

## Prediction of Incident Dementia Using Patient Temporal Health Status

Sunyang Fu M.H.I<sup>a</sup>, Omar A. Ibrahim, Ph.D.<sup>a</sup>, Yanshan Wang, Ph.D.<sup>a</sup>, Maria Vassilaki M.D.,Ph.D.<sup>b</sup>, Ronald C. Petersen, M.D., Ph.D.<sup>bc</sup>, Michelle M. Mielke, Ph.D.<sup>bc</sup>, Jennifer St Sauver Ph.D.<sup>b</sup>, Sunghwan Sohn Ph.D.<sup>a</sup>

<sup>a</sup> Department of Artificial Intelligence and Informatics Research, Mayo Clinic, Rochester, Minnesota, USA

<sup>b</sup> Department of Quantitative Health Sciences, Mayo Clinic, Rochester, Minnesota, USA

<sup>c</sup> Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA

### Abstract

*Dementia is one of the most prevalent health problems in the aging population. Despite the significant number of people affected, dementia diagnoses are often significantly delayed, missing opportunities to maximize life quality. Early identification of older adults at high risk for dementia may help to maximize current quality of life and to improve planning for future health needs in dementia patients. However, most existing risk prediction models predominantly use static variables, not considering temporal patterns of health status. This study used an attention-based time-aware model to predict incident dementia that incorporated longitudinal temporal health conditions. The predictive performance of the time-aware model was compared with three traditional models using static variables and demonstrated higher predictive power.*

### Keywords:

dementia, deep learning, machine learning

### Introduction

Aging of the population has led to an increase in dementia. Alzheimer's disease is the most common cause of dementia, with more than 6 million people in the United States currently affected, and cases are expected to increase to 15 million by 2060 [1]. These trends have resulted in a tremendous burden for patients, their families, society, and healthcare systems, with an annual estimate of 18.5 billion hours of unpaid care at a value of \$234 billion [2]. Despite the huge number of people affected by dementia, clinicians are not aware of relevant cognitive impairment (i.e., mild cognitive impairment [MCI] and dementia) in more than 40% of their patients [3]. Further, clinical diagnoses often occur late in the process of cognitive decline or after opportunities for maximizing quality of life have long passed [4, 5]. Given this imminent growth of older adults with dementia and its significant underdiagnosis, predicting dementia risk and understanding its progression is crucial to help the aging population with their health needs.

Dementia is usually a slowly progressing disease over time. One promising approach for dementia prediction is to examine longitudinal trajectories of various clinical assessments (e.g., gait speed, activities of daily living, neuropsychological characteristics) among cognitively normal patients and patients with dementia. Studies have shown that trajectory analyses may detect signals much earlier than the dementia is first clinically diagnosed [6, 7]. Our preliminary studies also demonstrated distinct temporal patterns of activities of daily living between

cognitively normal and impaired older adults years before clinical diagnosis [8, 9].

Existing models for risk prediction have been largely static-time models (i.e., not considering temporal patterns of health status) [10-12]. Despite some early success in modeling patients' future risk, the informative longitudinal patterns and dynamic changes in patient's health status were not captured. With the recent advancement in sequential learning, recurrent neural networks (RNN) and its variants have been applied to a wide range of temporal-based datasets leveraging their advantage of learning complex nonlinear relationships and sequential patterns [13].

One important characteristic of electronic health record (EHR) data is the availability of a detailed record of longitudinal disease progression. Because dementia is a slowly progressing disease, the changes in patients' clinical assessment during their long-term follow-up visits may provide crucial information to predict dementia. However, since previous experiments only considered diagnosis codes as the model input in the experiment [14-16], the applicability and feasibility of these models may be limited in a real-world EHR setting given the fact that patient information can be multimodal, diverse, and dynamic [17].

A well-known model incorporating time information includes time-aware long short-term memory neural network (T-LSTM), a variation of LSTM proposed by Baytaş et al that can capture the time interval between two consecutive elements of a visit [14]. Another approach is through the attention mechanisms for risk prediction, such as RETAIN [15] and RetainEX [18], which leveraged the reverse-time attention mechanism to consolidate historical visits and significant clinical variables.

