



## OPEN ACCESS

## EDITED BY

Roozbeh Sadeghian,  
Harrisburg University of Science and  
Technology, United States

## REVIEWED BY

Wenqi Shi,  
Georgia Institute of Technology, United States  
Shinya Tasaki,  
Rush University Medical Center, United States

## \*CORRESPONDENCE

Sylvester O. Orimaye  
✉ sylvester.orimaye@uhsp.edu

RECEIVED 29 December 2023

ACCEPTED 21 March 2024

PUBLISHED 05 April 2024

## CITATION

Orimaye SO and Schmidtke KA (2024)  
Combining artificial neural networks and a  
marginal structural model to predict the  
progression from depression to Alzheimer's  
disease. *Front. Dement.* 3:1362230.  
doi: 10.3389/frdem.2024.1362230

## COPYRIGHT

© 2024 Orimaye and Schmidtke. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Combining artificial neural networks and a marginal structural model to predict the progression from depression to Alzheimer's disease

Sylvester O. Orimaye<sup>1\*</sup> and Kelly A. Schmidtke<sup>2</sup>

<sup>1</sup>College of Global Population Health, University of Health Sciences and Pharmacy, St. Louis, MO, United States, <sup>2</sup>College of Arts and Sciences, University of Health Sciences and Pharmacy, St. Louis, MO, United States

**Introduction:** Decades of research in population health have established depression as a likely precursor to Alzheimer's disease. A combination of causal estimates and machine learning methods in artificial intelligence could identify internal and external mediating mechanisms that contribute to the likelihood of progression from depression to Alzheimer's disease.

**Methods:** We developed an integrated predictive model, combining the marginal structural model and an artificial intelligence predictive model, distinguishing between patients likely to progress from depressive states to Alzheimer's disease better than each model alone.

**Results:** The integrated predictive model achieved substantial clinical relevance when using the area under the curve measure. It performed better than the traditional statistical method or a single artificial intelligence method alone.

**Discussion:** The integrated predictive model could form a part of a clinical screening tool that identifies patients who are likely to progress from depression to Alzheimer's disease for early behavioral health interventions. Given the high costs of treating Alzheimer's disease, our model could serve as a cost-effective intervention for the early detection of depression before it progresses to Alzheimer's disease.

## KEYWORDS

depression, major depressive disorder, Alzheimer's disease, MDD to AD progression, marginal structural model, artificial neural network, predictive model

## 1 Introduction

Decades of research in population health have established depression as a likely precursor to Alzheimer's disease (AD) along with internal and external mediating mechanisms (Chen et al., 1999; Jorm, 2001; Kessing and Andersen, 2004; Ownby et al., 2006; Dotson and Beydoun, 2010; Goveas et al., 2011; Köhler et al., 2016; Almeida et al., 2017; Steffens, 2017). Some of these mechanisms are modifiable if intervened early, e.g., deficient serotonin bindings or inaccessible services (Köhler et al., 2016). Further literature shows that depression is highly prevalent among patients with mild cognitive impairment,

which may then progress to AD (Figure 1; Barnes et al., 2006; Geda et al., 2006; Dotson and Beydoun, 2010; Goveas et al., 2011; Snowden et al., 2015; Ismail et al., 2017; Ding et al., 2023). Research suggests that depression among patients with mild cognitive disorders substantially accelerates cognitive decline many years before the onset of dementia (Snowden et al., 2015). These findings support the direction of causation (depression then AD) and suggest a window for treatment (Almeida et al., 2017; Steffens, 2017; Dafsari, 2020). Further research could identify those patients most likely to develop AD so that timely treatment could be offered (Thapar et al., 2022). However, investigating each factor individually is resource-intensive. Further understanding how each factor interacts with all the other factors is not feasible without computational assistance. Artificial intelligence models could assist (Orimaye et al., 2020a).

The present study builds on existing literature seeking to predict progression from Major Depressive Disorder (MDD) to AD by integrating artificial intelligence models with more traditional statistical methods, specifically a marginal structural model (MSM) and a Multi-Layer Perceptron (MLP) model (Ding et al., 2023). MSMs are commonly used in healthcare to assess associative relationships and causal relationships between variables in observational data. The use of artificial intelligence models in behavioral health is relatively new (Orimaye et al., 2017; Orimaye and Wong, 2018). While artificial intelligence models can capture complex underlying patterns in healthcare data (Orimaye and Wong, 2018), how those data should be interpreted is not always clear—i.e., the black box concern. In addition, while artificial intelligence training data are adequate for determining associations, they may be limited for identifying causal relationships between variables without a more rigorous methodology. In healthcare, such data are often incomplete, raising concerns regarding whether their findings generalize to the broader population. For instance, commonly used Centers for Medicare and Medicaid Services (CMS) Limited Data Sets (LDS) contain incomplete patient data as, e.g., patients move in and out of enrolment or data are censored. MSMs often apply inverse probability weighting to address such confounding biases. Thus, combining MSMs with an artificial intelligence model may increase prediction power while reducing common concerns about interpretability and generalization (Athey, 2015; Wang and Chen, 2019; Prosperi et al., 2020; Balzer, 2021).

The present study aims to develop an integrated predictive model (IPM), combining MSM and MLP models, that distinguishes between patients likely to progress from depressive states to Alzheimer's disease better than each model alone. In addition to including patient factors (e.g., gender, comorbidities, etc.), we include exogenous factors (e.g., healthcare accessibility) in this model that research already suggests influence progression. Our predictive model has the potential to be at the core of an artificial intelligence-based behavioral analysis and intervention clinical screening tool that identifies patients who are likely to progress from depression to AD for early behavioral health intervention. Using such an integrative approach has previously proved effective for diabetes care provisioning (Kalia et al., 2022). To our knowledge, this is the first time such an integrative approach has been used to predict the progression from depression to AD.

## 2 Materials and methods

### 2.1 Study population/datasets

The primary data sources for our study are the de-identified Medicare inpatient and skilled nursing facility files of the Centers for Medicare and Medicaid Services (CMS) Limited Use Datasets (LDS). The dataset contains 6,567,071 Medicare beneficiary claims from 2012 to 2019, which comprise 5% of the total Medicare beneficiaries. CMS is a federal health care program in the United States that is mainly composed of residents who are 65 and older. From this data set, we identified 282 MDD patients who progressed to AD and a randomly selected equal number of patients with MDD who never progressed to AD between 2012 and 2019.

We used the International Classification of Diseases, Clinical Modification (ICD-CM) coding systems, including ICD-9-CM and ICD-10-CM, to identify depression and AD diagnoses. The ICD-9-CM diagnosis applies to data from 2012 to 2015, and the ICD-10-CM diagnosis applies to data from 2016 to 2019.

Together with the CMS LDS, we used the county-level Health Professional Shortage Areas (HPSA) classification to identify the shortage of health professionals in each patient's county and the 2013 county-level Economic Research Service Rural-Urban Continuum Codes (RUCCs) to classify whether an observation is from a designated rural or urban county (Hardeman et al., 2022; Tan et al., 2022). The Office of Management and Budget uses the RUCC to classify large metros with at least 1 million residents and small metro areas with <1 million residents as urban. Other non-metropolitan regions are classified as rural.

