

RESEARCH ARTICLE

Predicting malnutrition from longitudinal patient trajectories with deep learning

Boyang Tom Jin¹, Mi Hyun Choi², Meagan F. Moyer³, David A. Kim^{4*}

1 Department of Computer Science, Stanford University, Stanford, California, United States of America, **2** Department of Bioengineering, Stanford University, Stanford, California, United States of America, **3** Department of Digital Health Care Integration, Stanford Health Care, Stanford, California, United States of America, **4** Department of Emergency Medicine, Stanford University, Stanford, California, United States of America

* davidak@stanford.edu

Abstract

Malnutrition is common, morbid, and often correctable, but subject to missed and delayed diagnosis. Better screening and prediction could improve clinical, functional, and economic outcomes. This study aimed to assess the predictability of malnutrition from longitudinal patient records, and the external generalizability of a predictive model. Predictive models were developed and validated on statewide emergency department (ED) and hospital admission databases for California, Florida and New York, including visits from October 1, 2015 to December 31, 2018. Visit features included patient demographics, diagnosis codes, and procedure categories. Models included long short-term memory (LSTM) recurrent neural networks trained on longitudinal trajectories, and gradient-boosted tree and logistic regression models trained on cross-sectional patient data. The dataset used for model training and internal validation (California and Florida) included 62,811 patient trajectories (266,951 visits). Test sets included 63,997 (California), 63,112 (Florida), and 62,472 (New York) trajectories, such that each cohort's composition was proportional to the prevalence of malnutrition in that state. Trajectories contained seven patient characteristics and up to 2,008 diagnosis categories. Area under the receiver-operating characteristic (AUROC) and precision-recall curves (AUPRC) were used to characterize prediction of first malnutrition diagnoses in the test sets. Data analysis was performed from September 2020 to May 2021. Between 4.0% (New York) and 6.2% (California) of patients received malnutrition diagnoses. The longitudinal LSTM model produced the most accurate predictions of malnutrition, with comparable predictive performance in California (AUROC 0.854, AUPRC 0.258), Florida (AUROC 0.869, AUPRC 0.234), and New York (AUROC 0.869, AUPRC 0.190). Deep learning models can reliably predict malnutrition from existing longitudinal patient records, with better predictive performance and lower data-collection requirements than existing instruments. This approach may facilitate early nutritional intervention via automated screening at the point of care.

OPEN ACCESS

Citation: Jin BT, Choi MH, Moyer MF, Kim DA (2022) Predicting malnutrition from longitudinal patient trajectories with deep learning. PLoS ONE 17(7): e0271487. <https://doi.org/10.1371/journal.pone.0271487>

Editor: Vijayalakshmi Kakulapati, Sreenidhi Institute of Science and Technology, INDIA

Received: July 9, 2021

Accepted: July 4, 2022

Published: July 28, 2022

Copyright: © 2022 Jin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data used in the study are available upon request from the California Department of Health Care Access and Information (CDHCAI, <https://hcai.ca.gov/data-and-reports/request-data>) and the Healthcare Cost and Utilization Project (HCUP, https://www.hcup-us.ahrq.gov/tech_assist/centdist.jsp). Data use agreements with CDHCAI and HCUP expressly prohibit researchers uploading their datasets to separate repositories. The exact datasets used in this study are available directly from CDHCAI and HCUP.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Malnutrition, in which inadequate or imbalanced nutritional availability leads to deleterious changes in body composition and function [1], may develop as a result of deficient intake, inflammation, hypermetabolism, or malabsorption [2]. 15–60% of hospitalized patients are estimated to be malnourished [3], but the diagnosis is often missed or delayed, leading to preventable adverse outcomes [4].

Many hospital accrediting agencies, such as the Joint Commission, mandate screening of all hospital patients for malnutrition [5]. Most hospitals rely on questionnaires such as the Nutritional Risk Screening 2002 (NRS-2002) [6] or the Malnutrition Universal Screening Tool (MUST) [7], but no universal standard exists [8,9]. These questionnaires commonly measure recent weight loss, body mass index, acute disease effects, and dietary intake. Screening questions are susceptible to subjective interpretation [9], and fail to identify many cases of malnutrition [10]. Therefore, malnutrition is often under-identified or undocumented.

Early detection and treatment of malnutrition has a major impact on clinical, functional, and economic outcomes [10]. Compared to non-malnourished patients, malnourished patients experience hospital admissions of twice the duration and cost. Malnourished inpatients have up to 4.7 times the mortality of the general inpatient population [11]. A malnourished patient is 1.6 times more likely to be readmitted within 30 days of discharge than a non-malnourished patient, and these readmitted patients are twice as likely to be diagnosed with a serious infection, compared to readmitted non-malnourished patients [11].

Critically, malnutrition, once identified, can be treated with evidence-based and relatively low-cost nutritional interventions [12]. Accurate identification and documentation of malnutrition is also necessary for coding, reimbursement, risk-adjustment, and overall accuracy of patient data [13]. In this way, early and accurate identification of malnutrition can improve patient and health system outcomes [5], raising the need for universal and accurate screening.

Despite the importance of early detection, existing methods to screen for malnutrition have substantial limitations. One common screening instrument, MUST, achieved an area under the receiver-operating characteristic curve (AUROC) of 0.662 in one recent evaluation [14]. Recent attempts to improve the detection or prediction of malnutrition have focused on the evaluation of diagnostic criteria, or the optimization of information flow within the hospital [15–18]. Two recent studies have applied random forests, a type of ensemble machine learning model [19], to the prediction of malnutrition. MUST-Plus, a random forests model relying heavily on anthropometric and laboratory values, achieved an AUROC of 0.835 for the detection of malnutrition in one tertiary health center [14]. Talukder and Ahammed applied several conventional machine learning algorithms to the prediction of malnutrition in a high-prevalence pediatric cohort in Bangladesh. Though this study did not report AUROC, their best-performing model was a random forests model trained on demographic and anthropometric information, achieving an overall accuracy of 68.51% [20]. Notably, both of these studies used conventional, cross-sectional machine learning models (random forests), depended heavily on laboratory or anthropometric measurements not collected in all health care encounters, and failed to demonstrate external generalizability.

The present study reports the development and external validation of deep learning models for the prediction of malnutrition from longitudinal patient records. To maximize potential applicability in screening, these models use only passively collected demographic, diagnostic, and procedural information, and the evolution of these features over time, to identify patients who will receive a first diagnosis of malnutrition. To our knowledge, this is the first study to predict malnutrition with demographic and diagnostic information alone, without requiring laboratory or anthropometric results, and to specifically characterize the time interval between

prediction of malnutrition and actual diagnosis, during which a focused nutritional assessment and intervention could be delivered.

