing - https://www.crossref.org/06members/50go-live.html

Patient Similarity Analysis Using the Hierarchical Temporal Memory with Adaptive Look-Ahead Function

[1]Enyindah, Promise, [2]Oghenekaro, Linda Uchenna & [3]Osegi, Emmanuel Ndidi
Department of Computer Science, University of Port Harcourt, Port Harcourt, Nigeria
Department of Computer Science, National Open University of Nigeria, Lagos, Nigeria
E-mail: promise.enyindah@uniport.edu.ng, linda.oghenekaro@uniport.edu.ng, emmaosegi@gmail.com

ABSTRACT

This paper presents an approach to Patient Similarity Analysis (PSA) based on a machine intelligence system and technology called the Hierarchical Temporal Memory (HTM). The Patient Similarity Metric (PSM) used is a novel scoring metric known as the overlap. In order to prove this concept, the overlap scoring metric in a HTM Spatial Pooler (HTM-SP) prediction system was applied to real world patient's medical data. Two external HTM parameters, the context and the look-ahead were introduced. A variant of HTM-SP was also introduced based on an adaptive look-ahead parameter. The performance of the standard and adaptive HTM-SP were evaluated and compared with a Neuro Case Rule (NCR) in terms of Percentage Correct Value (PCV). The reported PCV (TRUE, FALSE) estimates showed (80.83, 46.66), (87.41, 3.33), and (100.00, 40.00), for the standard HTM-SP, the adaptive HTM-SP and the NCR approaches respectively.

Keywords — Machine Intelligence, Patient Similarity Analysis, Prediction, Overlap

Enyindah, P., Oghenekaro, L.U. & Osegi, E.N. (2022): The Feasibility of using E-Learning System for Educational Delivery in Post Covid-19 Era: A Case Study of Selected Public Secondary Schools in Niger-Delta, Nigeria. Journal of Advances in Mathematical & Computational Sciences. Vol. 9, No. 2. Pp 51-60. DOI: dx.doi.org/10.22624/AIMS/MATHS/V10N1P6
Available online at www.isteams.net/mathematics-computationaljournal.

1. INTRODUCTION

The Patient Similarity Analysis (PSA) is an important step in the field of medical diagnosis, where Electronic Medical Records (EMR) are used for effective, timely and personalized level based treatment of diseases. The time for diagnosis of a diseased patients and the need for laboratory screening can be reduced marginally by using the similarity of symptomatic features previously recorded in patient's medical data.

PSA techniques uses tools of machine learning, artificial intelligence algorithms combined with pure statistical or logical models to reduce the feature size and in the process discover useful features that can assist in making precise medical decision. Most recent research directions in PSA focuses on prediction accuracy and computational run time of the similarity or matching algorithm used in Patient Similarity Metric (PSM). Thus, the performance evaluation of any PSA scheme is determined by the effectiveness of its PSM algorithm. Also artificial intelligence algorithms that have implicit support for similarity relations with a good foundation in biology and neuroscience holds better promises for automated medical diagnosis.

In this paper, an emerging neural technique that have the potential to solve a diversity of medical related problems while not losing focus on some of the core biological components of the brain, is presented. This technique, framework and tool for machine intelligence is called the Hierarchical Temporal Memory (HTM) and has been widely applied by many researchers in diverse fields of human endeavor.

2. PREVIOUS WORKS

The field of Patient Similarity Analysis (PSA) is currently receiving serious attention. In [1], a Patient Similarity Metric (PSM) called SQL-cosine algorithm was proposed for analysis of medical datasets. Their PSM method was implemented on the PostgreSQL and MonetDB Databases using both the column and row oriented structures respectively. The Optimization technique was based on batching to speed up computation. Their system was evaluated on two widely utilized Electronic Medical Records (EMR) and the result showed that the column oriented structures though fast in speed still have challenges of batch improvement factors. In [2], the prediction of occurrence of Acute Hypotensive Episode (AHE) using Patient Similarity Metric (PSM) based on Locally Supervised Metric Learning (LSML) and the Mahalanobis distance was proposed.