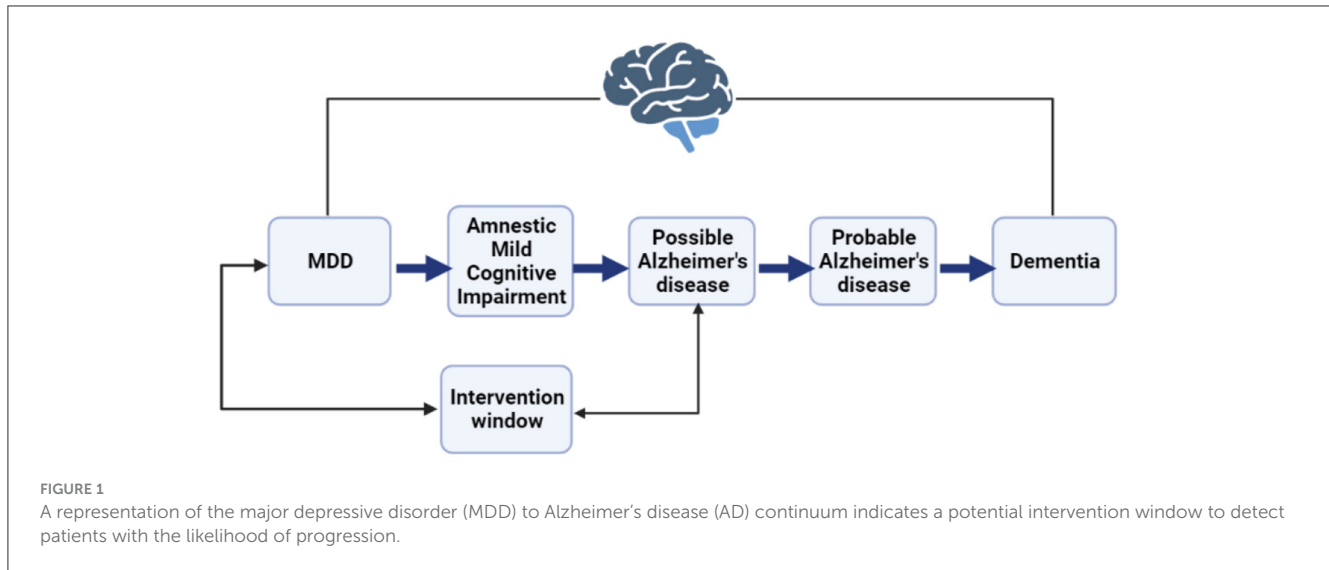
### 2.2 Study variables

#### 2.2.1 Outcome variable

We measured the outcome variable as a binary representation of patients who, between 2012 and 2019, made at least one claim for depression and later made at least one claim for AD. We excluded observations where claiming an AD diagnosis precedes a depression diagnosis. Thus, each patient's MDD diagnosis must precede the AD diagnosis within the longitudinal 8-year period. Further, the underlying unique pattern for predicting a future outcome (progressed to AD or not) is captured in the variability of the predicting variables, many of which increase or decrease over a longitudinal period.

#### 2.2.2 Predictor variables

Predictor variables include demographic variables: *age*, *gender*, *race*, and *rural-urban* location. The patient's location was included because there is growing evidence of increasing disparities in the prevalence of AD between the urban and rural parts of the United States (Bradford et al., 2009; Mehta, 2017; Zissimopoulos et al., 2018; Jack et al., 2019; Orimaye et al., 2020b). Additional predictors include the HPSA classification that captures counties with shortages of health professionals for access to early and preventive behavioral health services,



including mental and behavioral health practitioners, dentists, and primary care providers. The HPSA classification includes a whole shortage area (the entire county is underserved with health professionals), a partly shortage area (the county has some health professionals), and no shortage (the county has substantial health professionals). Research has shown that accessibility to early detection through professional healthcare services provides effective treatment strategies for preventing AD (Eichler et al., 2015; Orimaye et al., 2017, 2020a; Dafsari, 2020).

Other variables include Medicare *utilization days count after MDD diagnosis* through inpatient hospital admission or residence at a skilled nursing facility, any *comorbidity diagnoses* secondary to or co-occurring with the primary major depressive disorder (MDD) diagnosis, the *count of claims for MDD diagnosis* indicating the probable number of patient-provider visits for pharmacological or non-pharmacological intervention, and a *follow-up visit* variable that indicates whether patients made subsequent claims or visits after their first MDD diagnosis.

### 2.3 Assessment and diagnosis of depression

We operationalized MDD as clinical depression that includes all three severity levels of depression (mild, moderate, and severe) for single (F.32) and recurrent (F.33), including recurrent brief depressive episodes (Hasin et al., 2018). We consider 296.2X for ICD-9-CM and the F.32.X for ICD-10-CM (Association, 2015). According to the DSM-5, MDD is characterized by a minimum of five depressive mood symptoms presenting as feelings of sadness, emptiness, hopelessness, tremendous weight loss or daily bidirectional changes in appetite, insomnia or hypersomnia, agitation, fatigue, feelings of worthlessness, reduced cognitive or executive functions, and thoughts of suicide.

### 2.4 Assessment and diagnosis of Alzheimer's disease

Using the ICD codes in the CMS LDS, we identified AD with ICD-9-CM code 331.0 and ICD-10-CM codes G30.0, G30.1, G30.8, and G30.9 (Sachdev et al., 2014). Based on the DSM-5, AD is a progressive neurodegenerative disease that causes the death of nerve cells in various parts of the brain, mainly in the frontotemporal region. This loss of function results in a decline of cognitive abilities such as memory and language.

### 2.5 Multi-layer perceptron

We learned the underlying representation of the different levels of abstraction of the covariates and predictors using an efficient backward propagation (backpropagation) method for the Multi-Layer Perceptron (MLP) deep learning algorithm (Zhou et al., 2021). The backpropagation method uses the chain rule, enabling the MLP to systematically understand the underlying structure in reverse order, beginning from the output layer of the network (outcome variable) to the input layer (covariates and predictors). Thus, the MLP considers the complex underlying representation of the observational data and the contextual interaction among the covariates and predictors, often challenging to identify by standard regression models and incredibly difficult to calculate manually. The underlying abstraction of the data is captured as the weights connecting several internal layers of the predictor and covariate structures. A unique underlying weight structure adequately predicts the progression from depression to AD. Similarly, the control group, non-progressive depression (no progression from depression to AD), has an underlying weight structure, which is inherently different from the progression group.

We performed a grid search to identify optimal values for the MLP. Parameter values include the rectified linear

unit (ReLU) activation function, the limited-memory Broyden–Fletcher–Goldfarb–Shanno (lbfgs) solver, an alpha of 1e-8, tolerance level of 1e-3, learning rate initialization of 0.1, hidden layers of 6, max iteration of 1,200, and a random state of 4. ReLU is known for its exemplary performance in computing the predicted probabilities for the underlying weight structure (Hara and Saito, 2015). Also, the lbfgs is known for its efficiency on small to medium datasets, as observed in our training set. Finally, we trained the MLP model on 80% of the dataset and tested it with the remaining 20%.

The proposed fully connected structural representation of the observational data using the MLP architecture is presented in this section. With the predictors and covariates fed as inputs to the network structure, the hidden layer helps compute the weights using a specified activation function. The weights capture the underlying pattern within the data. The MLP then classifies the output as one of the outcome categories. Figure 2 depicts a simple three-layer structure of an MLP model.

