Individualized Prediction of Depressive Disorder in the Elderly: A Multitask Deep Learning Approach

Article *in* International Journal of Medical Informatics · September 2019

DOI: 10.1016/j.ijmedinf.2019.103973

5 authors, including:

Zhongzhi Xu
City University of Hong Kong
14 PUBLICATIONS **41** CITATIONS
SEE PROFILE

Paul Siu fai Yip
The University of Hong Kong
694 PUBLICATIONS **63,076** CITATIONS
SEE PROFILE

International Journal of Medical Informatics 132 (2019) 103973

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Individualized prediction of depressive disorder in the elderly: A multitask deep learning approach

Zhongzhi Xu[a], Qingpeng Zhang[a,*], Wentian Li[b], Mingyang Li[c], Paul Siu Fai Yip[d]

[a] School of Data Science, City University of Hong Kong, Hong Kong, China
[b] Wuhan Hospital for Psychotherapy, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei, China
[c] Department of Industrial and Management Systems Engineering, The University of South Florida, Tampa, USA
[d] Centre for Suicide Research and Prevention and the Faculty of Social Sciences, The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:
Depressive disorder prediction
Depression
Deep learning
Patient progression model

ABSTRACT

Introduction: Depressive disorder is one of the major public health problems among the elderly. An effective depression risk prediction model can provide insights on the disease progression and potentially inform timely targeted interventions. Therefore, research on predicting the onset of depressive disorder for elderly adults considering the sequential progression patterns is critically needed.
Objective: This research aims to develop a state-of-the-art deep learning model for the individualized prediction of depressive disorder with a 22-year longitudinal survey data among elderly people in the United States.
Methods: We obtain the 22-year longitudinal survey data from the University of Michigan Health and Retirement Study, which consists of information on 20,000 elderly people in the United States from 1992 to 2014. To capture temporal and high-order interactions among risk factors, the proposed deep learning model utilizes a recurrent neural network framework with a multitask structure. The C-statistic and the mean absolute error are used to evaluate the prediction accuracy of the proposed model and a set of baseline models.
Results: The experiments with the 22-year longitudinal survey data indicate that (a) machine learning models can provide an accurate prediction of the onset of depressive disorder for elderly individuals; (b) the temporal patterns of risk factors are associated with the onset of depressive disorder; and (c) the proposed multitask deep learning model exhibits superior performance as compared with baseline models.
Conclusion: The results demonstrate the capability of deep learning-based prediction models in capturing temporal and high-order interactions among risk factors, which are usually ignored by traditional regression models. This research sheds light on the use of machine learning models to predict the onset of depressive disorder among elderly people. Practically, the proposed methods can be implemented as a decision support system to help clinicians make decisions and inform actionable intervention strategies for elderly people.

1. Objective

Depressive disorder (or clinical depression) is a prevalent and serious mood disorder worldwide. By 2020, depressive disorder is expected to be the second contributor to disease burden [1]. By 2030, depressive disorder will be the leading cause of disability [2]. Among the elderly people, depressive disorder is the most common psychiatric disorder and it has been becoming increasingly prevalent [3]. Depressive disorder also reduces an elderly person's ability to rehabilitate [4]. Accordingly, the National Institute of Mental Health considers depression in older adults as a major public health problem that accounts for significant and growing health care expenditures [5]. In addition, depression disorder is usually associated with the elevated risk of other diseases (e.g., cardiac diseases) and mortality among the elderly [6]. Therefore, individualized early detection of depression is critical for the mental and physical well-beings in elderly people [4]. Effective and individualized prediction of the onset of depressive disorder can timely inform intervention strategies to prevent depressive disorder in the elderly and further reduce healthcare costs.

Before a clinical diagnosis of depressive disorder can be made, elderly people's pathophysiologic changes may already distinguish those who will eventually become depressed from those who will not [4]. However, less than 10% of depressed elderly patients eventually receive appropriate treatment [7], due to the low recognition rate of depressive disorder of both the patients and the primary care physicians [8]. Therefore, for the sake of effective and proactive disease management,

* Corresponding author.
E-mail address: qingpeng.zhang@cityu.edu.hk (Q. Zhang).

https://doi.org/10.1016/j.ijmedinf.2019.103973
Received 25 March 2019; Received in revised form 27 August 2019; Accepted 17 September 2019

there is a critical need for predicting symptoms and risk of depression for individual elderly people. However, as many risk factors for depressive disorder (such as recent stressful events, early-life stressors and social status changes) [9–12] does not exhibit observable physical symptoms and evidence, the prediction of depression remains a challenging task [4]. In addition, capturing event sequences has been useful in identifying the risk of certain diseases (e.g., heart failure) [13] and may be useful in predicting the onset of depressive disorder.

This research aims (a) to explore the feasibility of predicting the onset of depressive disorder for elderly adults by identifying critical pathophysiologic pathways that lead to depression from a large-scale 22-year longitudinal household survey data in the United States and (b) to further investigate whether capturing the temporal relations among variables can improve the performance in predicting depressive disorder compared with conventional methods that ignore such temporal information.