Wang et al [3] developed a framework called the SimProX system for the early prognosis of disease using Euclidean distance as similarity metric. In [4], a Bag-of-Words features framework for automatic annotation of disease labels of multi-source patient data in Intensive Care Units (ICU). The proposed model framework focused on learning disease correlation locally in the patient data, learning similarity between patients in the label space rather than the feature space. Their system uses disease correlations in similarity measurements, and an alternating algorithm to optimize an objective function based on the Squared Hinge Loss (SHL).

Morid et al [5] applied the K-Nearest Neighbor algorithm to ICU Mortality predictions. They developed the Patients-based Prediction Modeling Framework (PPMF) which utilized aggregation classifiers. Their similarity metric uses a direct access technique to sample data instances in a multidimensional Euclidean space. Their proposed technique was shown to outperform a machine learning technique used in a competition (MIMIC-III) dataset. In order to build mortality prediction models, Lee et al [6] used a cosine similarity patient similarity metric (PSM) for the identification of patient similarities. Their reported results showed that the proposed PSM technique will outperform existing ICU severity of illness scores; however, this may be at the cost of the computational complexity of their proposed PSM technique.

In [7], a similarity based approach is compared to a predictive modeling based approach for Mortality prediction using MIMIC-II ICU dataset. The similarity approach used was the K-Nearest Neighbor which requires the computation of a Euclidean distance while the predictive modeling approach used was the Logistic regression technique with L2-regularization. They reported results using Area Under the Curve (AUC) and the Cox model for earliest prediction time analysis. Their results showed that the prediction model will outperform the similarity model.

In [8], a supervised and unsupervised clustering framework was used for medication planning. Their PSA approach used the Mahalanobis distance metric for making ICU predictions on a sample cohort data. Performance results were based on Specificity, Sensitivity, F1 measure, Accuracy and AUC. The results in [8.Table 2] indicate that the supervised approach is better than the unsupervised. Gottlieb et al [9] predicted the eventual discharge diagnoses using the Jaccard, Euclidean distance with the MATLAB logistic regression classifier including 10-fold cross-validation tests on Electronic Health Record (EHR) data. They reported results using AUC and F-measure in the MATLAB language [9.Table 1]. Their results indicate that combining two different EHR datasets leads to a degraded performance which is attributed to a phenomenon called “information loss” or feigned knowledge.

A Neuro-Case-Rule approach that combines fuzzy inference techniques with neural networks has been proposed in [10] for the diagnosis of Hepatitis A virus using a real world patient data. The proposed NCR technique was able to give a 100% Percentage Correct Value (PCV) true estimate and a PCV false estimate of 40%. Among these techniques proposed in the literature, the neural approaches have potential usefulness in EMR systems and remain favorable candidates for performing PSA tasks; however, the question remains:

Can we have a neural technique that is general enough to be applied across all medical domains and not limited by excessive hyper-parameter tuning in addition to having desirable property of online (continual) learning of streaming medical data?

In this research, we propose a PSA scoring metric called the “overlap”. The overlap is based on the notion that the similarity or correlatedness between two entities or data vectors may be measured by the sum of all the dot products of the individual constituent elements greater than a pre-specified overlap threshold. This approach is very favorable to data vectors that possess sparse distributed binary representations. This concept will be further described in the subsequent section.

3. PROPOSED METHODOLOGY

In this section we present the details of the proposed methodology including the patient similarity scoring metric, a method for building causal relations and some specific system parameters.

3.1 Hierarchical Temporal Memory

The Hierarchical Temporal Memory (HTM) is a machine-intelligence but biologically constrained computational model of the neocortex, the seat of intelligence in the brain [11]. The HTM is specifically good in solving real-world continual learning (or streaming) tasks. The primary operation of HTM is based on the formation of Sparse Distributed Representations (SDRs) and then learning to make predictions from these representations using ideas of biology and neuroscience.