## 2.6 Marginal structural model

MSM is a class of causal models on observational data, which accounts for time-varying confounders when examining the effect of a time-dependent exposure. Its use has been encouraged to eliminate bias in observational data (Athey, 2015; Prospero et al., 2020).

Thus, we present the causal model as a directed acyclic graph (DAG). The DAG indicates the time-varying covariates, confounders, and the underlying connection with MDD diagnosis and the risk of progression to AD. In the DAG,  $X$  indicates the factors affecting the progression from MDD to AD,  $E$  denotes the exposure for having been diagnosed with MDD,  $U$  is the unknown factor(s) affecting the outcome, and  $Y$  is the outcome, which is the progression from MDD to AD. The DAG shows that the probability of being diagnosed with MDD at  $E(0)$  is conditioned on covariates and confounding factors  $X(0)$  in the data.  $E(1)$  and  $X(1)$  are conditioned on  $E(0)$  and  $X(0)$ . The inverse of these conditional probabilities of receiving exposure  $E(1)$  given  $E(0)$  and  $X(0)$  are referred to as the inverse probability of treatment weights (IPTW; Austin, 2016). *A priori*, the IPTW for the MSM is calculated using the following equation:

$$IPTW = \prod_{t=0}^T \frac{P(E_t|E_{t-1}, X_0)}{P(E_t|E_{t-1}, X_0, X_{t-1})} \quad (1)$$

where  $t$  indicates the time lag in years for  $T$  number of years with  $t = 0$  as the baseline year and  $t - 1$  as the 1-year time lag,  $E_t$  indicates the MDD diagnosis at a particular time or time lag in years.  $X$  is a set of time-varying and non-time-varying confounders. The conditional probabilities for the numerator and denominator were computed using the predicted probabilities from a logistic regression model that controls for the non-time-varying and time-varying confounders. Finally, the IPTW computes a stabilized weight that controls for time-varying confounding variables. It provides better causal estimates on observational data. Figure 3 depicts a simple MSM.

## 2.7 Integrated predictive model (IPM)

Our integrated model combines the MSM and MLP models using an ensemble learning method called the Gradient Boosting Classifier (GBC; Wade, 2020). This method uses an additive model to reflect the contribution of each model. Hyperparameter tuning was performed to control overfitting by introducing randomness to the model using a 60% subsample of the training set for each boosting iteration without replacement. Thus, the parameters for the GBC include `random_state = 28`, `n_estimators = 100`, `learningrate = 0.01`, `max_depth = 3`, and `subsample = 0.6`. Also, each model in the ensemble model used the same training data. The integrated prediction for progression from MDD to AD is computed using Equation 2 as follows:

$$IPM_i = [pANN_i, pMSM_i] \quad (2)$$

where  $IPM_i$  is the integrated prediction model for the  $i$ -th observation,  $pANN_i$  indicates the MLP model prediction for the  $i$ -th observation, and  $pMSM_i$  is the MSM prediction for the  $i$ -th observation. Thus, we used the IPM to classify unseen observations likely to progress from MDD to AD. The IPM weighs the predictions from the  $pANN$  and  $pMSM$  models and effectively integrates them as features for the final predictive outcome. The predictive features from each model represent varying aspects of the underlying patterns of the data. More importantly, the MSM features capture marginal relationships between the variables, while the MLP features capture more complex, non-linear patterns from the data. Therefore, the IPM leverages the diverse feature representations to make the final predictions.

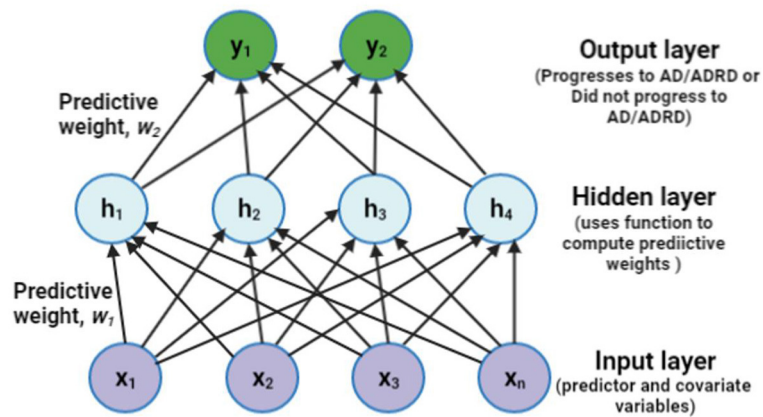
Further, we evaluated the model using the Area Under the Curve (AUC), accuracy, precision, recall, and the f1-score (Jiao, 2016; Carrington et al., 2022).

## 2.8 Sensitivity analysis

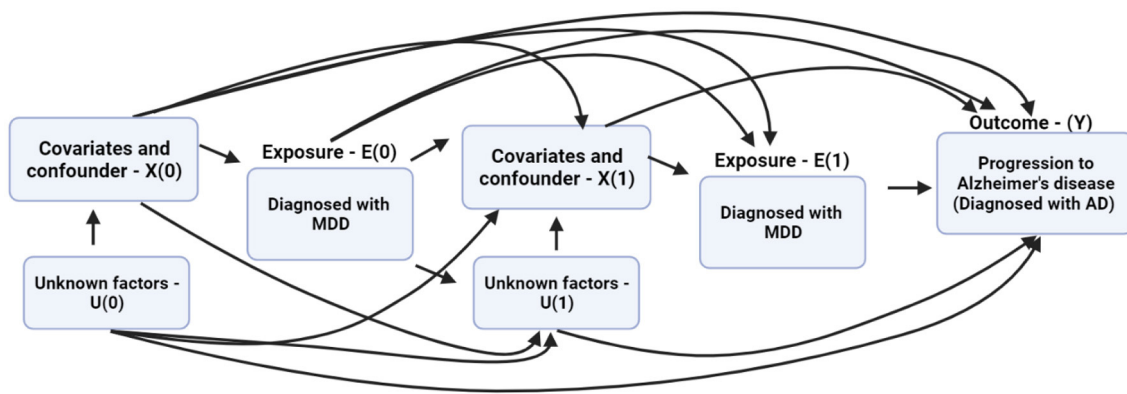
We performed sensitivity analysis (SA) as part of the safeguards for prediction modeling and explainable artificial intelligence models from observational data (Smith, 2021). Specifically, we demonstrated the robustness and precision of the IPM using external validation techniques (Peters and Bühlmann, 2016). Thus, we deliberately distorted the original dataset by randomly replacing 40% of the data with missing values across all the predictors. Further, we used the Markov Chain Monte Carlo multiple imputation technique to impute the missing values using the original data's predictive distribution (multivariate normal distribution; Rezvan and Lee, 2015). We then evaluated the IPM on the imputed dataset and compared findings to the original dataset.

## 2.9 Statistical analysis and machine learning platforms

To show the similarity between the characteristics of patients who remained MDD and progressed to AD, we conducted the chi-square test of independence on all the categorical variables



**FIGURE 2**  
A sample representation of the proposed Multi-Layer Perceptron (MLP) with one hidden layer of four units. ADRD means Alzheimer’s disease (AD) and related dementia, often used interchangeably with AD alone.