The recent advancement in hierarchical learning and Transformer architecture, hierarchical deep sequence models with attention mechanisms, has shown some early success in various EHR-related prediction tasks [16, 19, 20]. Therefore, this study used state of the art hierarchical attention-based time-aware model incorporating temporal patient health status to predict dementia. In addition, to address previous model limitations, the model used in this study considers heterogeneous input features in both static and dynamic characteristics. Finally, we evaluated the model in a population-based cohort with a comprehensive periodic cognitive assessment of MCI and dementia.

## Materials and Methods

### A. Data

This study was approved by the Mayo Clinic Institutional Review Board and the Olmsted Medical Center Institutional Review Board. We used data from the Mayo Clinic Study on Aging [21]. The MCSA is a prospective population-based cohort study with comprehensive periodic cognitive assessments (at baseline and repeated every 15 months), initiated in 2004 to investigate the epidemiology of MCI. Eligible persons from Olmsted County, Minnesota, population, were randomly selected and evaluated comprehensively in person using the clinical dementia rating scale, a neurological evaluation, and neuropsychological testing. A consensus committee used previously published criteria to diagnose the participants with normal cognition, MCI, or dementia. MCI is diagnosed according to published criteria [22] and dementia is diagnosed according to DSM-IV criteria [23]. In addition to cognitive assessment, other data elements are abstracted from the medical records or through interviews and questionnaires (e.g., education, body mass index, comorbid conditions, neuropsychiatric symptoms, vascular risk factors). The MCSA cohort comprises 6,185 unique patients with total 26,807 visits (4.3 average visits per patient). Among these, 3,070 patients were female (49.6%) and 729 patients (11.6%) have progressed to dementia. The median age of the cohort is 73.

The variables collected from the MCSA were used to predict the status of dementia. Table 1 summarizes 40 input variables into five different categories, including patient demographics, physical characteristics, psychological characteristics, social characteristics, and functional status. All physical characteristics, psychological characteristics, and functional status were considered time-dependent variables.

Table 1– Input Variables

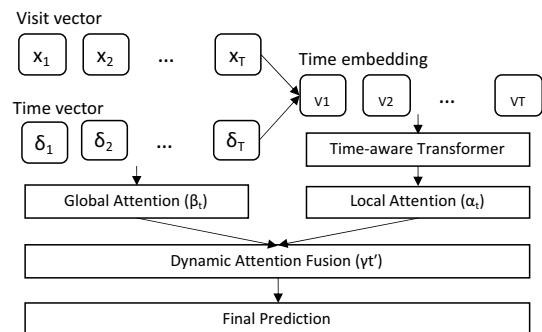
<b>Patient Demographics</b>
Age, Sex, Race, Ethnicity
<b>Physical characteristics</b>
BMI, Smoking status, Alcohol problem
Sleep apnea, Hypertension, Dyslipidemia, Atrial fibrillation, Angina, Congestive heart failure, Coronary artery disease, Myocardial infarction, Coronary artery bypass graft, Diabetes, ESSscore
<b>Psychological characteristics</b>
Delusions, Hallucinations, Agitation, Depression, Anxiety, Euphoria, Apathy, Disinhibition, Irritability/lability, Motor behavior, Nighttime behavior, Appetite/eating change, BDI-II grand total, BDI depression (Total $\geq 13$ ), BAI total (0-63)
<b>Social characteristics</b>
Education, Occupation, Marital status, Personal care
<b>Functional status</b>
FAQ Total Score (0-30), ECog-12

**Abbreviations:** ESSscore: Hypersomnolence ESS score (0-24), BDI: Beck Depression Inventory scores, BAI: Beck Anxiety Inventory, FAQ: 10-item questionnaire on instrumental activity of daily living, ECog-12: scales to measure multiple cognitively relevant everyday abilities, covering six domains