HTM is implemented as a suite of algorithms called the Cortical Learning Algorithms (CLA) which represent the core HTM technology or framework [11, 12]. The HTM-CLA consists of two key parts: a HTM Spatial Pooler (SP) part for forming Sparse Distributed Representations from real world or synthetic data and a HTM Temporal Pooler (TP) part for making predictions on the SDRs formed from the HTM-SP. The SDR architecture follow the notion of sparse coding earlier suggested for learning sensory inputs in [13, 14].

A typical neuron model used in a HTM system implementation is given in Fig.1. This model is inspired by neuroscientific studies of activity-dependent synaptogenesis – a phenomenon that hypothesizes the growth and origin of the biological synapses stimulated by external sensory signals [15]. In this diagram, the connections consisting of the green bulbs represent proximal synapses which are typically linearly summed to produce feed-forward activation; distal synapses are represented by segments of blue bulbs that are logically summed to give a dendrite spiking neuron activation when they exceed a certain recognition threshold value as indicated by the Sigma sign. Feedback and context experiences are formed using the distal connections through time.

The HTM-SP comprise of a number of operating parameters that assist in the discovery of a local or global solution space in the continual learning paradigm. In this study we are interested in the study of two additional key parameters: the look-ahead and context parameters; these parameters are described succinctly in a latter sub-section (see Section D).

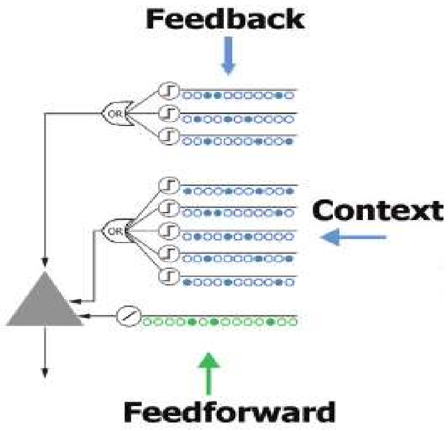

Fig. 1: An HTM neuron. Adapted from [16].

3.2 Patient Similarity Scoring Metric

As mentioned earlier, the Patient Similarity Metric (PSM) is used based on a scoring metric called the overlap. The concept of overlap is as shown in Fig.2. To compute the overlap or match, we require a set of internally generated HTM synapses formed from the spatial pooling and an input vector. In the figure (Fig.2), both HTM synapses and input vector are represented as binary sparse distributed representations (SDRs) – the input binary vector is actually generated from a real time dataset by using a suitable scalar encoder as described in [17].

The overlap is computed as in (1):

$$o_i = b_i \sum_j W_{ij} z_j \qquad (1)$$

where,

b_i = is a positive boost factor for exciting each HTM-SP mini-column

z_j = input pattern vector seen by the generative HTM neuron

i, j = mini-column indexes comprise of a number of operating parameters that assist in the discovery of a local or global solution space in the continual learning paradigm. In this study we are interested in the study of two additional key parameters: the look-ahead and context parameters; these parameters are described succinctly in a latter sub-section.

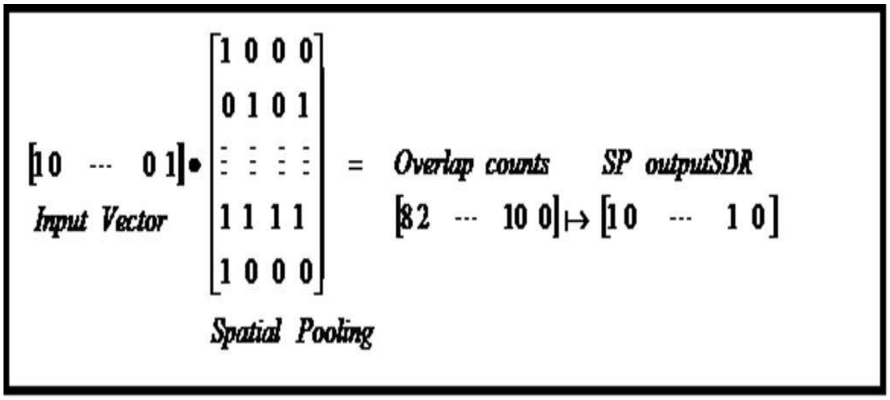

Fig 2: An illustrative example of Overlap processing in an HTM-SP; Source: Ahmad & Hawkins [18].