**FIGURE 3**  
The directed acyclic graph (DAG) describes the relationship between the exposure, covariates, time-varying predictors, and progression from major depressive disorder (MDD) to Alzheimer’s disease (AD).

(age, gender, race, rural-urban, HPSA, comorbidity, and follow-up visit). Further, we performed a two-sided independent sample *t*-test on all numerical variables (follow-up visit and count of claims for MDD). We conducted *t*-tests, chi-square tests, and logistic regression models for computing the IPTW stabilized weights (Equation 1) using the Statistical Analysis Software (SAS) version 9.4 on SAS Studio. All statistical tests were performed at a 95% confidence interval (CI). We also used Python 3.12.1 to develop the MSM with the Weighted Least Squares (WLS) estimation function of the statsmodels module version 0.15.0. The WLS function takes the IPTW as weights to fit the MSM and generates the predictive probabilities. Finally, we used the Scikit-learn version 1.1 Python machine learning module to implement the MLP model and the Gradient Boosting Classifier that performed the IPM. Note that 80% of the patients per group were used as the training set. There were 282 beneficiaries per group (remained MDD vs. progressed to AD), which equates to ~226 patients per group for training and ~56 patients for

testing. The exact proportions were used across the IPM, MSM, and MLP.

### 3 Results

#### 3.1 Sample characteristics

Table 1 shows the descriptive characteristics of the samples used in developing our model. A total of 282 unique patients met the definition of the outcome variable by progressing from MDD to AD diagnosis. A randomly selected group of 282 patients who remained MDD as of 2019 was used as a control group. Patient demographics were broadly similar across groups per the statistical significance tests at 95% CI. Descriptively, there were more older patients who progressed to AD compared to those who remained MDDs. Both groups contained a higher percentage of female than male patients (remained MDD 59.22% and progressed to AD 65.96%). White was the most common ethnicity in both groups (remained

TABLE 1 Descriptive characteristics of the sample.

| Predictors  | Remained MDD (N = 282) |       | Progressed to AD (N = 282) |       |
|---|------------------------|-------|----------------------------|-------|
|   | N                      | %     | N                          | %     |
| <b>Age*</b>                                       |                        |       |                            |       |
| <65   | 117                    | 41.49 | 32                         | 11.35 |
| 65–69   | 17                     | 6.03  | 30                         | 10.64 |
| 70–74   | 6                      | 2.13  | 44                         | 15.6  |
| 75–79   | 100                    | 35.46 | 66                         | 23.40 |
| 80–84   | 28                     | 9.93  | 62                         | 21.99 |
| >84   | 14                     | 4.96  | 48                         | 17.02 |
| <b>Gender<sup>a</sup></b>                         |                        |       |                            |       |
| Male  | 115                    | 40.78 | 96                         | 34.04 |
| Female  | 167                    | 59.22 | 186                        | 65.96 |
| <b>Race<sup>a</sup></b>                           |                        |       |                            |       |
| White   | 241                    | 85.46 | 245                        | 86.88 |
| Black   | 29                     | 10.28 | 27                         | 9.57  |
| Others  | 12                     | 4.26  | 10                         | 3.55  |
| <b>Rural-urban*</b>                               |                        |       |                            |       |
| Rural   | 233                    | 82.62 | 204                        | 72.34 |
| Urban   | 49                     | 17.38 | 78                         | 27.66 |
| <b>HPSA<sup>a</sup></b>                           |                        |       |                            |       |
| None  | 30                     | 10.64 | 32                         | 11.35 |
| Whole   | 94                     | 33.33 | 73                         | 25.89 |
| Partly  | 158                    | 56.03 | 177                        | 62.77 |
| <b>Comorbidity<sup>a</sup></b>                    |                        |       |                            |       |
| None  | 78                     | 27.66 | 64                         | 22.70 |
| One or more                                       | 204                    | 72.34 | 218                        | 77.30 |
| <b>Follow-up visit<sup>a</sup></b>                |                        |       |                            |       |
| At least two visits                               | 69                     | 24.47 | 66                         | 23.40 |
| No follow-up visit                                | 213                    | 75.53 | 216                        | 76.60 |
|   | <b>Mean ± St. dev</b>  |       |                            |       |
| <b>Utilization days count after MDD diagnosis</b> | 9.65 ± 9.55            |       | 10.90 ± 8.39               |       |
| <b>Count of claims for MDD<sup>a</sup></b>        | 1.78 ± 3.03            |       | 1.58 ± 2.63                |       |

\*Statistical significance difference observed between “Remained MDD” and “Progressed to AD” at  $p < 0.05$ .

<sup>a</sup>No statistical significance observed.

MDD 85.46% and progressed to AD 86.88%) groups. Most patients lived in rural locations (remained MDD 82.62% and progressed to AD 72.34%), experiencing partial healthcare shortages (remained MDD 56.03% and progressed to AD 62.77%). Both groups utilized healthcare services with similar frequencies (remained MDD 9.65 days and progressed to AD 10.90 days).

### 3.2 Model performance and clinical relevance

Table 2 shows the performance of the MSM, MLP, and IPM on 20% held-out test data. In terms of predicting MDD

patients who progressed to AD, the MSM outperformed the MLP across all the statistical measures except precision. However, the IPM showed improved model performance across all the statistical measures compared to the MSM or MLP. We observed that the IPM showed a higher precision in detecting patients who remained MDD than those who progressed to AD. Figure 4 shows the graphical representation of the area under the curve (AUC) measure of clinical relevance, depicting the differences between the predictive effectiveness of the MSM, MLP, and IPM. Only the IPM achieved substantial clinical relevance when considering the AUC measure above 80% (Verbakel et al., 2020; Carrington et al., 2022).

TABLE 2 Model performance for predicting progression from depression to AD on a smaller dataset.

| Statistical measures | MSM                    |                  | MLP                    |                  | IPM (MSM + MLP)               |                  |
|----------------------|------------------------|------------------|------------------------|------------------|-------------------------------|------------------|
|                      | Remained MDD           | Progressed to AD | Remained MDD           | Progressed to AD | Remained MDD                  | Progressed to AD |
| Precision            | 0.66                   | 0.67             | 0.58                   | 0.66             | 0.84                          | <b>0.71</b>      |
| Recall               | 0.61                   | <b>0.71</b>      | 0.70                   | <b>0.53</b>      | 0.59                          | <b>0.90</b>      |
| F1-score             | 0.63                   | <b>0.69</b>      | 0.63                   | <b>0.58</b>      | 0.70                          | <b>0.79</b>      |
| Accuracy             | 0.66                   |                  | 0.62                   |                  | <b>0.75</b>                   |                  |
| AUC                  | 0.70 (CI: 0.61 – 0.80) |                  | 0.68 (CI: 0.51 – 0.72) |                  | <b>0.91</b> (CI: 0.85 – 0.97) |                  |

Bold values highlight differences in the performance of the three models.

### 3.3 Influencing variables in the IPM

We evaluated the trained models to identify influencing features and their significant contributions to the IPM. Table 3 shows the ranking of features by importance to the ensemble model. We ranked the MSM features using the marginal structural model’s weighted least square regression coefficients. Additionally, we ranked the MLP features by computing the feature importance based on the sum of absolute weights for each input feature connected to the first hidden layer of the neural network. Significant coefficients and weights were identified at  $p < 0.05$ . In determining significant variables, the results of the MSM-fitted model generate the coefficients and the corresponding  $p$ -values. However, for computing  $p$ -values for the MLP, we used an independent  $t$ -test to reject the null hypothesis that the mean of the feature weights is zero.