### B. Model

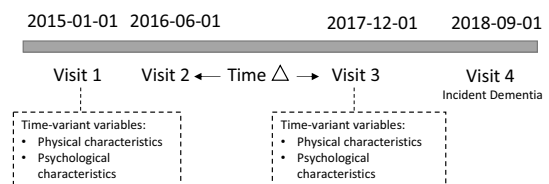
Transformer, originally proposed in the tasks in natural language processing, has demonstrated robustness and capability to capture long-term sequential events [24]. With recent research applying Transformer to model EHR data, the model can capture the dynamic interaction between consecutive visits for risk prediction leveraging the self-attention mechanism. We used time-aware Transformer to predict the risk of progression to dementia in the future state. In our study, we explored three different transformer-based architectures, including the original HiTANet (Hierarchical Time-Aware Attention Networks - H-Net) proposed by Luo et al [16] and its variants Transformer Time Embedding (T-EMD) and Transformer Time Attention (T-ATT). For the given visit, the H-Net considers two vector representations: a diagnosis code vector  $x_i$  and time interval  $\delta_i$ . This representation is able to learn a visit-specific weight through local attention score  $\alpha_i$ . At the global level, H-Net uses a time-aware key-query attention mechanism to study the overall disease progression. Additionally, the model uses dynamic attention fusion (DAF), the fusion mechanism to combine the local and global level attention score  $\alpha_i$  and  $\beta_i$  for each visit. T-EMD, and T-ATT are the simplified version of H-Net that only used either time embedding (local visit analysis) or time-aware key-query attention for global weight adjustment (Figure 1).

Figure 1– Architecture of H-Net



The original HiTANet was designed for risk prediction based on terminology codes only. To comprehensively capture the complex patient disease status, we used the input representation as time-dependent generic features. Let  $p = \{1, 2, \dots, P\}$  where the input vector consists of a sequence of follow-up visits and each visit contains concepts such as functional status, comorbidities, and demographic information (Figure 2). All variables were then converted to ordinal scales using either integer encoding or a binary vector.

Figure 2– Illustration of Longitudinal Disease Progression Patterns



### C. Patient Representation (temporal vs. static)

We used variables in all MCSA visits prior to incident dementia and all visits for non-dementia cases to develop temporal-based models incorporating dynamic changes of health status (H-Net, T-EMD, T-ATT). Static models, which used variables in the last visit prior to incident dementia and the last visit for non-dementia cases, were implemented for comparison (see D. Experiment section).

### D. Experiments

To evaluate the effectiveness of the proposed model, three baseline models (static models using static variables) were developed, such as multilayer perceptron (MLP), random forest (RF), and gradient boosting machine (GBM). For all models, we used a random split of 70% for development (8:2 ratio of training and validation) and 30% as the test set for evaluation. We used accuracy, precision, recall, the area under the ROC curve (AUC), and F1-score as a performance metric. F1-score is a popular metric for imbalanced data with a focus on the positive class by combining precision and recall. Missing values were imputed using missForest [25].

### Results

The summarized performances for experiments across different models are shown in Table 2. We observed that the temporal models (H-Net and T-ATT) yielded higher performance than static models in recall, AUC, and F1-score, except for very close AUC to RF. T-EMD yielded the highest performance in accuracy, however, given that the distribution of the data set was imbalanced (dementia cases < 12%), accuracy may not reflect the overall performance reasonably. The temporal models produced relatively higher recall but lower precision than those of static models. The use of temporal information (i.e., use of all visit information over time) in temporal models may enable to capture more dementia cases despite the sacrifice of the model precision.

Table 2 – Model Performance in Dementia Prediction

	Acc	Pre	Rec	AUC	F1
Static Model					
MLP	0.784	<b>0.710</b>	0.420	0.658	0.528
RF	0.781	<b>0.750</b>	0.398	<b>0.684</b>	0.520
GBM	<b>0.786</b>	0.708	0.391	0.665	0.504
Temporal Model					
H-Net	0.773	0.662	<b>0.463</b>	<b>0.681</b>	<b>0.545</b>
T-EMD	<b>0.827</b>	0.676	0.401	0.673	0.503
T-ATT	0.757	0.595	<b>0.501</b>	<b>0.681</b>	<b>0.544</b>

\*Acc: accuracy, Pre: precision, Rec: recall, F1: F1-score, AUC: Area under the ROC Curve, bold font denotes top two highest performed models