3.3 Temporal Aggregation Technique

The temporal aggregation technique (TAT) for data pre-processing is based on the concept of cause-and-effect data relations; in the TAT, the SDR formed from the Hierarchical Temporal Memory Spatial Pooler (HTM-SP) stage is transformed into a pair-wise mixed integer representation and then passed over to an overlapping temporal classification (OTC) stage for post-processing.

The temporal aggregation technique is performed as follows:

- **Step1:** A single-column vector matrix of length 1:*N* having a width of 1 is formed. The variable *N* represents the number of sampled sequences (sample size of SDRs in row-wise orientation) obtained from the HTM-SP stage. The elements in this matrix also contain the indexes for temporal aggregation.
- **Step2:** For each element formed in the matrix generated in Step 1 that is greater than 1, a modulus operation is performed such that when there exists a remainder, the element in consideration is skipped, otherwise it is selected; the operation in this step results in single-column matrix of length equivalent to half the sample size of the SDRs used in the

aforementioned step. The elements in this matrix also contain a set of even indexes in the matrix obtained from Step1 at time instance, *t*. This set is denoted $A_{t(1)}$.

- **Step3:** For all elements in the set $A_{t(1)}$, form a concatenation of $A_{t(1)}$ with $A_{t-1(1)}$ 1-step behind as $\{A_{t(1)} \;\; A_{t-1(1)}\}$; this concatenation step generates the temporal aggregator index set. We call this set of indexes $A_{t(agg)}$.
- **Step4:** Using $A_{t(agg)}$ as index sequence, extract SDR patterns obtained from the HTM-SP stage in a temporal aggregated fashion and then perform overlap temporal classification through time.

3.4 System Parameters under Study

3.4.1 Adaptation of the Look-Ahead parameter

The look-ahead parameter plays the important role of performing multiple-step forecast in a prediction system. In the HTM-SP prediction system, look-ahead parameter is adaptively fine-tuned in an iterative manner as sample sequences are formed using the model formula in (2) as:

$$n_{step(adapt)} = n_{step}^{1/8} \qquad (2)$$

where,

n_{step} = the HTM-SP system look-ahead parameter

By close inspection, we see that this formula is equivalent to taking the square root of the look-ahead parameter three times.
Similarity Metric (PSM) is used based on a scoring metric called the overlap.

3.4.2 Context Parameter

The context parameter represents the number of past sequences to consider during spatial pooling. As the spatial pooling is a continual learning system, the context parameter is applied continually during this phase. It is also possible to apply explicit context to the spatial pooling phase i.e. we may bias temporally the considered sequences from some other external source.

4. EXPERIMENTAL DETAILS

4.1 Systems Architecture and Performance metrics

The proposed systems architecture for hepatitis predictions is shown in Fig.3. In this architecture, the encoder transforms real world input data signals (comprising patient hepatitis data) into a binary representation suitable for spatial pooling; then the Hierarchical Temporal Memory Spatial Pooler (HTM-SP) forms sparse distributed representations (SDRs) of the binary representations using the similarity matching metric described in Section III sub-section B. The SDRs are temporally aggregated and predicted using the Overlapping Temporal Classification (OTC) scheme described in Section III sub-section C. The metrics chosen for evaluating performance includes the Mean Absolute Percentage Error (MAPE), and the classification percentage value (PCV) i.e. the percentage of correct prediction states (TRUE and FALSE states). The Mean Absolute Percentage Error (MAPE) was chosen as the metric for evaluating the performance of the HTM-SP system.

The use of MAPE is due to its insensitivity to outliers so it presents an unbiased metric when compared to other metrics such as Mean Squared Error (MSE) or absolute errors which exhibit very large estimated deviations for increasing/decreasing predicted values - an obvious disadvantage when performing comparative evaluations of different techniques.