To model time-stamped pathophysiologic patterns in the survey data, we adopt the long short-term memory (LSTM) recurrent neural network, a popular deep learning framework for modeling event sequences. LSTM was initially proposed as a language model to capture word sequence in text [14]. It has been successfully applied to a wide variety of healthcare/medical problems, such as early detection of heart failure onset [13], identification of relations in clinical notes [15], and mining e-cigarette-related adverse events [16]. We also compared the performance of LSTM with that of traditional machine learning models. In doing so, we demonstrate the advantage of such sequential models over traditional machine learning models in the temporal prediction of depressive disorder.

2. Background and significance

Extensive works on detecting depressive disorder exist [4,10,12–14], but most of them are either based on behavioral logs collected by pervasive computing devices, such as smartphones and straps, or on current pathophysiologic data. Although the studies have been successful, the predictability of depression disorder has not been well examined. In addition, existing research on depression prediction has focused on predicting the prevalence of depression for a population, and not on individual patients [15–18]. Population-based predictions are useful for high-level decision/policy making, such as resource allocation, budget planning, and national campaigns. However, a population-based prediction cannot directly help the identification of individuals who are at a high risk of becoming depressed. Therefore, research on individualized prediction of depressive disorder is urgently needed [4].

Moreover, temporal information, which is commonly existent in medical and healthcare datasets, is often ignored in existing regression-based depression prediction research [21–24]. A recent research [4] successfully applied the LSTM model to predict severely depressed moods (including depression) based on self-reported histories. However, this research relied on a limited set of information collected through a smartphone application in a relatively short period of time (three weeks). This factor caused two limitations: (a) many depression-related risk factors, such as socioeconomic status and historical mental health conditions, were not considered and (b) the long-term effect of various factors were not captured.

To the best of our knowledge, no research on using long-term longitudinal data to predict the depressive disorder for elderly individuals is available. Thus, the present research aims to fill this gap by developing an individualized LSTM depression prediction model in elderly people using a 22-year longitudinal household survey data in the United States.

3. Materials and methods

3.1. Description of HRS data

The Health and Retirement Study (HRS) is a longitudinal household survey data set for the study of retirement and health among elderly people in the United States. HRS has 12 waves every two years from 1992 to 2014 with the exception of an extra wave in 1995. It contains clean variables with consistent naming conventions. The survey contains rich information of demographics, income, assets, health, cognition, family structure and connections, healthcare utilization and costs, housing, job status and history, expectations, and insurance (refer to Table 1 for samples of partial variables).

3.2. Risk for depressive disorder

HRS consists of a representative set of items from a concise version of the commonly used Center for Epidemiologic Studies Depression Scale (CES-D) to measure the risk for depressive disorder [33]. The CES-D measure in HRS data ranges from 0 to 8, indicating the existence of eight risk factors for depressive disorder. The higher the CES-D score, the higher the risk for depressive disorder. In practice, we usually identify people whose CES-D score is among the top 20 percentile as those at risk for depressive disorder [34]. In the HRS data, the distribution of CES-D score (see the distribution of Wave 12 in Fig. S2(A)) indicates that the threshold for the top 20 percentile is 3. Thus, we label the participants whose CES-D score is smaller than 3 as risk free and others as depressed. A cutoff value of 3 has also been used in other studies with the same CES-D scoring system (eight-item CES-D score) [34].

3.3. Multitask LSTM model

The recent advances of deep learning techniques have introduced new opportunities to deal with many healthcare problems that involve complex high-dimensional clinical and behavioral data. [14–35], The original form of deep learning is the multilayer perceptron (MLP), a class of feedforward artificial neural network with one or more hidden layers. However, MLP is not suitable to deal with the problems considering sequential data (such as predicting depression disorder with longitudinal survey or clinical data), because of its inability to capture the inter-dependency of hidden states sequentially. To capture such sequential inter-dependency, the recurrent neural network (RNN) with hidden long short-term memory (LSTM) units is proposed [36]. The LSTM-based RNN model is capable of capturing the patterns of the entire sequences of data, and has been proved to be a powerful model for learning from sequential data [37]. Details of the mechanisms of LSTM model are included in Supplementary Information.

We have two prediction tasks on an elderly person:

- Task 1: To predict whether he or she will become depressed in two years.
- Task 2: To predict the actual CES-D score.

Instead of building two models to separately solve the two tasks, we adopt the multitask learning framework in which multiple tasks are trained together with the same model. The advantages of this framework include the following: (a) the estimation of parameters is better generalized because part of the model is shared across different tasks and (b) the model structure is simple [38]. Fig. 1 shows the architecture of our model. Risk factors that were used to indicate the incidence of depression disorder by previous clinical studies [22–32] are recursively used as the inputs for the LSTM component. They are also observed to be statistically associated with the current stage depression conditions of individuals in the HRS data (Table 1). The outputs of the LSTM component are concatenated with the auxiliary non-temporal variables (*Age* and *Gender*). The newly concatenated feature vector is then fed into a MLP component to capture nonlinear interactions among the features. The components up to this part are shared by Tasks 1 and 2. Then, we feed the output of the first MLP component (shared) into two separated MLP components. The two separated MLP components are used to generate the prediction results for Task 1 and Task 2, respectively. In order to address potential overfitting problems, we use a 0–1

Table 1
Summary statistics of the training data and test data.