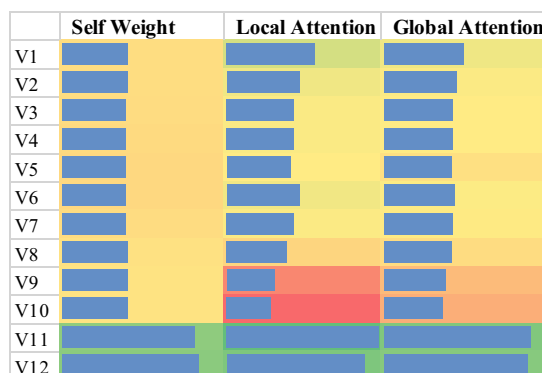
### Discussion

Dementia is highly prevalent and associated with severe health outcomes in the aging population. However, its diagnoses are often significantly delayed. To facilitate the early detection and risk prediction of dementia using patients' longitudinal health conditions, we have explored an attention-based time-aware mechanism and transformer architecture and compared them with other traditional models.

The evaluation results indicated the importance of leveraging time-variant information for modeling dementia risk in longitudinal data. Transformer-based models performed higher than other models in an F1-score, which is more appropriate than accuracy or AUC to measure the performance in imbalanced data. Compared to other studies for prediction of incident dementia, our models produced promising results; other studies using deep neural networks [26] produced an F1-score of 25% – 30% (3 to 8 years prior to the index date), and using Lasso logistic regression had an AUC 0.69 (1.5% dementia cases), sensitivity 9.9% and specificity 99.9% (4 to 5 years in advance) [27].

To further understand the impact of different attention (local vs. global) mechanisms on the attention weights learned by three Transformer models, we compared and visualized three attention weights in the training process: 1) self-weight, 2) self-weight adjusted by the local attention (allow to focus on local region), and 3) self-weight adjusted by the global attention (trained based on global time vector). As shown in Figure 3, all three representations demonstrated a non-linear decreasing order of the weights from the most recent visit to the first visit. The local attention demonstrated a more aggressive pattern of adjusting the weight sequence for both recent and long-distance visits. The global attention alleviates some high degree adjustments (areas with color changes) such as v1, v9, and v10 and neutralizes the overall attention weights. Based on the performance evaluation (Table 2), we believe that the combination of two attention mechanisms may potentially improve the generalizability of the model.

Figure 3 – Architecture of H-Net



\*V1-12: visit sequence prior to the incident dementia, color red: lower weight assigned, color green: higher weight assigned, blue bar: weight value

We also tested an unsupervised stacked denoising autoencoder [28] – i.e., a deep patient representation to reconstruct the input vector from noisy and sparse data using a deep sequence of non-linear transformations – to examine its efficacy on the models. However, we did not observe a performance gain compared

with the original vector representation. This may be due to relatively less complexity/variability of the MCSA data that can be benefited by denoising transformation.

The limitation of this study includes the use of variables in the MCSA based on comprehensive assessments that may not be readily available in other institutions. We plan to extract relevant variables from routine EHRs to facilitate broad applicability. The model was developed using the data in a single institution and thus the external validity is warranted to assess its generalizability.

## Conclusions

The transformer-based time-aware models using longitudinal visit information demonstrated higher performance in dementia prediction compared to traditional models that used static variables. This warrants the use of temporal patterns of health conditions in dementia prediction, reflecting the nature of a slowly progressing disease over time. In the future, we also plan to explore a cohort-specific model (i.e., age, sex, and socioeconomic status) to investigate its efficacy in dementia prediction. Considering the significant growth of dementia and its underdiagnosis, this predictive model of dementia risk could help the aging population with their health needs and better planning.

## Acknowledgments

This study was supported by NIA R01 AG068007 and NIAID R21 AI142702. The Mayo Clinic Study of Aging was supported by National Institutes of Health (NIH) Grants U01 AG006786, P50 AG016574, R01AG057708, the GHR Foundation, the Mayo Foundation for Medical Education and Research and was made possible by the Rochester Epidemiology Project (R01 AG034676).