The percentage correct value (PCV) estimates was used for comparing with the results obtained in (Obot & Uzoka, 2009). The PCV estimate gives a good indication of the percentage classification and is simple to interpret by non-experts. It can be interpreted as PCV (TRUE) for the percentage of correctly classified logical 1 states and PCV (FALSE) for the percentage of correctly classified logical 0 states where the logical 1 and 0 states represent the desired hepatitis state.

Accordingly, the experimental tests were conducted using the patient dataset provided in (Obot & Uzoka, 2009: Appendix A). which was obtained from a real world survey of three hospitals located in Nigeria. The parameters used for the HTM-SP simulations in the proposed system can be found in the Appendix.

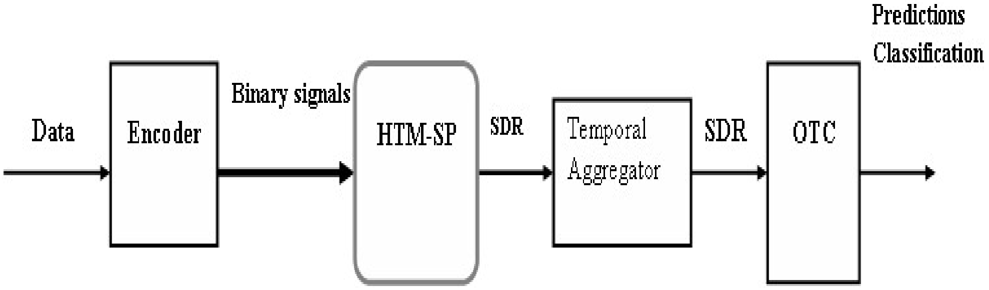

Fig 3: HTM-SP System for performing the PSA simulations.

4.2 Prediction Classification Task

The classification uses the hepatitis dataset provided in (Obot & Uzoka, 2009). The task is to continually and adaptively predict n-step ahead, the condition of a patient (has hepatitis or not) using a set of similarity relations built from a history of past patient data comprising encoded hepatitis data attributes. In addition, the dataset is presented sequentially instead of using traditional techniques such as bagging or random split into a separate and test dataset as in [4].

5. RESULTS AND DISCUSSIONS

5.1 Results

The prediction result of a typical simulation run using HTM-SP in MATLAB is as shown in Fig.4. The plot describes a moving Mean Absolute Percentage Error (mMAPE) for the considered medical dataset using default settings (see Appendix). The percentage average correct value (PCV estimates) results of the standard HTM-SP using the best setting of the context and look-ahead parameters are as shown in Table 1. These values consist of average PCV (TRUE) and PCV (FALSE) estimates. Also shown is the average PCV estimate of the HTM-SP with adaptive look-ahead and the average PCV estimate using the NCR technique as reported in [10]. These estimates are obtained after 10 simulation trial runs.