Risk factors	Cases (N = 2843) No. (%)	Controls (N = 2843) No. (%)	Odds ratio (95% CI)	P-value	Descriptions
Demographic					
Gender [25,26]	2073 (72.96%)	1821 (64.09%)	1.511 (1.350–1.691)	< .001	This is a binary variable indicating if the participant is a male (value is 1) or female (value is 0).
Age [26]			1.009 (1.003–1.015)	0.005	This variable denotes the age of the participant at the current stage.
Marriage status [26]	1212 (42.65%)	1728 (60.81%)	0.479 (0.431–0.533)	< .001	A respondent's marriage status is binary. It is 1 if the participant is married, but 0 for those widowed/divorced/separated participants.
Health related rick factors					
The Mobility Index [27]			1.611 (1.554–1.671)	< .001	This index is the number of activities that the participant has difficulty in doing, including walking one block, walking several blocks, walking across a room, climbing one flight of stairs, and climbing several flights of stairs activities. For each task, 1 refers to difficulty and 0 refers to no difficulty.
Smoking cigarettes [28]	313 (11.01%)	210 (7.39%)	1.551 (1.292–1.863)	< .001	This variable indicates whether a respondent smokes at the current stage (1 = yes; 0 = no).
Self-report health [15,17,22]			2.365 (2.229–2.509)	< .001	This variable is the respondent's self-reported general health status, ranged from 1 (excellent) to 5 (worst).
Chronic illness related risk factors					
Cancer [30]	661 (23.25%)	575 (20.23%)	1.195 (1.053–1.356)	0.006	This variable indicates whether or not a doctor has ever told the respondent he/she got cancer (1 = yes; 0 = no).
Diabetes [31]	973 (34.22%)	683 (24.02%)	1.646 (1.466–1.848)	< .001	This variable indicates whether or not a doctor has ever told the respondent he/she got diabetes (1 = yes; 0 = no).
Socioeconomic status related risk factors					
Employment status [32]					
Not retired	254 (8.93%)	369 (12.98%)	Ref	< .001	This variable is 0 if a respondent has a full-time job.
Completely retired or unemployed	2283 (80.30%)	2038 (71.68%)	1.618 (1.485–1.838)		This variable is 2 if a respondent is completely retired or unemployed.
Partly retired	306 (10.77%)	436 (15.34%)	0.887 (0.696–1.073)		This variable is 1 if a respondent is partly retired.

norm to normalize the value of inputs. Table S1 shows the information regarding the parameters that we used to tune. To determine the best model, we try all possible combinations of parameters listed in Table S1 and select the best performed ones. In this study, the parameters with bond font in Table S1 are adopted. Although the parameters are tuned towards experiment with 12 look-back years, we apply the tuned parameters to experiments with varying look-back years as well.

To demonstrate the effectiveness of the proposed model, we adopt a fully connected MLP-based neural network, the single-task LSTM, the support vector machine (SVM), and the Dynamic Bayesian Network (DBN) models as the baselines for Task 1. In addition, we adopt a fully connected MLP-based neural network, the single-task LSTM, the LASSO-based generalized linear regression, and DBN models as the baselines for Task 2. To take time into account, in the DBN model, the risk for the current time period is not only dependent on the current values of the variables, but also the risk in the previous time periods. The detailed DBN structure is presented in the Supplementary Fig. S3. Given that other baseline models cannot model temporal data, we concatenate time series feature values to create a single feature vector, allowing them to learn a predictive model based on the same information.

The performance of the proposed multitask LSTM model is related to the length of look-back years. Empirically, the mental and physical health conditions of a person is related to the progression of his or her conditions in the past several years [23]. However, how far into the past we should look back is unclear. If the performance improves with more look-back years, then the significance of the benefits with the inclusion of more historical data needs to be investigated, because having access to fine-grained long-term longitudinal data in practice is usually difficult. Therefore, we characterize the relationship between the prediction performance of the multitask LSTM model and number of look-back years included in the model.

4. Experiment results

4.1. Data preprocessing

The data in the first four waves are excluded because the values of certain critical variables are missing. We use the data in waves 5–11 as the training set and those in wave 12 as the test set. More specifically, we train the parameters in the model using the data in waves 5–10 (number of look-back years is 6) to predict depression-related outcomes (Tasks 1 and 2) in wave 11 (year of 2012). Then, we use the trained model to predict depression-related outcomes in wave 12 (year of 2014) using the data in waves 6–11.

As shown in Fig. S2, the distribution of CES-D score is right-skewed, with the majority of individuals labeled as not depressed. Models trained with such imbalanced data can be biased and inaccurate. To solve this problem, we adopt a common under-sampling approach to randomly remove the data of individuals whose CES-D score is lower than 3. Eventually, we have 50% depressed individuals in both training sets (2690 individuals) and testing sets (2996 individuals), as shown in Table 1.

Different from traditional statistical models, which assume that variables are independent of one another, deep learning models can capture high-order interactions between variables and automatically select variables during the convergence process. Thus, there is no additional feature selection step for the proposed single-task LSTM and multitask LSTM models. For baseline models, LASSO regression can perform variable selection and regularization through a penalization scheme [39]. SVM and DBN can use an exhaustive search to perform feature selection and model regularization [40].