## Address for correspondence

\*Corresponding Author: Sunghwan Sohn, Ph.D., E-mail: Sohn.Sunghwan@mayo.edu

## Disclosures

Maria Vassilaki has received research funding from Roche and Biogen; she currently consults for Roche, receives research funding from NIH, and has equity ownership in Abbott Laboratories, Johnson and Johnson, Medtronic, and Amgen. Jennifer St. Sauver has received research funding from Exact Sciences to study colorectal cancer. Michelle M. Mielke has consulted for Biogen and Brain Protection Company and receives research funding from NIH and DOD.

## References

- [1] Brookmeyer, R., Abdalla, N., Kawas, C.H., and Corrada, M.M.: 'Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the United States', *Alzheimer's & Dementia*, 2018, 14, (2), pp. 121-129
- [2] Gaugler, J., James, B., Johnson, T., Marin, A., and Weuve, J.: '2019 Alzheimer's disease facts and figures', *ALZHEIMERS & DEMENTIA*, 2019, 15, (3), pp. 321-387
- [3] Chodosh, J., Petitti, D.B., Elliott, M., Hays, R.D., Crooks, V.C., Reuben, D.B., Galen Buckwalter, J., and Wenger, N.: 'Physician recognition of cognitive impairment: evaluating the need for improvement', *Journal of the American Geriatrics Society*, 2004, 52, (7), pp. 1051-1059
- [4] Prince, M., Bryce, R., and Ferri, C.: 'Alzheimer's Disease International World Alzheimer Report 2011 the benefits of early diagnosis and intervention Executive Summary', Published by Alzheimer's Disease International (ADI), 2011
- [5] Iliffe, S., Manthorpe, J., and Eden, A.: 'Sooner or later? Issues in the early diagnosis of dementia in general practice: a qualitative study', *Family Practice*, 2003, 20, (4), pp. 376-381
- [6] Dodge, H.H.: 'Temporal patterns of change in clinical variables leading to MCI', *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 2010, 6, (4), pp. S92-S93
- [7] Hu, K., Riemersma-Van Der Lek, R.F., Patxot, M., Li, P., Shea, S.A., Scheer, F.A., and Van Someren, E.J.: 'Progression of dementia assessed by temporal correlations of physical activity: results from a 3.5-year, longitudinal randomized controlled trial', *Scientific reports*, 2016, 6, pp. 27742
- [8] Goudarzvand, S., Sauver, J.S., Mielke, M.M., Takahashi, P.Y., and Sohn, S.: 'Analyzing Early Signals of Older Adult Cognitive Impairment in Electronic Health Records', in Editor (Ed.) (Eds.): 'Book Analyzing Early Signals of Older Adult Cognitive Impairment in Electronic Health Records' (IEEE, 2018, edn.), pp. 1636-1640
- [9] Goudarzvand, S., St Sauver, J.L., Mielke, M.M., Takahashi, P.Y., and Sohn, S.: 'Early Temporal Characteristics of Elderly Patient Cognitive Impairment in Electronic Health Records', *BMC Med Inform Decis Mak*, 2019, 19, (149)
- [10] Kunutsor, S., Whitehouse, M., Blom, A., and Beswick, A.: 'Systematic review of risk prediction scores for surgical site infection or periprosthetic joint infection following joint arthroplasty', *Epidemiology & Infection*, 2017, 145, (9), pp. 1738-1749
- [11] Sohn, S., Larson, D.W., Habermann, E.B., Naessens, J.M., Alabbad, J.Y., and Liu, H.: 'Detection of clinically important colorectal surgical site infection using Bayesian network', *Journal of Surgical Research*, 2017, 209, pp. 168-173
- [12] Chen, D., Afzal, N., Sohn, S., Habermann, E.B., Naessens, J.M., Larson, D.W., and Liu, H.: 'Postoperative bleeding risk prediction for patients undergoing colorectal surgery', *Surgery*, 2018, 164, (6), pp. 1209-1216
- [13] Che, Z., Purushotham, S., Cho, K., Sontag, D., and Liu, Y.: 'Recurrent neural networks for multivariate time series with missing values', *Sci*, 2018, 8, (1), pp. 1-12
- [14] Baytas, I.M., Xiao, C., Zhang, X., Wang, F., Jain, A.K., and Zhou, J.: 'Patient subtyping via time-aware lstm networks', in Editor (Ed.) (Eds.): 'Book Patient subtyping via time-aware lstm networks' (2017, edn.), pp. 65-74
- [15] Choi, E., Bahadori, M.T., Kulas, J.A., Schuetz, A., Stewart, W.F., and Sun, J.: 'Retain: An interpretable predictive model for healthcare using reverse time attention mechanism', *arXiv preprint arXiv:1608.05745*, 2016
- [16] Luo, J., Ye, M., Xiao, C., and Ma, F.: 'HiTANet: Hierarchical Time-Aware Attention Networks for Risk Prediction on Electronic Health Records', in