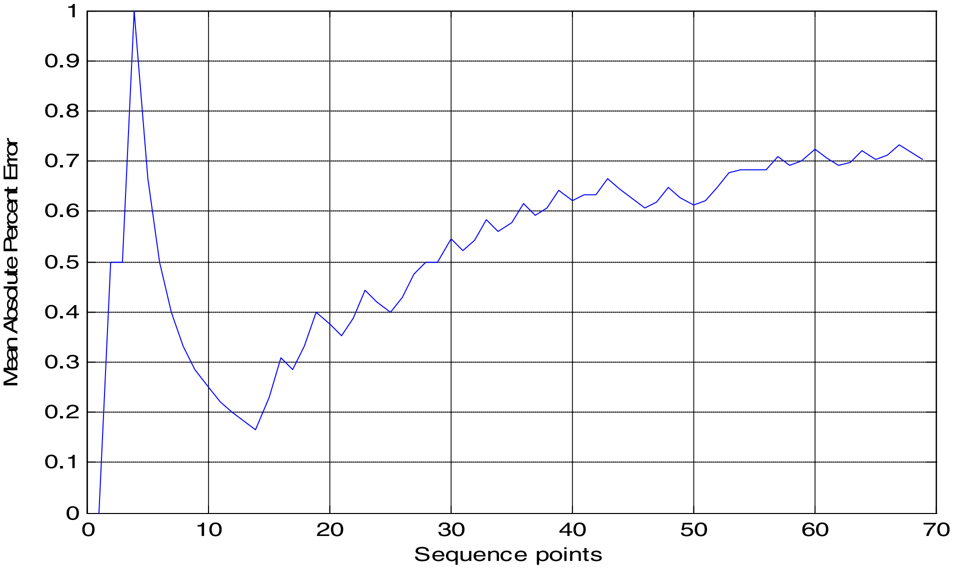

Fig 4: Moving average plot of the HTM-SP for 1-step ahead prediction

TABLE 1: PCV (TRUE, FALSE) ESTIMATES COMPARING THE HTM, ADAPTIVE HTM AND THE NCR TECHNIQUES

PCV VALUES (%)		
PCV	PCV	PCV
(80.83. 46.66)	(87.41, 3.33)	(100.00, 40.00)

Table 2; HTM-SP SYSTEM PARAMETERS

Parameter	*Default Values*
Number of columns	250
Initial Synaptic permanence	0.21
Dendritic segmant activation threshold	2
Boost factor	100
Synapse permanent increment	0.1
Synapse permanent decrement	0.01
Context	2
Look-ahead	3

5.2 Discussion of Results

The prediction results in Table 1 show that the proposed standard and adaptive HTM technique compares favorably with the Neuro-Case-Rule (NCR) approach reported in [10]. As indicated by the (PCV) true estimates, NCR attained the best performance with 100% PCV true estimate. The standard HTM-SP outperforms both the adaptive HTM-SP and the NRC with higher PCV false estimates.

However, the adaptive HTM-SP is promising and has better generalization ability in PCV true estimate over the standard one when the context and look-ahead parameters are set to 2 and 3 respectively. It remains a matter of further research if the standard/adaptive HTM-SP techniques will play a useful role for next generation patient similarity analysis.

6. CONCLUSION AND FUTURE WORK

This presentation proposed an emerging state-of-the-art machine intelligence technique called the Hierarchical Temporal Memory (HTM) for patient similarity analysis (PSA) of Electronic Medical Records (EMR). HTM does not require a separate training and test dataset and has internal mechanisms for handling continual predictions – a very important property for real time medical analysis. The implementation uses a special HTM Spatial Pooler (HTM-SP) system with a unique temporal pooling strategy based on a matching technique called overlap.

In addition, two explicit HTM-SP parameters called the context and look-ahead parameters with a useful technique for building cause and effect relations called the temporal aggregation technique (TAT) were introduced. These techniques were applied to the similarity analysis of a reported hepatitis patient dataset and compared with the results obtained from the Neuro-Case-Rule. In particular, HTM generalized better than the Neuro-Case-Rule technique as long as the look-ahead and context number parameters are carefully chosen. HTM also has inherent mechanisms for generating a similarity process using the concept of overlap scoring.

In future, the performance of HTM overlap similarity metric including the influence of the context and look-ahead parameters for PSA on diverse medical datasets will further be investigated and study on other kinds of datasets requiring some sort of similarity measurements conducted. Other interesting research directions may also include anomaly detection in similarity measurements or similarity data graphs [19], predictive representational learning systems using deep learning [20], developing strategies for handling missing data particularly for unstructured (textual) EMR data (see the works of Wells et al [21]) and including the use of swarm prediction and optimization algorithms for finding reliable and highly accurate patient similarity predictions.

REFERENCES

[1] A. Tashkandi, I. Wiese, and L. Wiese, "Efficient In-Database Patient Similarity Analysis for Personalized Medical Decision Support Systems," Big Data Research, May 2018 (in press).