4.2. Model evaluation

Experimental results are shown in Fig. 2. We use the C-statistic, also named as area under the ROC curve, to evaluate the performance in

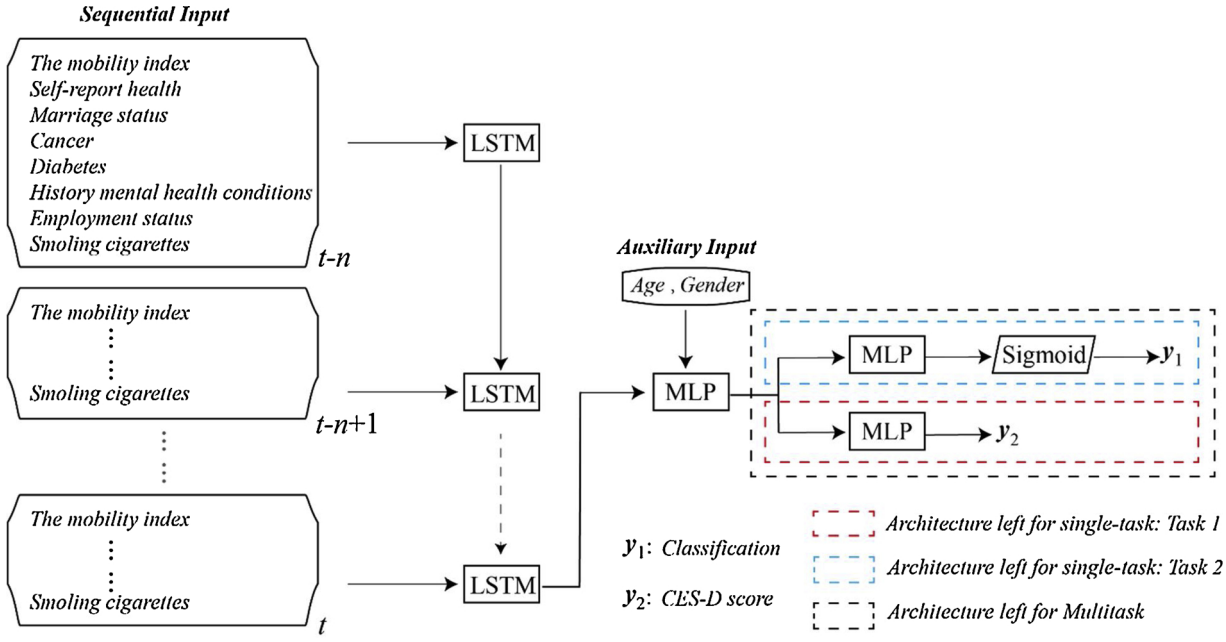

Fig. 1. The architecture of the single-task/multitask LSTM models.

Task 1 and the mean absolute error (MAE) to evaluate the performance in Task 2. To evaluate the confidence level of the predictions, the Bootstrap analysis technique (with replacement) is adopted [41]. Specifically, 80% of the predicted probability values and the corresponding labels are randomly bootstrapped. The sample size is then expanded to the original size through randomly picking from previous bootstrapped samples. This procedure is repeated 1000 times. We then have 1000 performance measures (C-statistic or MAE) from which we could calculate the 95% confidence interval.

In general, both the proposed deep learning models and baseline

A

Model	Task 1 (C-statistic)	Model	Task 2 (MAE)
MLP classifier	0.856(0.850-0.863)	MLP regression	1.469(1.438-1.491)
SVM classifier	0.846(0.838-0.851)	LASSO regression	1.347(1.319-1.377)
DBN	0.816(0.814-0.819)	DBN	1.465(1.434-1.487)
Single-task LSTM	0.868(0.862-0.873)	Single-task LSTM	1.322(1.299-1.353)
Multitask LSTM (without auxilary variables)	0.869(0.863-0.874)	Multitask LSTM (without auxilary variables)	1.292(1.275-1.307)
Multitask LSTM (full model)	**0.873**(0.868-0.878)	Multitask LSTM (full model)	**1.287**(1.270-1.301)