- Editor (Ed.)^(Eds.): ‘Book HiTANet: Hierarchical Time-Aware Attention Networks for Risk Prediction on Electronic Health Records’ (2020, edn.), pp. 647-656
- [17] Fu, S., Leung, L.Y., Raulli, A.-O., Kallmes, D.F., Kinsman, K.A., Nelson, K.B., Clark, M.S., Luetmer, P.H., Kingsbury, P.R., and Kent, D.M.: ‘Assessment of the impact of EHR heterogeneity for clinical research through a case study of silent brain infarction’, *BMC Med. Informatics Decis. Mak.*, 2020, 20, pp. 1-12
- [18] Kwon, B.C., Choi, M.-J., Kim, J.T., Choi, E., Kim, Y.B., Kwon, S., Sun, J., and Choo, J.: ‘Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records’, *IEEE Trans Visual Comput Graphics*, 2018, 25, (1), pp. 299-309
- [19] Li, Y., Rao, S., Solares, J.R.A., Hassaine, A., Ramakrishnan, R., Canoy, D., Zhu, Y., Rahimi, K., and Salimi-Khorshidi, G.: ‘BEHRT: transformer for electronic health records’, *Sci*, 2020, 10, (1), pp. 1-12
- [20] Choi, E., Xu, Z., Li, Y., Dusenberry, M., Flores, G., Xue, E., and Dai, A.: ‘Learning the graphical structure of electronic health records with graph convolutional transformer’, in Editor (Ed.)^(Eds.): ‘Book Learning the graphical structure of electronic health records with graph convolutional transformer’ (2020, edn.), pp. 606-613
- [21] Roberts, R.O., Geda, Y.E., Knopman, D.S., Cha, R.H., Pankratz, V.S., Boeve, B.F., Ivnik, R.J., Tangalos, E.G., Petersen, R.C., and Rocca, W.A.: ‘The Mayo Clinic Study of Aging: design and sampling, participation, baseline measures and sample characteristics’, *Neuroepidemiology*, 2008, 30, (1), pp. 58-69
- [22] Petersen, R.C.: ‘Mild cognitive impairment as a diagnostic entity’, *Journal of internal medicine*, 2004, 256, (3), pp. 183-194
- [23] Association, A.P.: ‘Diagnostic and statistical manual of mental disorders. 4th ed.’ (Washington: American Psychiatric Association, 1994. 1994)
- [24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I.: ‘Attention is all you need’, *arXiv preprint arXiv:1706.03762*, 2017
- [25] Stekhoven, D.J., and Bühlmann, P.: ‘MissForest—non-parametric missing value imputation for mixed-type data’, *Bioinformatics*, 2012, 28, (1), pp. 112-118
- [26] Nori, V.S., Hane, C.A., Sun, Y., Crown, W.H., and Bleicher, P.A.: ‘Deep neural network models for identifying incident dementia using claims and EHR datasets’, *Plos one*, 2020, 15, (9), pp. e0236400
- [27] Nori, V.S., Hane, C.A., Martin, D.C., Kravetz, A.D., and Sanghavi, D.M.: ‘Identifying incident dementia by applying machine learning to a very large administrative claims dataset’, *PLoS One*, 2019, 14, (7), pp. e0203246
- [28] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., and Bottou, L.: ‘Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion’, *J*, 2010, 11, (12)