Materials and methods

Data sources and transformations

The study data consisted of comprehensive, statewide administrative databases of emergency department (ED) visits and hospital admissions for California (California Department of Health Care Access and Information), Florida, and New York (Healthcare Cost and Utilization Project). These data sources were selected because they represent large, diverse states, allow tracking of individual patient trajectories across time, and by virtue of their comprehensive, all-payer nature, reduce the risk of results biased to a specific population (such as a specific hospital system or commercial insurer). The study data included 43M Emergency Department (ED) visits and hospital admissions, representing the trajectories of 5.9M unique adult patients in California (CA), Florida (FL), and New York (NY), who had at least four visits between October 1, 2015, when the 10th International Statistical Classification of Diseases and Related Health Problems (ICD-10) diagnostic coding went into effect, and December 31, 2018. Each de-identified visit record included patient demographics (age, sex, race, state income quartile of patient's zip code), visit characteristics (primary payer, length of stay, number of days since last visit), up to 24 ICD-10 diagnosis codes (truncated to the first three characters), and up to 20 procedure codes (ICD-10 for admissions, CPT-4 for ED visits), which were standardized to Clinical Classification Software (CCS) procedure categories [21]. Patients with inconsistent ages were excluded.

Patients were classified as malnourished upon receipt of one or more diagnoses of protein-calorie malnutrition or cachexia (S1 Table), selected based on the 2016 HCUP report on malnutrition, and reviewed by a Stanford Health Care registered dietitian [11]. For malnourished patients, trajectories consisted of 3–5 visits preceding the first diagnosis of malnutrition. The visit during which malnutrition was diagnosed was excluded, with the goal of developing models to facilitate early identification of malnutrition. Control trajectories consisted of ED visits and admissions without a malnutrition diagnosis during the study period. As for case trajectories, the last visit was excluded, and trajectories consisted of 3–5 visits.

Training, validation, and test sets

CA and FL trajectories were combined to produce training (51,710 patients; 219,796 visits) and internal validation datasets (11,101 patients; 47,155 visits) with similar numbers of malnourished (39.6%) patients.

To simulate the effectiveness of models as clinical screening instruments, test sets were produced for each state, randomly sampling malnourished and non-malnourished patient trajectories in proportion to the prevalence of malnourished patients in the complete statewide datasets: CA (6.2%), FL (4.9%), and NY (4.0%). The size of the test sets (CA = 63,997, FL = 63,122, NY = 62,472 trajectories) were selected for computational tractability. Although models were trained on data from CA and FL, none of the test sets contained patients used in training or validation (S1 Fig).

Diagnosis and procedure code representations

In the primary models, diagnoses, procedures, sex, race, income quartile, and payer were represented using one-hot encoding (binary indicators). Age, days since last visit, and admission

length-of-stay were normalized to the range [0,1]. Each visit was thus represented as a 2029-dimensional vector (S2 Fig).

Auxiliary models used pre-trained embeddings developed based on coincident diagnosis and procedure codes (S1 Methods). Each code in each patient visit was represented by an embedding vector (S3 Fig), which was averaged to produce a dense embedding in place of the one-hot encodings (S4 Fig). Each visit could then be represented using a smaller, dense vector of length 53–277 (32–256 length embeddings + 21 one-hot encoded demographic features), depending on the size of the embedding used.

Predictive models

Classifiers were trained to predict whether a patient would receive a malnutrition diagnosis. The demographic baseline models consisted of logistic regression, random forest (n_estimators = 100, max_depth = None) [22], and gradient-boosted tree ensemble classifiers (XGBoost: n_estimators = 100, max_depth = 3) [23], trained on the demographic variables from the last visit of each patient trajectory. An XGBoost model was also trained using all codes from all visits in the trajectory (i.e., “flattened” codes, with temporal and sequence information discarded).

Deep learning models were developed using the long short-term memory (LSTM) architecture [24], a type of recurrent neural network. The LSTM architecture was selected for its performance on time-series data, and demonstrated utility for medical trajectories [25,26]. In this application, LSTMs are used to model the temporal relationships between visits. LSTM models consisted of a bidirectional LSTM layer with 32 units, followed by two fully connected layers, the first with 128 units and a dropout rate of 0.2, and the second with a binary outcome (malnourished or non-malnourished). Binary cross-entropy loss and the Adam optimizer were used with a learning rate of 0.001. LSTM models were developed on both one-hot encoded and embedded visit representations. A cross-sectional model was also trained, without an LSTM layer, on the last visit in each trajectory, with the same hyperparameters except for the learning rate, which was set to 0.0001.

Model predictions were interpreted using model-agnostic Shapley values, which reflect the marginal contribution of each variable in each patient’s trajectory to the model’s prediction [27].

Stratified analyses, ablation studies, lead time experiments

In order to understand the performance of the models in specific populations, the test datasets were stratified on demographic features. For instance, the test patient trajectories were split into elderly (≥ 65 years) and non-elderly cohorts (< 65 years), and predicted malnutrition for each of these cohorts using the same model.

In order to understand the relative importance of different features, beyond the Shapley values described above, ablation studies were conducted, retraining the same models but excluding a specific feature category, such as all diagnosis codes, for each retrained model.

Prediction *lead time* refers to the number of days between the last visit observed for each patient (i.e., the point of prediction), and the visit containing the first malnutrition diagnosis. For instance, to assess model performance in predicting malnutrition with at least 30 days lead time, any visits occurring less than 30 days before the malnutrition diagnosis were excluded from trajectories. To control for the reduction in trajectory lengths with increasing lead time, performance with trajectories restricted to exactly five visits, after lead time filtering, was also evaluated.

Statistical analysis of model results

Model performance was characterized using area under the receiver-operating characteristic curve (AUROC), reflecting the relationship between sensitivity and specificity, and area under the precision-recall curve (AUPRC), reflecting the relationship between positive predictive value (PPV, 'precision') and sensitivity ('recall'), which may better reflect the prediction of rare outcomes than AUROC [28]. The 95% confidence intervals (CI) of AUROC and AUPRC were estimated using the Wilson Score interval method [29]. CIs were validated for the best-performing models with bootstrap sampling.

