



This is a repository copy of *Personalised treatment for cognitive impairment in dementia : development and validation of an artificial intelligence model.*

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/183736/>

Version: Published Version

Article:

Liu, Q., Vaci, N., Koychev, I. et al. (4 more authors) (2022) Personalised treatment for cognitive impairment in dementia : development and validation of an artificial intelligence model. *BMC Medicine*, 20 (1). 45.

<https://doi.org/10.1186/s12916-022-02250-2>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.




eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

RESEARCH ARTICLE

Open Access



Personalised treatment for cognitive impairment in dementia: development and validation of an artificial intelligence model

Qiang Liu^{1*} , Nemanja Vaci², Ivan Koychev¹, Andrey Kormilitzin^{1,3}, Zhenpeng Li¹, Andrea Cipriani^{1,4} and Alejo Nevado-Holgado^{1,5,6}

Abstract

Background: Donepezil, galantamine, rivastigmine and memantine are potentially effective interventions for cognitive impairment in dementia, but the use of these drugs has not been personalised to individual patients yet. We examined whether artificial intelligence-based recommendations can identify the best treatment using routinely collected patient-level information.

Methods: Six thousand eight hundred four patients aged 59–102 years with a diagnosis of dementia from two National Health Service (NHS) Foundation Trusts in the UK were used for model training/internal validation and external validation, respectively. A personalised prescription model based on the Recurrent Neural Network machine learning architecture was developed to predict the Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) scores post-drug initiation. The drug that resulted in the smallest decline in cognitive scores between prescription and the next visit was selected as the treatment of choice. Change of cognitive scores up to 2 years after treatment initiation was compared for model evaluation.

Results: Overall, 1343 patients with MMSE scores were identified for internal validation and 285 [21.22%] took the drug recommended. After 2 years, the reduction of mean [standard deviation] MMSE score in this group was significantly smaller than the remaining 1058 [78.78%] patients (0.60 [0.26] vs 2.80 [0.28]; $P = 0.02$). In the external validation cohort ($N = 1772$), 222 [12.53%] patients took the drug recommended and reported a smaller MMSE reduction compared to the 1550 [87.47%] patients who did not (1.01 [0.49] vs 4.23 [0.60]; $P = 0.01$). A similar performance gap was seen when testing the model on patients prescribed with AChEIs only.

Conclusions: It was possible to identify the most effective drug for the real-world treatment of cognitive impairment in dementia at an individual patient level. Routine care patients whose prescribed medications were the best fit according to the model had better cognitive performance after 2 years.

Keywords: Dementia, Personalised treatment, Acetylcholinesterase inhibitor, Memantine, Artificial intelligence

* Correspondence: qiang.liu@psych.ox.ac.uk

¹Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK

Full list of author information is available at the end of the article



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Dementia is an age-related neurodegenerative syndrome [1], currently affecting approximately 55 million people worldwide, with just under 10 million new cases every year [2]. Medications approved for the treatment of dementia in the US, UK and Canada are the acetylcholinesterase inhibitors (AChEIs; donepezil, rivastigmine and galantamine) and memantine [3]. The use of these medications has been indicated in several guidelines across these countries [4–6]. Generally, AChEIs are recommended for patients with mild to moderate dementia and memantine for patients with moderate to severe dementia [4]. The guidelines are informed by double-blind placebo-controlled randomised trials showing that decline in cognitive performance stabilises for a period of 3–6 months after treatment initiation and have been replicated in a large real-world observational study [7]. However, previous studies that assessed the efficacy of these medications on dementia focused on the average effect of these interventions and did not investigate the potential differences at the level of individual patients [8]. There is a strong movement in clinical research advocating for a more personalised approach in medicine, using more advanced analytical approaches so that non-linear relationships among multiple factors can be explored together [9].