Four variables significantly influence the MSM model, with three of the four ranked as the top three features. They include, in order, any *comorbidity* diagnoses secondary to or co-occurring with MDD, residing in a *rural location*, *age*, and Medicare *utilization days count after MDD diagnosis*. On the other hand, *race* is the only significant feature in the MLP model, but it contains two of the four MSM top-ranked features in its top four ranked features (*comorbidity* and *rural location*). Thus, we show that the MSM features compensate for the bias in the IPM as they capture the underlying patterns in the data.

### 3.4 Error analysis

We performed an error analysis for the three models using the respective confusion matrices generated on the test data alone. Again, each model was trained on a separate training set independent of the test set. We used each model’s confusion matrix to compute the models’ overall precision, recall (sensitivity), and specificity. The IPM shows 89.8% precision (53 true positives and 6 false positives), showing a minimal rate of false positives compared to MSM’s 71.2% (42 true positives and 17 false positives) and MLP’s 52.5% (31 true positives and 28 false positives). However, the IPM has a slightly low recall, 75.7%, but 86.0% specificity compared to its precision value. MSM and MLP showed the same recall values at 66.7 and 66.0%, but the MSM has a moderately higher specificity (66%) than the MLP (57.6%). Thus, the IPM shows evidence of a low rate of false positives (wrong predictions) among patients who remained MDD while correctly identifying patients who are likely to progress to AD the majority of the time.

### 3.5 Results of the sensitivity analyses

Tables 4A, B show the results of the Sensitivity Analysis (SA) performed on the imputed data of the original sample using the IPM and an MLP-only model. The IPM maintained its effectiveness and clinical relevance in predicting patients progressing to AD. The AUC clinical relevance measure remained above 80% (Figure 5A). Also, the IPM showed a high precision value of 82% despite the distortion in the data.

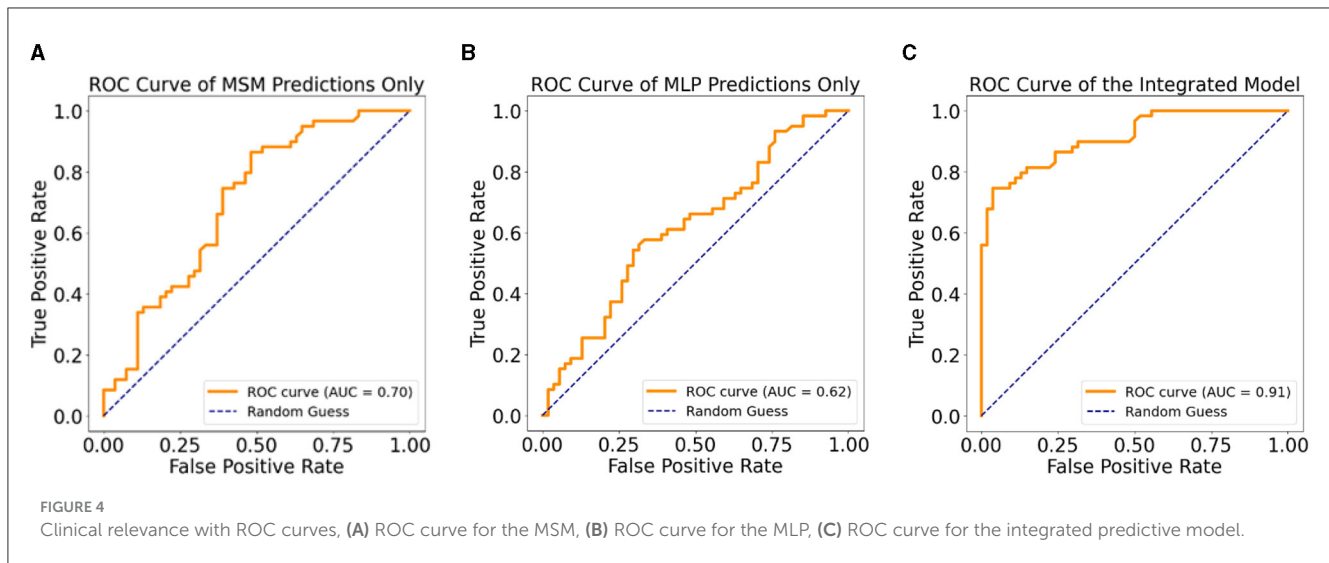


TABLE 3 Ranked features by importance to the ensemble model.

| Rank # | MSM features       | MSM coefficient | MLP features       | MLP abs. weight |
|--------|--------------------|-----------------|--------------------|-----------------|
| 1      | Comorbidity*       | 0.145           | Comorbidity        | 4.600           |
| 2      | Rural-urban*       | 0.142           | Follow-up visit    | 4.303           |
| 3      | Age*               | 0.089           | HPSA               | 4.234           |
| 4      | HPSA               | 0.036           | Rural-urban        | 3.731           |
| 5      | Gender             | 0.015           | Gender             | 3.453           |
| 6      | Utilization days*  | 0.005           | # of MDD diagnosis | 3.417           |
| 7      | Follow-up visit    | 0.001           | Race*              | 2.874           |
| 8      | # of MDD diagnosis | -0.005          | Age                | 2.854           |
| 9      | Race               | -0.030          | Utilization days   | 1.279           |

\* $p < 0.05$ .

Furthermore, Figures 5A, B demonstrate how the MSM helps with bias removal in two ways: first, by modeling with the IPTW for estimating treatment effects. The MSM features comprise IPTW, which computes a stabilized weight that controls for time-varying confounding variables. We emphasize that using IPTW has been proven to eliminate bias compared to using ordinary variables as feature representation (Athey, 2015; Prospero et al., 2020). Second, the sensitivity analysis confirms the effect of the estimated treatment effect on the IPM by removing the IPTW from the model and testing on the imputed dataset. Compared to the sensitivity analysis results with the IPM, the MLP-only model showed a lower AUC (Figure 5B). This shows the benefit of the removed IPTW features as a bias-removing mechanism.

## 4 Discussion

In this study, we demonstrated the effectiveness of a clinically relevant integrated predictive model (IPM) over a marginal structural model (MSM) alone or the Multi-Layer Perceptron (MLP) model alone. Our findings suggest that the IPM can combine the strength of the other models to differentiate patients diagnosed

with major depressive disorder (MDD) who are likely to progress to Alzheimer’s disorder (AD) from those who do not. Below, we briefly review the relevance of these findings in light of previous research and highlight the potential benefits of future clinical practice.

The effectiveness of using artificial neural network models, like our MLP algorithm, in predicting health outcomes is an attractive idea (Orimaye and Wong, 2018; Zhou et al., 2021). However, the strength of an MLP predictive model is better realized on large datasets. In this study, the temporal gap between MDD and AD diagnosis limits the number of patients that meet the definition of the outcome variable ( $N = 282$ ). Nonetheless, the statistical measure of precision demonstrated by the MLP model (Table 2) is on par with the MSM (67 vs. 66%). Further, the AUC measure of clinical relevance for the MLP model is only two points lower than the traditional MSM model (68 vs. 70%).