Results

Patient characteristics

Malnourished patients had similar characteristics in all states (S2 Table). Among admitted patients, malnourished patients had longer hospital stays, with median length-of-stay between 6 (CA) and 8 (NY) days, compared to 3 (CA, FL) or 4 (NY) days for non-malnourished patients. Malnourished patients were older than non-malnourished patients by a median of 20 (CA) to 23 years (FL, NY). Malnourished patients had more frequent visits, with median intervals between visits ranging from 24 (CA) to 28 (NY) days, compared to 57 (NY) to 77 (CA) days for non-malnourished patients. Women were diagnosed with malnutrition at higher rates than men. By primary payer, Medicare was associated with the highest rate of malnourished patients, consistent with the advanced age of these patients. Among racial/ethnic groups, Asians had the highest prevalence of malnutrition, and Hispanics had the lowest. The proportion of patients with malnutrition diagnoses was highest for zip codes in the highest (FL, NY) and second highest (CA) income quartiles. Tables 1 and S2 show patient characteristics of the subsets and the full datasets, respectively.

Prediction of malnutrition. The best-performing model for the prediction of malnutrition was the longitudinal LSTM using one-hot encoding of diagnosis and procedure codes (Fig 1). This model, trained on patient trajectories from CA and FL, and tested on distinct cohorts from CA and FL and a completely independent dataset from NY, achieved uniformly strong predictive performance on independent test cohorts, with AUROCs ranging from 0.854 (CA, 95% CI 0.851–0.857) to 0.869 (FL, NY, 95% CI 0.866–0.872), and AUPRCs ranging from 0.190 (NY, 95% CI 0.187–0.193) to 0.258 (CA, 95% CI 0.255–0.261). Model performance is summarized in S3 Table, with bootstrap validation of CIs described in S4 Table. The best-performing embeddings (length 256, S5 Table) were slightly less performant by AUPRC, compared to the one-hot encoding strategy.

Models using longitudinal trajectories performed substantially better than models using only data from the patient's last visit. Of the cross-sectional models, the best-performing model was a fully connected neural network, achieving AUROCs from 0.831 (CA, 95% CI 0.828–0.834) to 0.843 (FL, 95% CI 0.840–0.846), and AUPRCs from 0.162 (NY, 95% CI 0.159–0.165) to 0.233 (CA, 95% CI 0.230–0.236).

A hybrid approach, using XGBoost and collapsing features from all visits, achieved performance between that of the cross-sectional and longitudinal models: AUROCs from 0.836 (CA, 95% CI 0.833–0.839) to 0.852 (NY, 95% CI 0.849–0.855), and AUPRCs from 0.177 (NY, 95% CI 0.174–0.180) to 0.238 (CA, 95% CI 0.235–0.241).

Patient-level risk factors

The implementation of Shapley values allowed determination of patient-level and aggregate predictors of malnutrition (Fig 2). Major predictors of malnutrition identified by the

Table 1. Patient characteristics in test sets used to assess predictive performance.

	California			Florida			New York		
	Total	Malnourished	Control	Total	Malnourished	Control	Total	Malnourished	Control
Sample size	63997	3997	60000	63122	3122	60000	62472	2472	60000
Length of stay, median (IQR), days	3 (2–6)	6 (3–11)	3 (2–6)	3 (2–6)	7 (4–13)	3 (2–5)	4 (2–8)	8 (4–15)	4 (2–7)
Age, median (IQR), years	52 (33–69)	70 (58–82)	50 (32–67)	50 (32–69)	70 (58–82)	48 (32–68)	49 (32–67)	71 (58–82)	48 (31–66)
Time from last visit, median (IQR), days	72 (15–199)	24 (7–75)	77 (16–207)	73 (18–198)	27 (11–77)	77 (18–204)	55 (14–146)	27 (11–68)	57 (14–150)
Income, n (%)									
Unknown	2527 (3.949)	158 (3.953)	2369 (3.948)	2169 (3.436)	114 (3.652)	2055 (3.425)	834 (1.335)	37 (1.497)	797 (1.328)
Quartile 1	13635 (21.306)	815 (20.390)	12820 (21.367)	17472 (27.680)	800 (25.625)	16672 (27.787)	26548 (42.496)	816 (33.010)	25732 (42.887)
Quartile 2	18903 (29.537)	1077 (26.945)	17826 (29.710)	15207 (24.091)	725 (23.222)	14482 (24.137)	10527 (16.851)	387 (15.655)	10140 (16.900)
Quartile 3	16815 (26.275)	1055 (26.395)	15760 (26.267)	17668 (27.990)	838 (26.842)	16830 (28.050)	11504 (18.415)	476 (19.256)	11028 (18.380)
Quartile 4	12117 (18.934)	892 (22.317)	11225 (18.708)	10606 (16.802)	645 (20.606)	9961 (16.602)	13059 (20.904)	756 (30.583)	12303 (20.505)
Payer, n (%)									
Medicaid	23993 (37.491)	751 (18.789)	23242 (38.737)	12113 (19.190)	321 (10.282)	11792 (19.653)	24015 (38.441)	414 (16.748)	23601 (39.335)
Medicare	22281 (34.816)	2679 (67.025)	19602 (32.670)	22682 (35.934)	2204 (70.596)	20478 (34.130)	21077 (33.738)	1682 (68.042)	19395 (32.325)
Other	1698 (2.653)	62 (1.551)	1636 (2.727)	3899 (6.177)	119 (3.812)	3780 (6.300)	1635 (2.617)	38 (1.537)	1597 (2.662)
Private	13446 (21.010)	471 (11.784)	12975 (21.625)	14263 (22.596)	370 (11.851)	13893 (23.155)	12250 (19.609)	315 (12.743)	11935 (19.892)
Self	2579 (4.030)	34 (0.851)	2545 (4.242)	10165 (16.104)	108 (3.459)	10057 (16.762)	3495 (5.595)	23 (0.930)	3472 (5.787)
Race, n (%)									
Asian	3504 (5.475)	323 (8.081)	3181 (5.302)	305 (0.483)	20 (0.641)	285 (0.475)	1502 (2.404)	56 (2.265)	1446 (2.410)
Black	8041 (12.565)	482 (12.059)	7559 (12.598)	15202 (24.084)	511 (16.368)	14691 (24.485)	15624 (25.010)	443 (17.921)	15181 (25.302)
Hispanic	20459 (31.969)	864 (21.616)	19595 (32.658)	11237 (17.802)	378 (12.108)	10859 (18.098)	9751 (15.609)	239 (9.668)	9512 (15.853)
Native	341 (0.533)	18 (0.450)	323 (0.538)	64 (0.101)	1 (0.032)	63 (0.105)	175 (0.280)	5 (0.202)	170 (0.283)
Other	2953 (4.614)	174 (4.353)	2779 (4.632)	1188 (1.882)	58 (1.858)	1130 (1.883)	6291 (10.070)	234 (9.466)	6057 (10.095)
White	28699 (44.844)	2136 (53.440)	26563 (44.272)	35126 (55.648)	2154 (68.994)	32972 (54.953)	29129 (46.627)	1495 (60.477)	27634 (46.057)
Sex, n (%)									
Female	37399 (58.439)	2017 (50.463)	35382 (58.970)	38343 (60.744)	1619 (51.858)	36724 (61.207)	36928 (59.111)	1245 (50.364)	35683 (59.472)
Male	26598 (41.561)	1980 (49.537)	24618 (41.030)	24779 (39.256)	1503 (48.142)	23276 (38.793)	25544 (40.889)	1227 (49.636)	24317 (40.528)

<https://doi.org/10.1371/journal.pone.0271487.t001>

model included advanced age, indicators of poor overall health (heart failure, COPD, malignancies), as well as conditions limiting nutritional intake (aphagia, food allergy, homelessness).