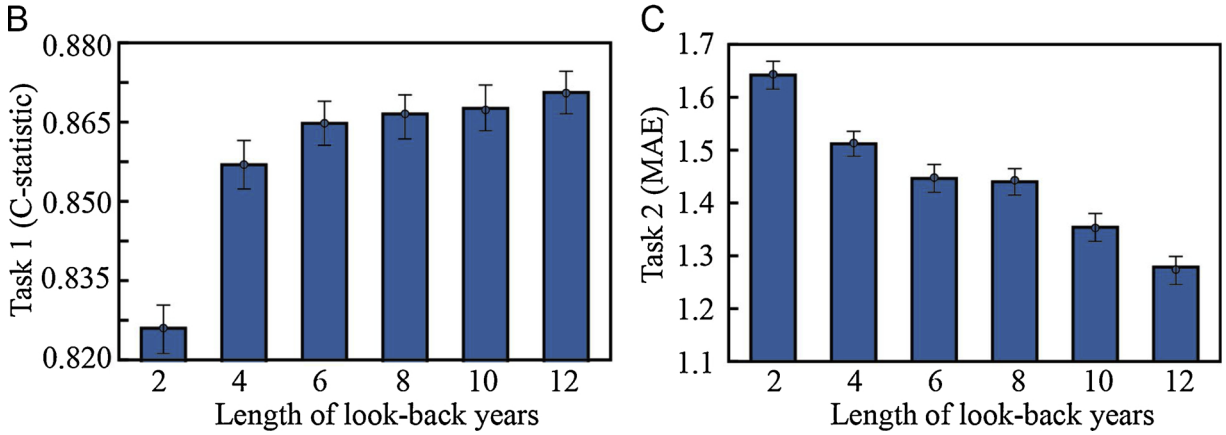

Fig. 2. Prediction performance: (A) The C-statistic for Task 1 and MAE for Task 2 of each model, with the 95% confidence interval; (B) The C-statistic of the proposed Multitask LSTM model with different lengths of look-back years; (C) The MAE of the proposed Multitask LSTM model with different lengths of look-back years.

Table 2
Predictions on each item.

Item	C-statistic
Felt Depressed	0.7764
Restless sleep	0.7489
Was happy	0.7548
Felt lonely	0.7906
Felt sad	0.8011
Could not get going	0.7705
Enjoyed life	0.7298
Everything was an effort	0.8098

models can perform prediction tasks with high accuracy, with a C-statistic of over 0.81 and MAE less than 1.47 (Fig. 2(A)). This observation indicates the feasibility of using machine learning models to predict depressive disorder among the elderly. Thus, the predictive power of the risk factors identified in the literature has been demonstrated.

The performance of MLP models is on par with the traditional statistical models (SVM, LASSO regression and DBN). In particular, MLP performs better than the SVM classifier in Task 1 (higher C-statistic) but worse than the LASSO regression in Task 2 (higher MAE). This finding indicates that a simple MLP framework without any calibration is not more advantageous than the traditional statistical models in these tasks, largely due to the limited sample size in healthcare and medical research. The DBN performed worse than the LSTM-based models, indicating that the temporal patterns in the data are more complex than an explicit DBN structure.

The single-task and multitask LSTM models outperform the MLP and statistical models in both tasks. LSTM's capability of modeling temporal patterns is proven to be helpful in predicting an individual's risk for becoming depressed. The multitask LSTM model has the best performance in both tasks, indicating that multitask framework can (a) simplify the learning process and more importantly (b) optimize the training of the shared LSTM layers by encoding the observed patterns in both tasks into the same structure.

To demonstrate the predictability of the auxiliary variables, we also performed the experiment using the Multitask LSTM model without these auxiliary variables. The resulted C-statistic and MAE are 0.869 and 1.292, respectively, indicating that the prediction performance is worse than the full model. From this experiment, we found that these auxiliary variables can enhance the predictability of the model.

In reality, the occurrence of depressive disorders among the elderly people is unbalanced. In the HRS dataset, there are 19.6% of elderly people labeled as with the risk of depressive disorders according to the CES-D score. This value is consistent with the overall depression rate among the elderly people in the US (15%–17%) [42,43]. To check the robustness of the Multitask LSTM model, we further evaluate its performance on an unbalanced test set, which includes all the valid data points in 2014: 1358 elderly people with the risk of depressive disorders and 5573 without the risk. The C-statistic (for Task 1) and MAE (for Task 2) are 0.869 and 1.086, respectively. The performance of the proposed model is further evaluated using precision, recall and Matthews correlation coefficient (MCC). Precision for positive/negative samples is 0.628/0.910; recall for positive/negative samples is 0.764/0.812. MCC is 0.696. It is worth noting that low precision and high recall on positive samples is fine because we won't miss many individuals at risk.

To characterize the influence of the number of look-back years utilized for prediction, we evaluate the performance of the proposed multitask LSTM model with multiple choices of look-back years (2, 4, 6, 8, 10, and 12 years). As shown in Fig. 2(B) and (C), as we increase the number of look-back years, the prediction performance improves. The improvement is greatly evident when the data in recent years are taken into consideration. This finding demonstrates that the mental health status of an elderly is not only related to his or her previous status, but also many years back. In practice, the inclusion of the past six years (three waves in the HRS data) can lead to a reasonably good prediction performance.

The eight items of the CES-D refer to the eight dimensions of depressive disorder. We conducted additional experiments to show that the proposed multitask LSTM model can predict the onset of depressive disorder in all dimensions. The value of the C-statistic ranges from 0.7298 to 0.8098, as presented in Table 2. Specifically, "Everything was an effort" is the most predictable item. This finding makes a lot of sense because the HRS data are focused on elderly people. The multifunctional (e.g., physical, mental and social) declines and losses of the elderly people make them highly vulnerable to such symptom of depression [44]. On the other hand, "Enjoyed life" and "Restless sleep" are relatively more difficult to predict partly because of the related factors (i.e. genetic causes for insomnia, non-marriage related social determinants of mental health) that are not included in the HRS data.