Clinical support for combining artificial intelligence models with other statistical models requires careful analysis and validation across different statistical measures. Previous research focused on diabetes (Kalia et al., 2022). Our study focused on predicting AD. Future studies may develop models for other diseases. Recent studies have attested to the effectiveness of using an ensemble of models to create an integrated predictive model (MacKay et al.,



2021). In particular, the gradient-boosting classification algorithm used in combining the MSM and MLP predictive probabilities has a track record of reliability in healthcare data (Zhang et al., 2019; Rufo et al., 2021). The literature attributes the effectiveness of gradient-boosting classifiers to the careful combination of several weak decision tree learning methods to create a more robust decision tree with better predictive accuracy (Wade, 2020).

The precision of the IPM in predicting progression from depression to AD supersedes that of MSM and MLP by at least 5 points. Effective classifiers tend to have similar precision and recall values. We believe that the limited sample size for the outcome can explain the gap between the precision and recall measures for the IPM (71 vs. 90%). Nevertheless, the F1-score, accuracy, and AUC statistical measures for the IPM demonstrated robust predictive performance compared to the MSM or MLP alone.

TABLE 4 (A, B) Model performance from the sensitivity analysis using imputed data (N = 564).

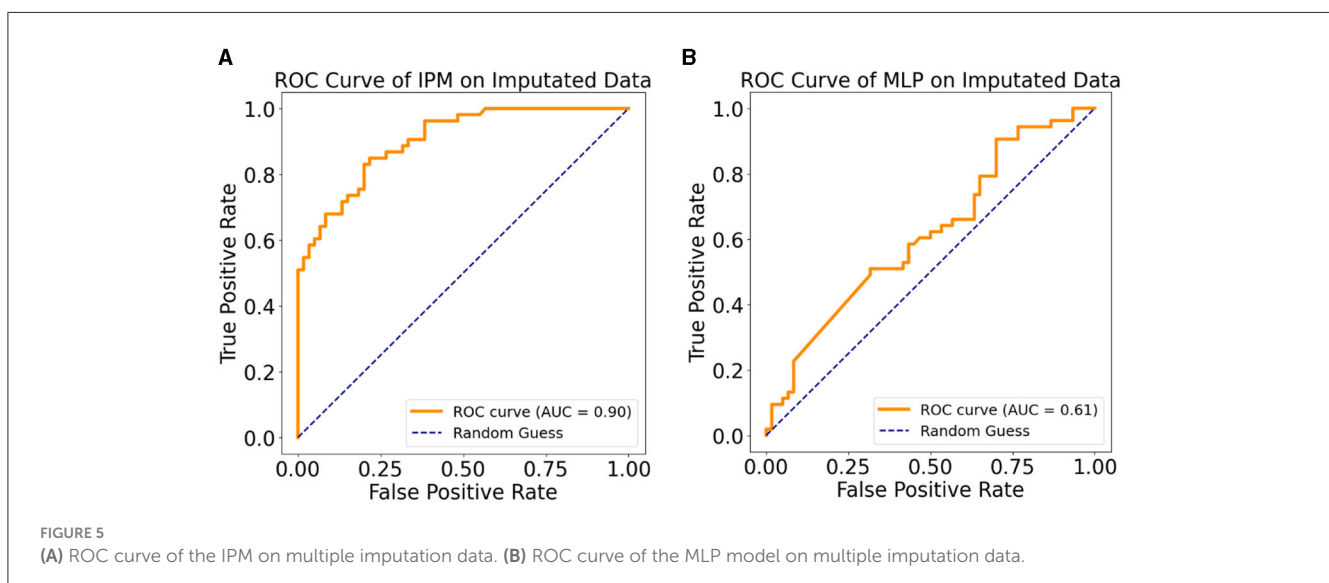
| Statistical measures                        | Remained MDD                  | Progressed to AD |
|---|-------------------------------|------------------|
| <b>(A) Sensitivity analysis—MLP and MSM</b> |                               |                  |
| Precision                                   | 0.76                          | <b>0.82</b>      |
| Recall                                      | 0.87                          | <b>0.68</b>      |
| F1-score                                    | 0.81                          | <b>0.74</b>      |
| Accuracy                                    | <b>0.78</b>                   |                  |
| AUC   | <b>0.90</b> (CI: 0.84 – 0.96) |                  |
| <b>(B) Sensitivity analysis—without MSM</b> |                               |                  |
| Precision                                   | 0.57                          | 0.49             |
| Recall                                      | 0.40                          | 0.66             |
| F1-score                                    | 0.47                          | 0.56             |
| Accuracy                                    | 0.52                          |                  |
| AUC   | 0.61 (CI: 0.51 – 0.72)        |                  |

Bold values highlight differences in the performance of the three models.

More importantly, findings from the sensitivity analysis further attest to the robustness of the IPM model. Despite the distortion introduced to the data, the IPM showed a substantial AUC close to the AUC measured on the original dataset (Table 4A). With only 60% of the distorted sensitivity analysis data maintaining an actual discriminant pattern, we believe the IPM model correctly emphasizes precision using only the portion of the data exhibiting discriminant features.

The IPM could form a part of a clinical intervention screening tool that identifies patients likely to progress from MDD to AD for early behavioral health interventions. In addition to other freely available datasets used in this study, we anticipate future research sponsorship by the CMS or one of its agencies such that access to the entire CMS Medicare data would be free for large-scale model development. Using the model on electronic medical records in real-time could save health systems and CMS costs by identifying patients potentially in need of early behavioral health intervention before they progress to AD. Given the high costs of treating AD (Lynch, 2018), our model could be cost-effective. However, how this tool is sustained across time and what interventions are offered are matters for future research.

This present study is not without limitations. First, the CMS LDS dataset used in this study has inherent limitations, which are popularly discussed in the literature (Lee et al., 2019; MacKay et al., 2021; Velasquez and Orav, 2023). The CMS LDS dataset does not provide the clinical diagnostic criteria and instruments for any disease outcomes (MDD, AD, and other comorbidities). Second, the lack of continuous enrollment for some eligible Medicare beneficiaries creates censoring problems that are challenging to capture. For instance, the data failed to capture why some patients diagnosed with MDD did not make subsequent claims or visits after their first diagnosis. Whether the patients stopped because an intervention (treatment) worked after their first diagnosis or they stopped because they did not want to pursue therapy is unclear. Third, the control group was never matched using statistical methods such as propensity score matching (Caliendo, 2008). Instead, we randomly selected an equal number of MDD patients.



In addition to the many practical limitations for creating a set of balanced covariates across demographics such as age and race in healthcare research (Reiffel, 2020), a matched set of observations will likely limit the variability of the model, thereby reducing the generalizability. Also, the intention of this study was not to estimate the effect of an intervention as often required in targeted quasi-experimental studies on observational data (Reiffel, 2020). Fourth, unlike the MSM, it is still challenging to describe the effectiveness of the learned weights of the artificial neural networks relative to the combined model. Finally, the CMS LDS dataset used in this study comprises only 5% of the total inpatient and skilled nursing facilities data. The dataset excludes outpatients and carrier files, which could provide additional observations relating to MDD patients progressing to AD. This explains the small sample size of 282 patients who progressed to AD over 8 years. Therefore, future work should endeavor to train the predictive model on a combination of inpatient, outpatient, and skilled nursing facilities and the carrier files data to improve the model's generalizability.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: The CMS Limited Datasets were analyzed in this study. Information about the CMS LDS data can be found at: <https://www.cms.gov/data-research/files-for-order/limited-data-set-lds-files>. The county-level Health Professional Shortage Areas (HPSA) classification can be found at: <https://bhw.hrsa.gov/workforce-shortage-areas/shortage-designation#hpsas>. The 2013 county-level Economic Research Service Rural-Urban Continuum Codes (RUCCs) can be found at: <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/>.