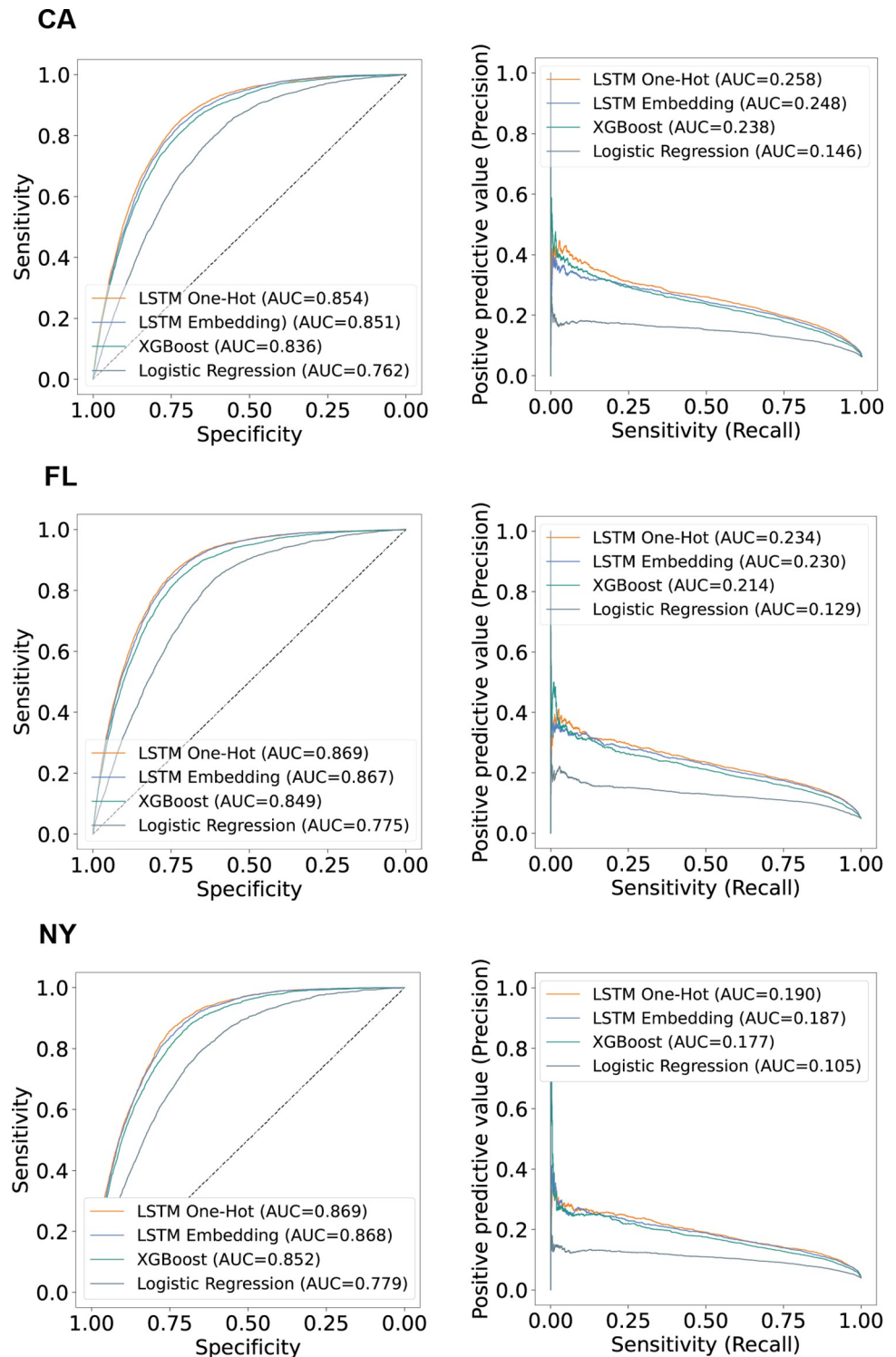


Fig 1. Receiver-operating characteristic (left) and precision-recall curves (right) for malnutrition prediction models, in three states. The long short-term memory (LSTM) models used longitudinal patient trajectories. The gradient-boosted tree model (XGBoost) used collapsed features from all visits. The logistic regression baseline model used only demographic data.