To inform actionable interventions and enhance mental well-being, we further examine the significance of the predictive variables. First, all variables are statistically significant with the current stage depression conditions with positive odds ratios (Table S3). In terms of prediction, CES-D_{t-1}, CES-D_{t-2}, CES-D_{t-3}, CES-D_{t-5} and CES-D_{t-4} are more predictive than other variables in the SVM model, and CES-D_{t-1}, CES-D_{t-2}, CES-D_{t-3}, CES-D_{t-5} and Self-report $health_{t-1}$ are more predictive in the LASSO model. No systematic way to unveil the high-order interactions within the neural networks exists due to the complex structure of deep learning models [45,46]. We then adopt a commonly used stepwise backward elimination approach [47] to evaluate the predictive power of these variables by excluding one type of variable each time. Then, we compare the resulting differences in prediction accuracy. The results are similar to the aforementioned tests for SVM and LASSO models (Table S4 in the Supplementary materials).

5. Discussions

The proposed multitask LSTM model exhibits superior performance in predicting the onset of depressive disorder (Task 1) and CES-D score (Task 2). Specifically, the deep neural network framework can characterize and incorporate temporal patterns and complicated high-order interactions among variables for model prediction. In addition, the shared LSTM layers can encode patterns observed in both tasks into the same structure. The analysis of look-back years suggests that access to the historical record of a patient's health condition is crucial for predicting the onset of depressive disorder in the future. Moreover, the improvement is evident when more recent data (within six years) are included.

Many risk assessment systems for various diseases exist, such as the CES-D for depression, Global Assessment of Functioning (GAF) system for a broad set of psychiatric illnesses [48], and systems for other diseases, such as Global Registry of Acute Coronary Events (GRACE) system [49] and a series of systems for diabetes [50]. These existing risk assessment systems are usually based on a simplified scoring function learned from statistical models. This simplified framework is effective in helping clinicians make decisions based on the existence of certain risk factor(s) (i.e., a person's risk of depressive disorder increases a certain degree when he or she has restless sleep). However, the progression of depressive disorder (and many other diseases) is based on the latent high-order interactions among multiple risk factors. Such interactions are too complicated to be captured by traditional statistical models but are essential for characterizing and predicting disease progression. Therefore, the proposed deep learning models, which have superior predictive power, can serve as a supplementary "gatekeeper" for clinical decision making.

In predicting the onset of depressive disorder (Task 1), we should identify a threshold to differentiate depressed and non-depressed individuals. Ideally, we would like to have a model with good precision (the percentage of identified depressed people who became indeed

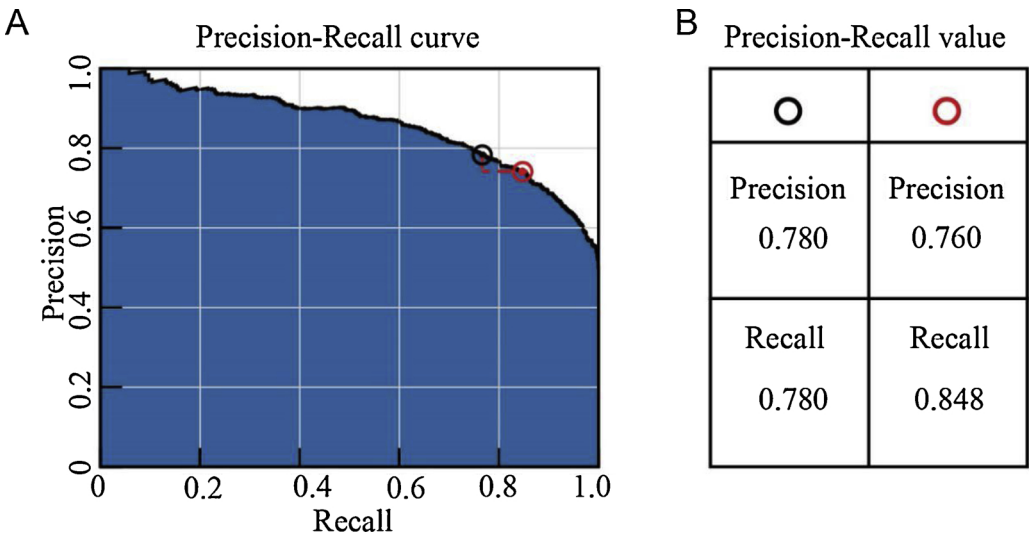

Fig. 3. Precision-Recall curve (A) and the precision and recall values of black dot and red dot respectively (B). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

depressed) as well as recall (the percentage of to-be depressed people who are identified). However, there is always a tradeoff between precision and recall. In clinical practice, the priority is to have a high recall, so that we do not miss many people at risk for depression. To help in clinical decision making, we present the precision-recall (P–R) curve in Fig. 3. In general, the curve is rather flat so that we can increase the recall significantly without sacrificing too much precision. For example, the recall can be increased by 8.7% from the black dot to the red dot. In the meanwhile, the precision only drops by 2.6%. In practice, clinicians and other decision makers (i.e., physicians, nurses, and social workers) can choose the appropriate threshold based on their domain knowledge.