## Ethics statement

The studies involving humans were approved by East Tennessee State University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

SO: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Methodology,

Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. KS: Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the Center for Perioperative Mental Health at Washington University in St. Louis under the R03 Pilot grant in conjunction with the National Institute of Mental Health grant P50MH122351, the College of Global Population Health at the University of Health Sciences and Pharmacy in St. Louis, and the East Tennessee State University Research Development Committee major research grant.

## Acknowledgments

We acknowledge the support of the Center for Perioperative Mental Health at Washington University in St. Louis, the Center for Health Outcomes Research and Education (SCORE) at the University of Health Sciences and Pharmacy in St. Louis, and the Centers for Medicare and Medicaid Services. We also acknowledge the contributions of the following researchers toward an ongoing research grant effort emanating from this study: Kate Beatty, Nathan L. Hale, Pamela K. Xaverius, and George T. Grossberg.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Almeida, O. P., Hankey, G. J., Yeap, B. B., and Golledge, J. (2017). Depression as a modifiable factor to decrease the risk of dementia. *Transl. Psychiatry* 7:e1117. doi: 10.1038/tp.2017.90
- Association, A. P. (2015). *Depressive Disorders: DSM-5<sup>®</sup> Selections*. Washington, DC: American Psychiatric Publishing.
- Athey, S. (2015). "Machine learning and causal inference for policy evaluation," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: Association for Computing Machinery), 5–6.
- Austin, P. C. (2016). Variance estimation when using inverse probability of treatment weighting (IPTW) with