<https://doi.org/10.1371/journal.pone.0271487.g001>

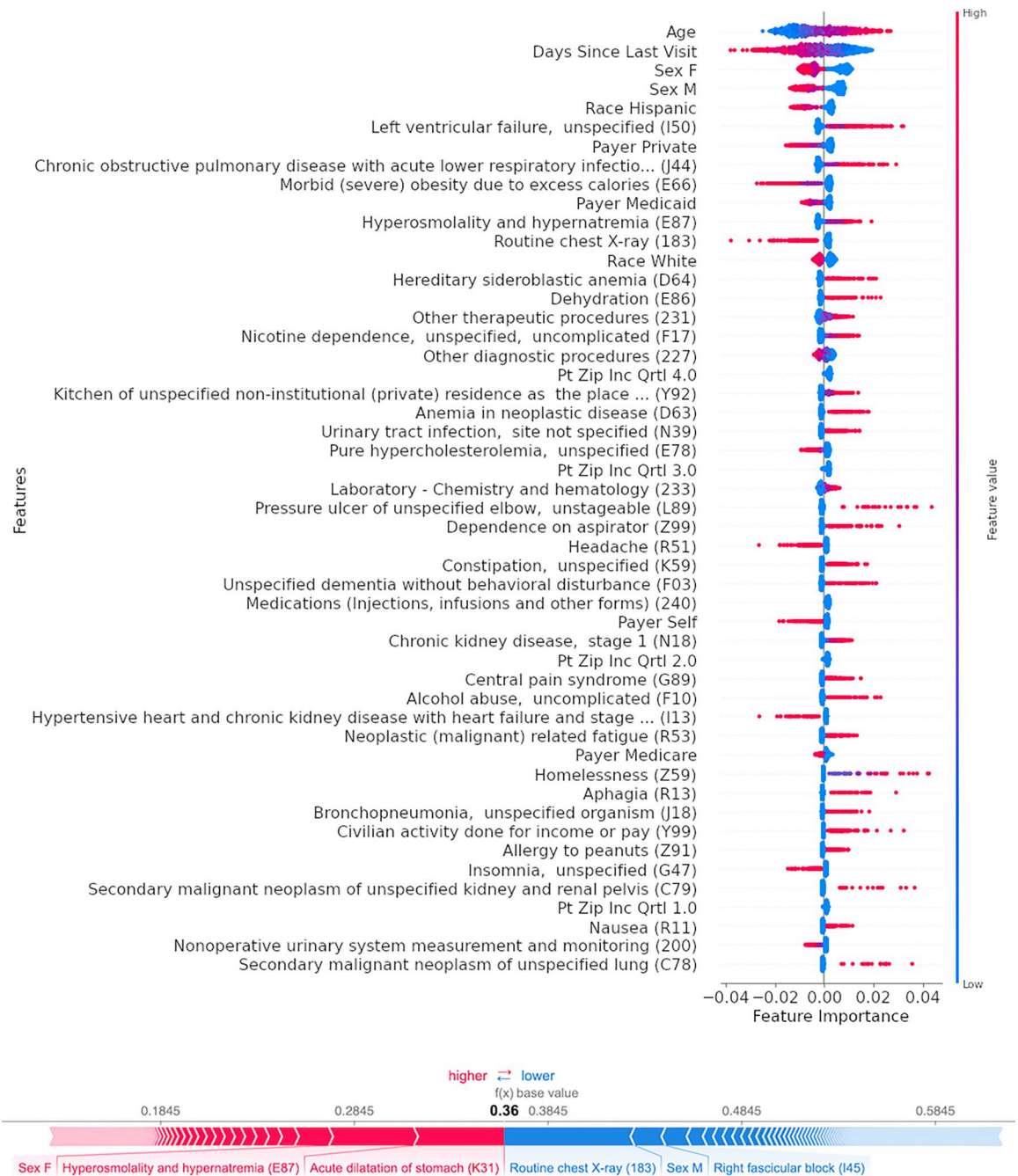


Fig 2. Overall and patient-specific features predictive of malnutrition. Upper panel shows the top 50 features predictive of malnutrition from the best performing model (LSTM with one-hot encoded trajectories) in descending order of influence (mean Shapley values over five visits for 1000 patients in California). Bottom panel shows risk factors for an individual malnourished patient.

<https://doi.org/10.1371/journal.pone.0271487.g002>

Stratified analyses

In stratified analyses (S6 Table), predictive performance reflected the size and the baseline rates of malnutrition in the population of interest. For instance, models had better PPV (reflected in AUPRC) in elderly patients compared to non-elderly patients, because PPV is dependent on base rates of malnutrition, which are higher in the elderly.

Ablation studies

When ablating categories of features to assess their relative impacts on model performance, diagnosis codes had the greatest collective effect. This was most prominent in CA, where the AUROC decreased from 0.854 (95% CI, 0.851–0.857) to 0.806 (95% CI, 0.803–0.809) with ablation of diagnostic information. Next most important were demographic data, then CCS procedure categories. Models exposed to procedure codes alone achieved an AUROC of 0.717 (95% CI, 0.714–0.720) in CA, lower than the baseline demographic models (S7 Table).

Effect of prediction lead time

Predictive performance declined with increasing lead time between the date of prediction (i.e., the last visit in each trajectory seen by the model) and the first diagnosis of malnutrition. Fewer visits are used for prediction as lead time increases, with up to 49% of the data in NY trajectories removed at a lead time of one year. Controlling for total number of visits, greater lead time remains associated with less accurate prediction (Fig 3). The steepest decline is seen between <1 month and 1–2 months, where, for instance, the AUPRC in CA drops from 0.258 (95% CI, 0.255–0.261) to 0.229 (95% CI, 0.226–0.232), though the corresponding AUROC dropped only from 0.854 (95% CI, 0.851–0.857) to 0.837 (95% CI, 0.834–0.840), suggesting a greater impact of lead time in lower-prevalence settings. AUPRC (reflective of positive predictive value) at over one year was 0.113 in CA (95% CI, 0.110–0.116), 0.104 in FL (95% CI, 0.101–0.107), and 0.075 in NY (95% CI, 0.072–0.078) (S8 Table).

Discussion

Malnutrition is a major global comorbidity, with deleterious effects on patient trajectories and healthcare costs. Evidence-based interventions can mitigate these effects [2], but malnutrition is often overlooked until its complications are already manifest, and more difficult to correct [30]. Moreover, non-standard screening practices can produce large variation in diagnosis

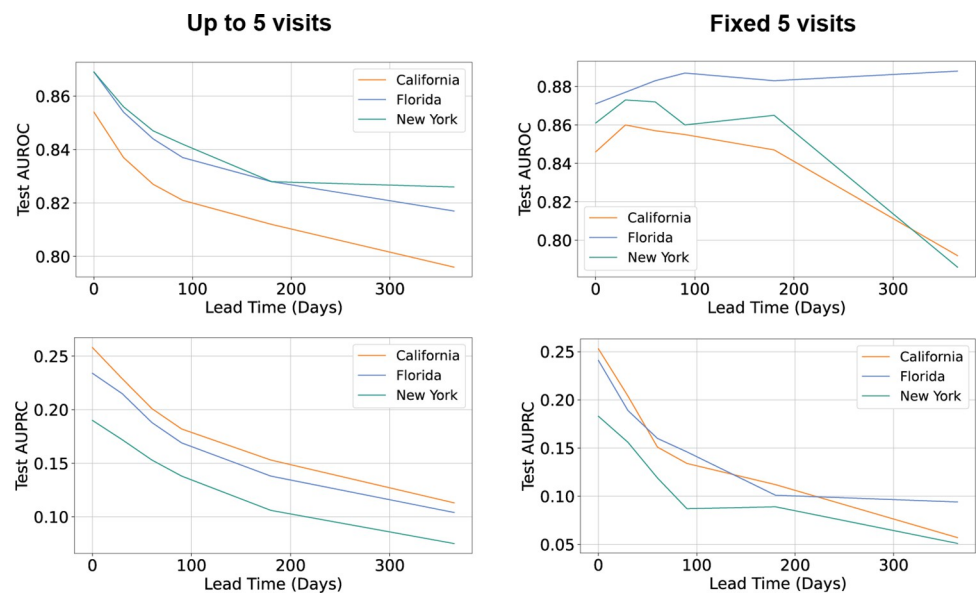


Fig 3. Effect of lead time on malnutrition prediction. Prediction performance of the best-performing model declines with longer intervals between prediction and diagnosis (left). Holding trajectory length constant, the detrimental effect of lead time is attenuated for AUROC but not for AUPRC (right).