This research has two limitations. First, the HRS data do not include information on hospital utilization and genetic information (e.g., family history), both of which are found to be predictive for mental disorder. Second, the HRS data are based on a cohort that is representative of the whole population in the United States. Whether health and disease progression patterns still hold for people living in a specific state or not remains unclear. In our future research, we aim to address the two limitations by (a) verifying the results in Florida through recruitment of a local cohort and (b) further improving the prediction performance by fusing the survey data and detailed hospitalization information from Skilled Nursing Facilities (SNF) in Southern Florida.[38]

6. Conclusion

This research sheds light on using machine learning models to predict the risk of depressive disorder among the elderly people with a 22-year longitudinal survey data. The proposed multitask LSTM model can successfully capture high-order and temporal patterns that traditional methods ignore. The proposed methods can be helpful for frontline clinicians, social workers, and other disease management groups to detect potential depressive disorders among the elderly people at their early stages, and to accordingly develop health education, preventive interventions, and clinical care for the patients as early as possible. The proposed deep learning methods also have the great potential to be implemented as a decision support system to help clinicians make clinical decisions and inform actionable intervention strategies for the elderly people. Furthermore, the methods are generic and can be easily applied to other risk assessment/prediction tasks for various diseases and clinical outcomes.

Declaration of Competing Interest

This research does NOT involve the experimentation with human subjects. The authors state that informed consent was obtained for analyzing the secondary and opensource survey data. The privacy rights of human subjects are always observed.

Acknowledgements

This research is supported in part by The National Natural Science Foundation of China (NSFC) Grant Nos. 71972164 and 71672163, in part by the Health and Medical Research Fund Grant (HMRF) No. 16171991, in part by the Li Ka Shing Foundation, in part by The Theme-Based Research Scheme of the Research Grants Council of Hong Kong Grant No. T32-102/14N, and in part by the National Science Foundation (NSF) Grant No. 1825761.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijmedinf.2019.103973.

References

[1] The World Health Report, Mental Health: New Understanding, New Hope, World Health Organization, 2001 2001.
[2] World Health Organization, The Global Burden of Disease: 2004 Update, [Online]. Available: (2008) https://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/.
[3] M.J. Prince, et al., The burden of disease in older people and implications for health policy and practice, Lancet 385 (9967) (2015) 549–562.
[4] Y. Suhara, Y. Xu, A.S. Pentland, DeepMood: forecasting depressed mood based on self-reported histories via recurrent neural networks, World Wide Web Conf. (2017), pp. 715–724.
[5] D. Blazer, Depression in the elderly, WebMD Medical Reference, [Online]. Available: (2018) https://www.webmd.com/depression/guide/depression-elderly#1.
[6] H.K. Koh, A.K. Parekh, Toward a United States of health: implications of understanding the US burden of disease, J. Am. Med. Assoc. 319 (14) (2018) 1438–1440.
[7] H.U. Wittchen, R. Lieb, U. Wunderlich, P. Schuster, Comorbidity in primary care: presentation and consequences, J. Clin. Psychiatry (1999).
[8] I.A. Parashos, S. Stamouli, E. Rogakou, R. Theodotou, I. Nikas, A. Mougias, Recognition of depressive symptoms in the elderly: what can help the patient and the doctor, Depress. Anxiety 15 (3) (2002) 111–116.
[9] M.N. Kuchibhatla, G.G. Fillenbaum, C.F. Hybels, D.G. Blazer, Trajectory classes of depressive symptoms in a community sample of older adults, Acta Psychiatr. Scand. 125 (6) (2012) 492–501.
[10] G.C. Martin, D. Nandini, Risk factors for depression among elderly community