Recently, artificial intelligence (AI) has been widely applied to dementia research. Deep neural network, one of the most sophisticated machine learning approaches, is commonly used with neuroimaging [10–12] and genetic data [13] as it can make predictions by discovering generalisable nonlinear latent patterns [14, 15] and detect early onset of dementia. Other clinical parameters, such as abnormal alterations in drawing [16], gait [17], and speech [18], have also been informed by AI to effectively monitor brain health and disease progression. In terms of treatment for dementia, AI-driven technologies, such as assistive robots [19] and smart sensors [20], have been developed, but mainly to provide caregiving and management support [21]. Some studies have investigated how to individualise non-pharmacological interventions, e.g. physical exercise recommendation [22], tailored interactive reminiscence session [23], etc., but little to no research has been carried out so far about the personalisation of pharmacological treatment in patients with dementia.

In this study, we aimed to test whether AI-based recommendations based on patient-level information can identify which is the best treatment for each patient and improve their clinical outcome. We tested the specific effect of the 4 recommended drugs for dementia (the three AChEIs and memantine), utilising a large observational dataset of real-world patients. This allowed us to investigate the differential effects of medications by

combining demographic data and longitudinal cognitive measures and building a cognitive score prediction model based on deep neural networks to identify the most effective drug for cognitive impairment in dementia at the individual patient level.

Methods

Study design and patients

We used de-identified electronic health record (EHR) datasets collected in two UK National Health Service (NHS) Foundation Trusts, namely Oxford Health NHS Foundation Trust (OHFT) and Southern Health NHS Foundation Trust (SHFT). The data are held on the UK CRIS platform [24]. All data used in this study were delivered via the Akriya Health (<https://akriyahealth.com/>, Oxford, UK) Data Research Platform under a service agreement between the NHS and Akriya Health. We selected patients aged 59–102 years with a diagnosis of dementia through both structured International Classification of Disease 10th revision (ICD-10) codes and mentions of dementia diagnosis in the clinical notes by 10 June 2019. The AI model was developed and internally validated on the data from SHFT, while OHFT data were used to externally validate the performance of the model. We identified 3905 individuals from SHFT and another 2899 individuals from OHFT for external validation. The SHFT and OHFT cohorts contributed to a total number of 12,905 and 5296 observations, respectively. In Table 1, we reported the descriptive statistics of cognitive tests in the two cohorts using the Mini-Mental State Examination (MMSE) [25] and the Montreal Cognitive Assessment (MoCA) [26].

We assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human subjects/patients were approved by the local CRIS oversight committees and are covered by approval for the CRIS database granted by the Oxfordshire and Southern Health Research Ethics Committee. Individual patient consent is not required for this use of anonymised, routine data.

Sample preparation

Data in UK CRIS consists of both structured fields, such as clinical diagnoses represented by ICD-10 codes, demographic factors and unstructured fields, i.e. clinical notes [27]. We first used a natural language processing (NLP) algorithm [28, 29] to decode the clinical notes into structured data including date of visit, diagnosis, medication and cognitive scale scores such as MMSE and MoCA. Here, we only used NLP to extract cognitive scores stated in the clinical notes as opposed to somehow reconstructing proxy scores. Rule-based NLP was

Table 1 Descriptive statistics of patients from Oxford Health NHS Foundation Trust (OHFT) and Southern Health NHS Foundation Trust (SHFT) based on the time when they received medication for the first time. The data for each Trust is presented separately according to the cognitive scale reported (Mini-Mental State Examination (MMSE) or Montreal Cognitive Assessment (MoCA))

	OHFT (MMSE)	OHFT (MoCA)	SHFT (MMSE)	SHFT (MoCA)
Number of patients	1772	1127	3358	547
Number of observations	3575	1721	12,007	898
Male	37.81%	42.41%	39.70%	42.96%
Age, mean [SD]	81.93 [6.81]	81.72 [6.52]	80.48 [7.61]	80.22 [7.76]
Score, mean [SD]	21.01 [5.11]	17.25 [5.39]	21.10 [5.41]	17.95 [5.51]
Ethnicity (white)	81.66%	55.46%	79.51%	76.05%
Ethnicity (Asian)	0.28%	6.21%	0.33%	0.18%
Ethnicity (black)	0.45%	4.44%	0.21%	0.18%
Ethnicity (others)	0.90%	6.21%	0.27%	NA
Ethnicity (not available)	16.70%	42.86%	19.68%	23.58%
Marital status (married)	40.86%	26.97%	44.16%	40.04%
Marital status (separated)	2.26%	1.77%	4.52%	3.47%
Marital status (single)	4.01%	1.24%	3.21%	3.47%
Marital status (widowed)	28.05%	17.13%	26.18%	21.02%
Marital status (not available)	24.83%	52.88%	21.92%	31.99%