<https://doi.org/10.1371/journal.pone.0271487.g003>

rates [31]. This study presents a deep learning approach to predicting malnutrition risk, using longitudinal modeling of ubiquitously recorded visit features (demographics, diagnoses, procedures). The best-performing models, which exploit the longitudinal structure of patient data, are highly accurate in predicting which patients will receive a diagnosis of malnutrition, and generalize extremely well to a population not used at all in model development. In many cases, predictions can be made with considerable lead time that could allow for focused evaluation and intervention.

The most comparable published model, MUST-Plus, is a random forest classifier trained on a single-center population used for both model development and validation, and achieved an AUROC of 0.835 [14]. In comparison, this study's longitudinal deep learning model achieves higher performance by AUROC in all three states (0.854–0.869), and demonstrates excellent generalizability to a large, diverse population not used in model development. Moreover, whereas existing models rely heavily on laboratory or anthropometric values requiring active collection at the point of screening [14,20], this study's approach derives most of its predictive ability from diagnostic trajectories that are passively and widely recorded, requiring no additional data collection for first-pass screening at the point of care.

While predictions were generally more accurate with shorter lead times between prediction and diagnosis (Fig 3), malnutrition diagnoses can still be predicted at clinically relevant levels even one year in advance. This provides ample opportunity to intervene on high-risk patients, which has clinical, operational, and economic benefits [32]. Somanchi et al., for instance, found that early nutritional intervention reduced average length of stay by 1.93 days, producing an annual savings of \$1.16 million for 1,275 patients [33].

Fig 4 illustrates how a model like those presented in this study could be applied in practice: upon a patient encounter (whether to clinic, ED, or hospital), the model is applied as a universal screen and flags high-risk patients without an existing malnutrition diagnosis for clinical review, using pre-existing demographic, diagnostic, and procedural data from the patient's electronic health record. Personalized patient-level risk factors as shown in Fig 2 provide at-a-glance model interpretation that can evolve over time as additional patient visits are recorded. The provider can then assess the patient and determine whether the patient would benefit

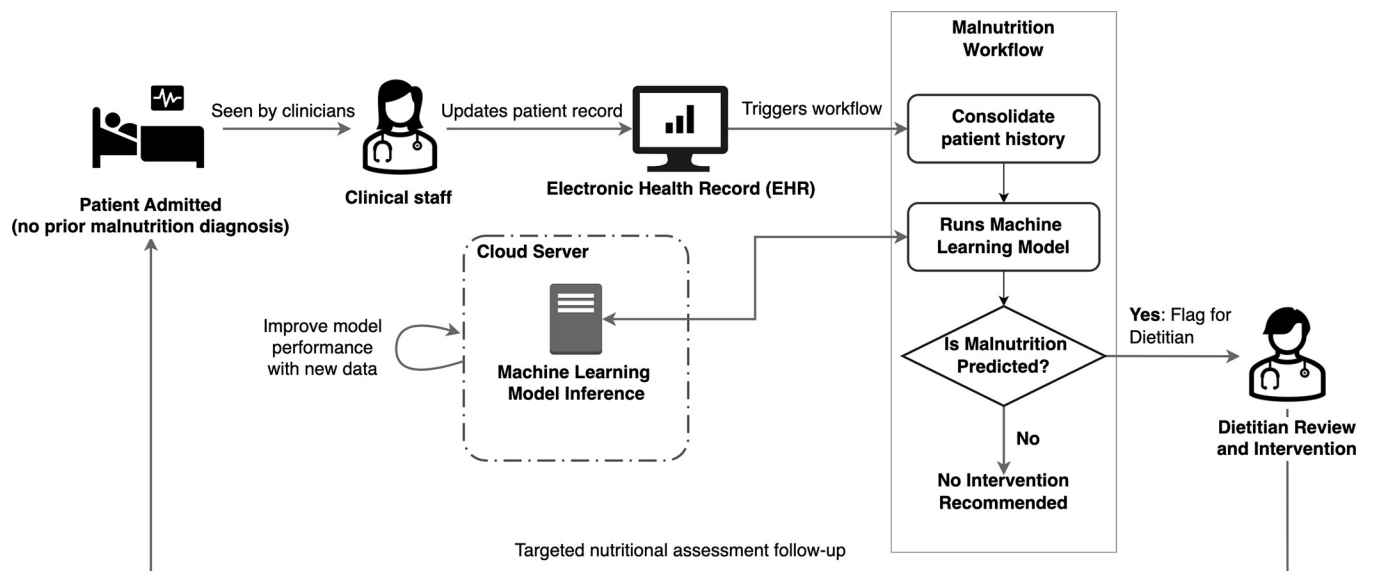


Fig 4. Proposed clinical implementation strategy to identify patients at risk for malnutrition and trigger dietitian referral or consultation.

<https://doi.org/10.1371/journal.pone.0271487.g004>

from consultation with or referral to a registered dietitian, who can then perform a targeted nutritional assessment, and suggest interventions tailored to the patient's specific needs and risks. In current practice, malnutrition is seldom diagnosed in the ED [34], despite its being a more common point of contact with the healthcare system than hospital admissions. Automating screening for malnutrition in the ED, where providers are otherwise preoccupied with more acute diagnoses, could expose a larger and more diverse population to nutritional screening and concomitant interventions.

This study has several limitations. Because this study relied on large administrative datasets and recorded diagnoses for identification of malnutrition, it was difficult to discern whether a patient identified by the model as high-risk had yet to develop malnutrition, or was already malnourished but had not yet received a diagnosis. For many practical purposes, however, the distinction may not be important, as malnourished patients without diagnoses of malnutrition are less likely to have received focused nutritional evaluation and intervention [35], so predicting either future or undiagnosed malnutrition may be of comparable clinical benefit. Though malnutrition includes inflammatory and hypermetabolic conditions in addition to deficient intake and malabsorption, these former are less commonly diagnosed, and are poorly characterized in the data. By considering only patient trajectories with at least three visits, patients with no or minimal prior medical history were excluded, which may bias the analyses towards sicker populations at higher risk for malnutrition, though the proportions of malnourished patients in this study's cohorts are still less than the 15–60% of hospitalized patients elsewhere estimated to be malnourished [3].