- survival analysis. *Stat. Med.* 35, 5642–5655. doi: 10.1002/sim.7084
- Balzer, L. B. (2021). Invited commentary: machine learning in causal inference—how do i love thee? Let me count the ways. *Am. J. Epidemiol.* 190, 1483–1487. doi: 10.1093/aje/kwab048
- Barnes, D. E., Alexopoulos, G. S., Lopez, O. L., and Williamson, J. D. (2006). Depressive symptoms, vascular disease, and mild cognitive impairment: findings from the Cardiovascular Health Study. *Arch. Gen. Psychiatry* 63, 273–279. doi: 10.1001/archpsyc.63.3.273
- Bradford, A., Kunik, M. E., Schulz, P., and Williams, S. P. (2009). Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. *Alzheimer Dis. Assoc. Disord.* 23:306. doi: 10.1097/WAD.0b013e3181a6bebc
- Caliendo, M. (2008). Some practical guidance for the implementation of propensity score matching. *J. Econ. Surv.* 22, 31–72. doi: 10.1111/j.1467-6419.2007.00527.x
- Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., and Osmani, V. (2022). Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. *IEEE Trans. Pat. Anal. Mach. Intell.* 45, 329–341. doi: 10.1109/TPAMI.2022.3145392
- Chen, P., Ganguli, M., and Mulsant, B. H. (1999). The temporal relationship between depressive symptoms and dementia: a community-based prospective study. *Arch. Gen. Psychiatry* 56, 261–266. doi: 10.1001/archpsyc.56.3.261
- Dafari, F. S. (2020). Depression—an underrecognized target for prevention of dementia in Alzheimer's disease. *Transl. Psychiatry* 10, 1–13. doi: 10.1038/s41398-020-0839-1
- Ding, H., Wang, B., Hamel, A. P., Melkonyan, M., Ang, T. F. A., Au, R., et al. (2023). Prediction of progression from mild cognitive impairment to Alzheimer's disease with longitudinal and multimodal data. *Front. Dement.* 2:1271680. doi: 10.3389/frdem.2023.1271680
- Dotson, V. M., and Beydoun, M. A. (2010). Recurrent depressive symptoms and the incidence of dementia and mild cognitive impairment. *Neurology* 75, 27–34. doi: 10.1212/WNL.0b013e3181e62124
- Eichler, T., Thyrian, J. R., Hertel, J., Michalowsky, B., Wucherer, D., Dreier, A., et al. (2015). Rates of formal diagnosis of dementia in primary care: the effect of screening. *Alzheimer's Dement.* 1, 87–93. doi: 10.1016/j.dadm.2014.11.007
- Geda, Y. E., Knopman, D. S., Mrazek, D. A., Jicha, G. A., Smith, G. E., Negash, S., et al. (2006). Depression, apolipoprotein E genotype, and the incidence of mild cognitive impairment: a prospective cohort study. *Arch. Neurol.* 63, 435–440. doi: 10.1001/archneur.63.3.435
- Goveas, J. S., Espeland, M. A., Woods, N. F., Wassertheil-Smoller, S., and Kotchen, J. M. (2011). Depressive symptoms and incidence of mild cognitive impairment and probable dementia in elderly women: the Women's Health Initiative Memory Study. *J. Am. Geriatr. Soc.* 59, 57–66. doi: 10.1111/j.1532-5415.2010.03233.x
- Hara, K., and Saito, D. (2015). "Analysis of function of rectified linear unit used in deep learning," in 2015 *International Joint Conference on Neural Networks (IJCNN)* (Killarney: IEEE), 1–8.
- Hardeman, R. R., Homan, P. A., Chantarat, T., and Davis, B. A. (2022). Improving the measurement of structural racism to achieve antiracist health policy: study examines measurement of structural racism to achieve antiracist health policy. *Health Aff.* 41, 179–186. doi: 10.1377/hlthaff.2021.01489
- Hasin, D. S., Sarvet, A. L., Meyers, J. L., Saha, T. D., Ruan, W. J., Stohl, M., et al. (2018). Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *J. Am. Med. Assoc. Psychiatry* 75, 336–346. doi: 10.1001/jamapsychiatry.2017.4602
- Ismail, Z., Elbayoumi, H., Fischer, C. E., Hogan, D. B., Millikin, C. P., Schweizer, T., et al. (2017). Prevalence of depression in patients with mild cognitive impairment: a systematic review and meta-analysis. *J. Am. Med. Assoc. Psychiatry* 74, 58–67. doi: 10.1001/jamapsychiatry.2016.3162
- Jack, C. R., Thorneau, T. M., Weigand, S. D., Wiste, H. J., Knopman, D. S., Vemuri, P., et al. (2019). Prevalence of biologically vs. clinically defined Alzheimer spectrum entities using the National Institute on Aging-Alzheimer's Association research framework. *J. Am. Med. Assoc. Neurol.* 76, 1174–1183. doi: 10.1001/jamaneurol.2019.1971
- Jiao, Y. (2016). Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quant. Biol.* 4, 320–330. doi: 10.1007/s40484-016-0081-2
- Jorm, A. F. (2001). History of depression as a risk factor for dementia: an updated review. *Austr. N. Zeal. J. Psychiatry* 35, 776–781. doi: 10.1046/j.1440-1614.2001.00967.x
- Kalia, S., Greiver, M., Sullivan, F., Sejdic, E., Escobar, M., Gronsbell, J., et al. (2022). Marginal structural models using calibrated weights with SuperLearner: application to longitudinal diabetes cohort. *Int. J. Popul. Data Sci.* 7:1783. doi: 10.23889/ijpds.v7i3.1783
- Kessing, L. V., and Andersen, P. K. (2004). Does the risk of developing dementia increase with the number of episodes in patients with depressive disorder and in patients with bipolar disorder? *J. Neurol. Neurosurg. Psychiatry* 75, 1662–1666. doi: 10.1136/jnnp.2003.031773
- Köhler, S., Cierpinsky, K., and Kronenberg, G. (2016). The serotonergic system in the neurobiology of depression: relevance for novel antidepressants. *J. Psychopharmacol.* 30, 13–22. doi: 10.1177/0269881115609072
- Lee, E., Gatz, M., Tseng, C., Schneider, L. S., Pawluczky, S., Wu, A. H., et al. (2019). Evaluation of Medicare claims data as a tool to identify dementia. *J. Alzheimer's Dis.* 67, 769–778. doi: 10.3233/JAD-181005
- Lynch, M. (2018). *New Alzheimer's Association Report Reveals Sharp Increases in Alzheimer's Prevalence, Deaths, Cost of Care*. Chicago, IL: Alzheimer's & Dementia.
- MacKay, E. J., Stubna, M. D., Chivers, C., Draugelis, M. E., Hanson, W. J., Desai, N. D., et al. (2021). Application of machine learning approaches to administrative claims data to predict clinical outcomes in medical and surgical patient populations. *PLoS ONE* 16:e0252585. doi: 10.1371/journal.pone.0252585
- Mehta, K. M. (2017). Systematic review of dementia prevalence and incidence in United States race/ethnic populations. *Alzheimer's Dement.* 13, 72–83. doi: 10.1016/j.jalz.2016.06.2360
- Orimaye, S. O., Goodkin, K., Riaz, O. A., Salcedo, J. M., Al-Khateeb, T., Awujoola, A. O., et al. (2020a). A machine learning-based linguistic battery for diagnosing mild cognitive impairment due to Alzheimer's disease. *PLoS ONE* 15:e0229460. doi: 10.1371/journal.pone.0229460
- Orimaye, S. O., Southerland, J. L., and Oke, A. O. (2020b). Increased prevalence in Alzheimer disease in the northeast Tennessee region of the United States. *South Med. J.* 113, 351–355. doi: 10.14423/SMJ.0000000000001116
- Orimaye, S. O., and Wong, J. S. M. (2018). Deep language space neural network for classifying mild cognitive impairment and Alzheimer-type dementia. *PLoS ONE* 13:205636. doi: 10.1371/journal.pone.0205636
- Orimaye, S. O., Wong, J. S. M., Golden, K. J., and Wong, C. P. (2017). Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. *BMC Bioinform.* 18:34. doi: 10.1186/s12859-016-1456-0
- Owby, R. L., Crocco, E., Acevedo, A., and John, V. (2006). Depression and risk for Alzheimer disease: systematic review, meta-analysis, and meta-regression analysis. *Arch. Gen. Psychiatry* 63, 530–538. doi: 10.1001/archpsyc.63.5.530
- Peters, J., and Bühlmann, P. (2016). Causal inference by using invariant prediction: identification and confidence intervals. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 2016, 947–1012. doi: 10.1111/rssb.12167
- Prosperi, M., Guo, Y., Sperrin, M., Koopman, J. S., Min, J. S., He, X., et al. (2020). Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nat. Mach. Intell.* 2, 369–375. doi: 10.1038/s42256-020-0197-y
- Reiffel, J. A. (2020). Propensity score matching: the 'devil is in the details' where more may be hidden than you know. *Am. J. Med.* 133, 178–181. doi: 10.1016/j.amjmed.2019.08.055
- Rezvan, P. H., and Lee, K. J. (2015). The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Med. Res. Methodol.* 15, 1–14. doi: 10.1186/s12874-015-0022-1
- Rufó, D. D., Debele, T. G., and Ibenhal, A. (2021). Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM). *Diagnostics* 11:1714. doi: 10.3390/diagnostics11091714
- Sachdev, P. S., Blacker, D., Blazer, D. G., Ganguli, M., Jeste, D. V., Paulsen, J. S., et al. (2014). Classifying neurocognitive disorders: the DSM-5 approach. *Nat. Rev. Neurol.* 10, 634–642. doi: 10.1038/nrneurol.2014.181
- Smith, L. H. (2021). *Bias Bounds and Target Trials for Causal Inference in Observational Epidemiology*.
- Snowden, M. B., Atkins, D. C., Steinman, L. E., Bell, J. F., Bryant, L. L., Copeland, C., et al. (2015). Longitudinal association of dementia and depression. *Am. J. Geriatr. Psychiatry* 23, 897–905. doi: 10.1016/j.jagp.2014.09.002
- Steffens, D. C. (2017). Late-life depression and the prodromes of dementia. *J. Am. Med. Assoc. Psychiatry* 74, 673–674. doi: 10.1001/jamapsychiatry.2017.0658
- Tan, S. B., DeSouza, P., and Raifman, M. (2022). Structural racism and COVID-19 in the USA: a county-level empirical analysis. *J. Racial Ethn. Health Disparit.* 9, 236–246. doi: 10.1007/s40615-020-00948-8
- Thapar, A., Eyre, O., and Patel, V. (2022). Depression in young people. *Lancet* 400, 617–631. doi: 10.1016/S0140-6736(22)01012-1
- Velasquez, D. E., and Orav, E. J. (2023). Enrollment and characteristics of dual-eligible medicare and medicaid beneficiaries in integrated care programs: study examines the enrollment and characteristics of dual-eligible medicare and medicaid beneficiaries in integrated care programs. *Health Aff.* 42, 683–692. doi: 10.1377/hlthaff.2022.01321
- Verbakel, J. Y., Steyerberg, E. W., Uno, H., De Cock, B., Wynants, L., Collins, G. S., et al. (2020). ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. *J. Clin. Epidemiol.* 126, 207–216. doi: 10.1016/j.jclinepi.2020.01.028
- Wade, C. (2020). *Hands-On Gradient Boosting With XGBoost and Scikit-Learn: Perform Accessible Machine Learning and Extreme Gradient Boosting With Python*. Birmingham: Packt Publishing Ltd.

Wang, L., and Chen, J. (2019). Deep learning based epidemic forecasting with synthetic information. *Proc. AAAI Conf. Artif. Intell.* 33, 9607–9612. doi: 10.1609/aaai.v33i01.33019607

Zhang, B., Ren, J., Cheng, Y., and Wang, B. (2019). Health data driven on continuous blood pressure prediction based on gradient boosting decision tree algorithm. *IEEE Access* 7, 32423–32433. doi: 10.1109/ACCESS.2019.2902217

Zhou, X., Zhang, W., Chen, Z., and Diao, S. (2021). Efficient neural network training via forward and backward propagation sparsification. *Adv. Neural Inf. Process Syst.* 34:5685. doi: 10.48550/arXiv.2111.05685

Zissimopoulos, J. M., Tysinger, B. C., and St. Clair, P. A. (2018). The impact of changes in population health and mortality on future prevalence of Alzheimer's disease and other dementias in the United States. *J. Gerontol.* 73, S38–S47. doi: 10.1093/geronb/gbx147