subjects: a systematic review and meta-analysis, Am. J. Psychiatry 160 (6) (2003) 1147–1156.
[11] D. Vink, M.J. Aartsen, R.A. Schoevers, Risk factors for anxiety and depression in the elderly: a review, J. Affect. Disord. 106 (1–2) (2008) 29–44.
[12] J. Ryan, A. Farré, K. Ritchie, M.L. Ancelin, C. Proust-Lima, I. Carrière, Chronic and remitting trajectories of depressive symptoms in the elderly. Characterisation and risk factors, Epidemiol. Psychiatr. Sci. 26 (02) (2016) 146–156.
[13] E. Choi, A. Schuetz, W.F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure onset, J. Am. Med. Inform. Assoc. 24 (2) (2017) 361–370.
[14] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[15] Y. Luo, Recurrent neural networks for classifying relations in clinical notes, J. Biomed. Inform. 72 (2017) 85–95.
[16] J. Xie, X. Liu, D.D. Zeng, Mining e-cigarette adverse events in social media using Bi-LSTM recurrent neural network with word embedding representation, J. Am. Med. Inform. Assoc. 25 (1) (2018) 72–80.
[17] H.A.H. Hosny, S.C.M. Srinivasan, J. Keenan, H. Fekry, Midterm results with Birmingham hip resurfacing/synergy stem modular metal-on-metal total hip arthroplasty, Acta Orthop. Belg. 79 (4) (2013) 386–391.
[18] D. Ben-Zeev, E.A. Scherer, R. Wang, H. Xie, Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health, Psychiatr. Rehabil. J. 38 (3) (2015) 218–226.
[19] A.A. Farhan, et al., Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data, 2016 IEEE Wireless Health, WH 2016, (2016), pp. 30–37.
[20] J. Rodda, Z. Walker, J. Carter, Depression in older adults, Annu. Rev. Clin. Psychol. 5 (2009) 363–398.
[21] S. Hirve, et al., Does self-rated health predict death in adults aged 50 years and above in India? Evidence from a rural population under health and demographic surveillance, Int. J. Epidemiol. 41 (6) (2012) 1719–1727.
[22] Y.-F. Tsai, S.-H. Yeh, H.-H. Tsai, Prevalence and risk factors for depressive symptoms among community-dwelling elders in Taiwan, Int. J. Geriatr. Psychiatry 20 (11) (2005) 1097–1102.
[23] G. Ambresin, P. Chondros, C. Dowrick, H. Herrman, J.M. Gunn, Self-rated health and long-term prognosis of depression, Ann. Fam. Med. 12 (1) (2014) 57–66.
[24] M. King, et al., Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees, Arch. Gen. Psychiatry 65 (12) (2008) 1368.
[25] M.T. Brown, D.A. Wolf, Estimating the prevalence of serious mental illness and sementia diagnoses among medicare beneficiaries in the health and retirement study, Res. Aging (Mdd) (2017) 1–19.
[26] J.A. Ailshire, E.M. Crimmins, Psychosocial factors associated with longevity in the United States: age differences between the old and oldest-old in the health and retirement study, J. Aging Res. 2011 (2011) 1–10.
[27] C.K. Suemoto, et al., Development and validation of a 10-year mortality prediction model: meta-analysis of individual participant data from five cohorts of older adults in developed and developing countries, J. Gerontol. Ser. A Biol. Sci. Med. Sci. 72 (3) (2016) 410–416.
[28] M. Cole, N. Dendukuri, Risk factors for depression among elderly community subjects: a systematic review, Am. J. Psychiatry 160 (6) (2003) 1147–1156.
[29] C.M. Perissinotto, I. Stijacic Cenzer, K.E. Covinsky, Loneliness in older persons: a predictor of functional decline and death, Arch. Intern. Med. 172 (14) (2012) 1078–1083.
[30] L. Ayalon, Perceived age discrimination: a precipitator or a consequence of depressive symptoms? J. Gerontol. B Psychol. Sci. Soc. Sci. 73 (5) (2018) 860–869.
[31] K.M. Utzschneider, J. Tong, B. Montgomery, R.N. Ms, J. Udayasankar, Diabetes, depression, and death: a randomized controlled trial of a depression treatment program for older adults based in primary care, Diabetes 44 (9) (2007) 1–15.
[32] M. Schoenbaum, J. Unützer, D. McCaffrey, N. Duan, C. Sherbourne, K. Wells, The effects of primary care depression treatment on patients' clinical status and employment, Health Serv. Res. 37 (5) (2002) 1145–1158.
[33] R.N. Jones, S.J. Fonda, Use of an IRT-based latent variable model to link different forms of the CES-D from the Health and Retirement Study, Soc. Psychiatry Psychiatr. Epidemiol. 39 (10) (2004) 828–835.
[34] C.L. Turvey, R.B. Wallace, R. Herzog, A revised CES-D measure of depressive symptoms and a DSM-based measure of major depressive episodes in the elderly, Int. Psychogeriatrics 11 (2) (1999) 139–148.
[35] A. Esteva, et al., A guide to deep learning in healthcare, Nat. Med. 25 (1) (2019) 24–29.
[36] S. Hochreiter, J. Schmidhuber, Long short term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[37] Z.C. Lipton, J. Berkowitz, C. Elkan, A Critical Review of Recurrent Neural Networks for Sequence Learning, (2015), pp. 1–38.
[38] R. Caruana, Multitask learning, Mach. Learn. 75 (3) (1997) 1–9.
[39] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B 58 (1) (1996) 267–288.
[40] K.P. Singh, N. Basant, S. Gupta, Support vector machines in water quality management, Anal. Chim. Acta 703 (2) (2011) 152–162.
[41] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, (1994).
[42] Mental Health of Older Adults, 2017. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/mental-health-of-older-adults.
[43] C.G. Cahoon, Depression in older adults, Am. J. Nurs. 112 (11) (2012) 22–30.
[44] L.F. Berkman, et al., Depressive symptoms in relation to physical health and functioning in the elderly, Am. J. Epidemiol. 124 (3) (1986) 372–388.
[45] B.E. Bejnordi, et al., Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer, J. Am. Med. Assoc. 318 (22) (2017) 2199–2210.
[46] V. Gulshan, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, J. Am. Med. Assoc. 316 (22) (2016) 2402–2410.
[47] D. Annane, V. Sébille, G. Troché, J.C. Raphaël, P. Gajdos, E. Bellissant, A 3-level prognostic classification in septic shock based on cortisol levels and cortisol response to corticotropin, J. Am. Med. Assoc. 283 (8) (2000) 1038–1045.
[48] M. Startup, M.C. Jackson, S. Bendix, The concurrent validity of the global assessment of functioning (GAF), Br. J. Clin. Psychol. 41 (4) (2002) 417–422.
[49] M. Moscucci, et al., Predictors of major bleeding in acute coronary syndromes: the global registry of acute coronary events (GRACE), Eur. Heart J. 24 (20) (2003) 1815–1823.
[50] D. Noble, R. Mathur, T. Dent, C. Meads, T. Greenhalgh, Risk models and scores for type 2 diabetes: systematic review, BMJ 343 (7836) (2011) 1243.