For similar reasons, relying on recorded diagnoses subjects the models to any biases in recording of diagnoses. For instance, if a group of patients had fewer diagnoses for a given rate of true malnutrition due to poorer access to healthcare, a model trained on this data could recapitulate this bias [36]. Indeed, the higher malnutrition rates in patients from wealthier communities could reflect such a bias, whereby patients from these communities are exposed to superior healthcare resources with greater capacity to diagnose and treat malnutrition. However, the limited reliance of the models on demographic compared to diagnostic features, and the excellent generalization of the models to independent test sets of different patient compositions, collectively suggest that this algorithmic screening approach is likelier to reduce than to exacerbate biases in the diagnosis of malnutrition.

Conclusions

This study suggests opportunities for early intervention in malnutrition via automated screening of longitudinal patient records with deep learning. The best-performing model outperformed existing instruments, and generalized well to an external validation cohort. Using only passively collected demographic, diagnostic, and procedural data, this model could be easily integrated into existing clinical systems. Despite external validation in retrospective data, the model presented in this study requires prospective clinical validation, and the development of a resource-efficient mechanism for dietitian consultation or referral, before consideration for clinical use.

Supporting information

S1 Fig. Cohort creation. ED visits and admissions are grouped by patient to form trajectories. A subset of the patient trajectories is randomly sampled to create training, validation, and test sets. The training and validation datasets are derived from Florida and California data, while New York is reserved as a fully independent test set. No patients in any of the test sets are used

in model development.

(PDF)

S2 Fig. One-hot encoding representation of visits. Patient trajectories are pre-processed into one-hot (binary) encoded representations.

(PDF)

S3 Fig. Diagnosis and procedure code embeddings. Pre-trained embeddings are created by training a two-layer fully connected network using a dataset created from target-context pairs of diagnostic and procedural codes from each visit.

(PDF)

S4 Fig. Dense embedding representation of visits. Patient trajectories are transformed into dense embeddings by averaging the pre-trained code embeddings within each visit.

(PDF)

S1 Table. ICD-10 diagnosis codes used as criteria for malnutrition.

(PDF)

S2 Table. Patient characteristics in full (non-sampled) datasets.

(PDF)

S3 Table. Comparison of model performance, by AUROC and AUPRC.

(PDF)

S4 Table. Comparison of analytical and bootstrap techniques for production of confidence intervals.

(PDF)

S5 Table. Effect of embedding length on prediction performance.

(PDF)

S6 Table. Prediction performance stratified by demographic characteristics.

(PDF)

S7 Table. Prediction performance in feature ablation studies.

(PDF)

S8 Table. Effect of lead time between prediction and diagnosis on prediction performance.

(PDF)

S1 Reference.

(PDF)

S1 Methods. Pre-trained embeddings.

(PDF)

Acknowledgments

Ethical approval

The study was approved by the Institutional Review Board of Stanford University.

Author Contributions

Conceptualization: Boyang Tom Jin, Mi Hyun Choi, David A. Kim.

Data curation: David A. Kim.

Formal analysis: David A. Kim.

Investigation: Boyang Tom Jin, Mi Hyun Choi, Meagan F. Moyer, David A. Kim.

Methodology: Boyang Tom Jin, Mi Hyun Choi, David A. Kim.

Project administration: David A. Kim.

Supervision: Meagan F. Moyer, David A. Kim.

Validation: Meagan F. Moyer.

Visualization: Boyang Tom Jin.

Writing – original draft: Boyang Tom Jin, Mi Hyun Choi, David A. Kim.

Writing – review & editing: Meagan F. Moyer, David A. Kim.

References

1. Soeters PB, Schols AMWJ. Advances in understanding and assessing malnutrition. *Curr Opin Clin Nutr Metab Care*. 2009; 12(5):487–494. <https://doi.org/10.1097/MCO.0b013e32832da243> PMID: 19512916
2. Barker LA, Gout BS, Crowe TC. Hospital malnutrition: prevalence, identification and impact on patients and the healthcare system. *Int J Environ Res Public Health*. 2011; 8(2):514–527. <https://doi.org/10.3390/ijerph8020514> PMID: 21556200
3. White JV, Guenter P, Jensen G, Malone A, Schofield M. Consensus statement: Academy of Nutrition and Dietetics and American Society for Parenteral and Enteral Nutrition: characteristics recommended for the identification and documentation of adult malnutrition (undernutrition). *JPEN J Parenter Enteral Nutr*. 2012; 36(3):275–283. <https://doi.org/10.1177/0148607112440285> PMID: 22535923
4. Keats EC, Jain RP, Bhutta ZA. Indirect health sector actions and supportive strategies to prevent malnutrition. *Curr Opin Clin Nutr Metab Care*. 2020; 23(3):190–195. <https://doi.org/10.1097/MCO.0000000000000653> PMID: 32167985
5. Patel V, Romano M, Corkins MR, DiMaria-Ghalili RA, Earthman C, Miller S et al. Nutrition Screening and Assessment in Hospitalized Patients: A Survey of Current Practice in the United States. *Nutr Clin Pract Off Publ Am Soc Parenter Enter Nutr*. 2014; 29(4):483–490. <https://doi.org/10.1177/0884533614535446> PMID: 24993585
6. Kondrup J, Rasmussen HH, Hamberg O, Stanga Z. Nutritional risk screening (NRS 2002): a new method based on an analysis of controlled clinical trials. *Clin Nutr Edinb Scotl*. 2003; 22(3):321–336. [https://doi.org/10.1016/s0261-5614\(02\)00214-5](https://doi.org/10.1016/s0261-5614(02)00214-5)
7. Stratton RJ, Hackston A, Longmore D, Dixon R, Price S, Stroud M et al. Malnutrition in hospital outpatients and inpatients: prevalence, concurrent validity and ease of use of the “malnutrition universal screening tool” (“MUST”) for adults. *Br J Nutr*. 2004; 92(5):799–808. <https://doi.org/10.1079/bjn20041258> PMID: 15533269
8. Anthony PS. Nutrition Screening Tools for Hospitalized Patients. *Nutr Clin Pract*. 2008; 23(4):373–382. <https://doi.org/10.1177/0884533608321130> PMID: 18682588
9. Reber E, Gomes F, Vasiloglou MF, Schuetz P, Stanga Z. Nutritional Risk Screening and Assessment. *J Clin Med*. 2019; 8(7). <https://doi.org/10.3390/jcm8071065> PMID: 31330781
10. Norman K, Pichard C, Lochs H, Pirlich M. Prognostic impact of disease-related malnutrition. *Clin Nutr*. 2008; 27(1):5–15. <https://doi.org/10.1016/j.clnu.2007.10.007> PMID: 18061312
11. Barrett ML, Bailey MK, Owens PL. Non-maternal and Non-neonatal Inpatient Stays in the United States Involving Malnutrition, 2016. ONLINE. August 30, 2018. U.S. Agency for Healthcare Research and Quality. Available: www.hcup-us.ahrq.gov/reports.jsp.
12. Gomes F, Baumgartner A, Bounoure L, Bally M, Deutz NE, Greenwald JL et al. Association of Nutritional Support With Clinical Outcomes Among Medical Inpatients Who Are Malnourished or at Nutritional Risk: An Updated Systematic Review and Meta-analysis. *JAMA Netw Open*. 2019; 2(11): e1915138. <https://doi.org/10.1001/jamanetworkopen.2019.15138> PMID: 31747030
13. Doley J, Phillips W. Coding for Malnutrition in the Hospital: Does It Change Reimbursement? *Nutr Clin Pract Off Publ Am Soc Parenter Enter Nutr*. 2019; 34(6):823–831. <https://doi.org/10.1002/ncp.10426> PMID: 31650622