[2] Sun, D. Sow, J. Hu, and S. Ebadollahi, "A System for Mining Temporal Physiological Data Streams for Advanced Prognostic Decision Support," 2010 IEEE International Conference on Data Mining, Dec. 2010, pp. 1061–1066

[3] F. Wang, J. Hu, and J. Sun, "Medical prognosis based on patient similarity and expert feedback," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 1799–1802.

[4] S. Wang, X. Li, L. Yao, Q. Z. Sheng, and G. Long, "Learning multiple diagnosis codes for ICU patients with local disease correlation mining," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, no. 3, p. 31, 2017.

[5] M. A. Morid, O. R. L. Sheng, and S. Abdelrahman, “PPMF: A Patient-based Predictive Modeling Framework for Early ICU Mortality Prediction,” *arXiv preprint arXiv:1704.07499*, 2017.

[6] J. Lee, “Patient-specific predictive modeling using random forests: An observational study for the critically ill,” *JMIR medical informatics*, vol. 5, no. 1, 2017.

[7] M. Hoogendoorn, A. El Hassouni, K. Mok, M. Ghassemi, and P. Szolovits, “Prediction using patient comparison vs. modeling: A case study for mortality prediction,” in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, 2016, pp. 2464–2467.

[8] M. Panahiazar, V. Taslimitehrani, N. L. Pereira, and J. Pathak, “Using EHRs for heart failure therapy recommendation using multidimensional patient similarity analytics,” *Studies in health technology and informatics*, vol. 210, p. 369, 2015.

[9] A. Gottlieb, G. Y. Stein, E. Ruppin, R. B. Altman, and R. Sharan, “A method for inferring medical diagnoses from patient similarities,” *BMC medicine*, vol. 11, no. 1, p. 194, 2013.

[10] O. U. Obot and F.-M. E. Uzoka, “A framework for application of neuro-case-rule base hybridization in medical diagnosis,” *Applied Soft Computing*, vol. 9, no. 1, pp. 245–253, 2009.

[11] J. Hawkins, S. Ahmad, and D. Dubinsky, “Hierarchical temporal memory including HTM cortical learning algorithms”. Techical report,Numenta,Inc,Palto-Alto. https://web.archive.org/web/20110714213347/http://www.numenta.com/htm-overview/education/HTM_CorticalLearningAlgorithms.pdf

[12] J. Hawkins, S. Ahmad, S. Purdy, and A. Lavin, “Biological and machine intelligence (bami),” *Initial online release 0.4*, 2016.

[13] B. A. Olshausen and D. J. Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” *Nature*, vol. 381, no. 6583, p. 607, 1996.

[14] B. A. Olshausen and D. J. Field, “Sparse coding of sensory inputs,” *Current opinion in neurobiology*, vol. 14, no. 4, pp. 481–487, 2004.

[15] K. Zito and K. Svoboda, “Activity-dependent synaptogenesis in the adult mammalian cortex,” *Neuron*, vol. 35, no. 6, pp. 1015–1017, 2002.

[16] Y. Cui, S. Ahmad, and J. Hawkins, “The HTM spatial pooler—a neocortical algorithm for online sparse distributed coding,” *Frontiers in computational neuroscience*, vol. 11, p. 111, 2017.

[17] S. Purdy, “Encoding data for HTM systems,” *arXiv preprint arXiv:1602.05925*, 2016.

[18] S. Ahmad and J. Hawkins, “Properties of sparse distributed representations and their application to hierarchical temporal memory,” *arXiv preprint arXiv:1503.07469*, 2015.

[19] Y. Sebastian, J. T. Chew, X. T. Tiong, V. Raman, A. Y. Y. Fong, and P. H. H. Then, “Anomaly detection from diabetes similarity graphs using community detection and Bayesian techniques,” in *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*, 2018, p. 32.

[20] Q. Suo *et al.*, “Deep Patient Similarity Learning for Personalized Healthcare,” *IEEE Transactions on NanoBioscience*, 2018.

[21] B. J. Wells, K. M. Chagin, A. S. Nowacki, and M. W. Kattan, “Strategies for handling missing data in electronic health record derived data,” *eGEMs*, vol. 1, No. 3, 2013.