14. Timsina P, Joshi HN, Cheng F-Y, Kersch I, Wilson S, Colgan C et al. MUST-Plus: A Machine Learning Classifier That Improves Malnutrition Screening in Acute Care Facilities. *J Am Coll Nutr*. 2021; 40(1):3–12. <https://doi.org/10.1080/07315724.2020.1774821> PMID: 32701397
15. Silver HJ, Pratt KJ, Bruno M, Lynch J, Mitchell K, McCauley SM. Effectiveness of the Malnutrition Quality Improvement Initiative on Practitioner Malnutrition Knowledge and Screening, Diagnosis, and Timeliness of Malnutrition-Related Care Provided to Older Adults Admitted to a Tertiary Care Facility: A Pilot Study. *J Acad Nutr Diet*. 2018; 118(1):101–109. <https://doi.org/10.1016/j.jand.2017.08.111> PMID: 29274640
16. Pratt KJ, Hernandez B, Blancato R, Blankenship J, Mitchell K. Impact of an interdisciplinary malnutrition quality improvement project at a large metropolitan hospital. *BMJ Open Qual*. 2020; 9(1). <https://doi.org/10.1136/bmjopen-2019-000735> PMID: 32213547
17. Green Corkins K. Nutrition-focused physical examination in pediatric patients. *Nutr Clin Pract Off Publ Am Soc Parenter Enter Nutr*. 2015; 30(2):203–209. <https://doi.org/10.1177/0884533615572654> PMID: 25739708
18. Jensen GL, Compher C, Sullivan DH, Mullin GE. Recognizing malnutrition in adults: definitions and characteristics, screening, assessment, and team approach. *JPEN J Parenter Enteral Nutr*. 2013; 37(6):802–807. <https://doi.org/10.1177/0148607113492338> PMID: 23969411
19. Breiman L. Random Forests. *Mach Learn*. 2001; 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
20. Talukder A, Ahammed B. Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. *Nutrition*. 2020; 78:110861. <https://doi.org/10.1016/j.nut.2020.110861> PMID: 32592978
21. Clinical Classifications Software (CCS) for ICD-10-PCS (beta version). Accessed April 26, 2021. <https://www.hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp>.
22. Couronné R, Probst P, Boulesteix A-L. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*. 2018; 19(1):270. <https://doi.org/10.1186/s12859-018-2264-5> PMID: 30016950
23. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. Association for Computing Machinery; 2016:785–794. <https://doi.org/10.1145/2939672.2939785>
24. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997; 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735> PMID: 9377276
25. Guo A, Beheshti R, Khan YM, Langabeer JR, Foraker RE. Predicting cardiovascular health trajectories in time-series electronic health records with LSTM models. *BMC Medical Informatics and Decision Making*. 2021; 21(5). <https://doi.org/10.1186/s12911-020-01345-1> PMID: 33407390
26. Liu S, Wang X, Xiang Y, Xu H, Wang H, Tang B. Multi-channel fusion LSTM for medical event prediction using EHRs. *Journal of Biomedical Informatics*. 2022; 127:10411. <https://doi.org/10.1016/j.jbi.2022.104011> PMID: 35176451
27. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. *ArXiv170507874 Cs Stat*. Published online November 24, 2017. Accessed April 26, 2021. <http://arxiv.org/abs/1705.07874>.
28. Ozenne B, Subtil F, Maucort-Boulch D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol*. 2015; 68(8):855–859. <https://doi.org/10.1016/j.jclinepi.2015.02.010> PMID: 25881487
29. Wei L, Hutson AD. A comment on sample size calculations for binomial confidence intervals. *J Appl Stat*. 2013; 40(2):311–319. <https://doi.org/10.1080/02664763.2012.740629> PMID: 26989292
30. Marik PE, Zaloga GP. Early enteral nutrition in acutely ill patients: a systematic review. *Crit Care Med*. 2001; 29(12):2264–2270. <https://doi.org/10.1097/00003246-200112000-00005> PMID: 11801821
31. Correia MITD. Addressing the Hidden Burden of Malnutrition for Hospitalized Patients. *J Acad Nutr Diet*. 2018; 118(1):37–39. <https://doi.org/10.1016/j.jand.2017.03.009> PMID: 28416433
32. Stratton RJ, King CL, Stroud MA, Jackson AA, Elia M. “Malnutrition Universal Screening Tool” predicts mortality and length of hospital stay in acutely ill elderly. *Br J Nutr*. 2006; 95(2):325–330. <https://doi.org/10.1079/bjn20051622> PMID: 16469149
33. Somanchi M, Tao X, Mullin GE. The facilitated early enteral and dietary management effectiveness trial in hospitalized patients with malnutrition. *JPEN J Parenter Enteral Nutr*. 2011; 35(2):209–216. <https://doi.org/10.1177/0148607110392234> PMID: 21378250
34. Lanctin DP, Merced-Nieves F, Mallett RM, Arensberg MB, Guenter P, Sulo S et al. Prevalence and Economic Burden of Malnutrition Diagnosis Among Patients Presenting to United States Emergency Departments. *Acad Emerg Med*. 2021; 28(3):325–335. <https://doi.org/10.1111/acem.13887> PMID: 31724782

35. Correia MITD, Waitzberg DL. The impact of malnutrition on morbidity, mortality, length of hospital stay and costs evaluated through a multivariate model analysis. *Clin Nutr Edinb Scotl*. 2003; 22(3):235–239. [https://doi.org/10.1016/s0261-5614\(02\)00215-7](https://doi.org/10.1016/s0261-5614(02)00215-7)
36. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019; 366(6464):447–453. <https://doi.org/10.1126/science.aax2342> PMID: 31649194