used because we aimed to identify and structure information that often follows certain linguistic patterns in clinical notes. In this situation, defining rules to extract medication prescriptions and cognitive scores was parsimonious in comparison to using statistical learning NER approaches. Moreover, it is more interpretable to clinicians [30]. The macro average F1 score of NLP extractions was 91.21%. Details can be found in Additional file 1: Tables S1 and S2. We then selected the patients who have either a diagnosis of dementia or mentions of dementia in the clinical notes. The selection criteria were further narrowed down to those that had either MMSE or MoCA score recorded. We kept clinical details of all selected patients from their very first observations until the last observations for our longitudinal analysis. The interval for subsequent visits was kept up to 2 years from the initial medication prescription. We excluded patients taking more than one anti-dementia drugs. The patients contributing to only one observation were also excluded since no follow-up data could identify their outcomes. We finally complemented the information of patients with structured demographic factors.

Predictors and outcomes

Two psychiatrists pre-defined a set of 7 variables from UK CRIS as predictors: age, sex, ethnicity, marital status, duration (time between two continuous observations), cognitive scores (MMSE or MoCA) and prescribed medication, which was considered as a minimal collection for building a personalised prescription model. Age [31] and cognitive score [7] were selected because,

among a wide range of clinical and demographic variables, they were found to be independently associated with progression to dementia [32]. Duration can be directly linked to cognitive decline [33], while sex [34], ethnicity [35] and marital status [36] were associated with dementia progression in multiple studies [37–39]. Since we aimed to investigate the specific response to different drugs, prescribed medication was also included as a predictor.

The primary outcome that our model predicted was the cognitive scale score post-drug initiation. MMSE and MoCA were adopted as they are the standard and most often used screening tools for an overall measure of cognitive impairment in both clinical and research settings [40, 41]. Studies have concluded they are sufficiently accurate to detect and monitor cognitive impairment [42, 43]. For both scales, a lower score indicates a more severe cognitive impairment. The treatment that resulted in the smallest decline in cognitive scores between prescription and the next visit was selected as the neural network treatment of choice (NNTToC).

Data analysis

To develop a model capable of identifying best responders given patient's clinical and demographic information, we built our personalised prescription model using a Recurrent Neural Network (RNN) machine learning architecture (Long Short Term Memory (LSTM) in particular), which has been widely applied in various fields [44–46]. The choice was based on a previous research study [47], where authors showed the

LSTM outperformed other traditional models when handling longitudinal EHR data. (We have also conducted analyses using ridge regression, random forest and one-dimensional Convolutional Neural Network, which are detailed in Additional file 1: Fig. S6, Tables S9 and S10.) The core of our model is a two-layer LSTM neural network. We used patients' cognitive scale change to evaluate the performance of the model at an individual level as this is the outcome reported by the previous studies reporting on the clinical response to these drugs [33, 48, 49].

The front-end stack of fully connected layers can be viewed as a feature extractor across the clinical and demographic information, similar to the principal component analysis, which identifies hidden correlations and patterns and summarises them into a feature vector. The rear-end RNN serves as a cognitive score predictor that captures temporal trends using encoded feature vectors and generates an estimated cognitive score for the four available drugs. Full details of the LSTM model and data pre-processing are reported in Additional file 1: Table S3, Fig. S1 and S2.

We carried out our main analysis using both MMSE and MoCA scores. For the subsequent subgroup analyses, we focused only on MMSE (5130 patients accounting for 15,582 observations from both sites) due to there being considerably fewer instances of MoCA scores (1674 patients with 2619 observations from both sites). The clinical data from SHFT was used for model development and internal validation, while the OHFT data was used for external validation.

Analysis using MMSE score

In the first step of the analysis, we identified 3358 patients from SHFT, all of whom had MMSE scores. We randomly selected 2015 [60.01%] individuals for model training and kept the rest for internal validation. For each patient, we randomly selected 4 continuous observations. We did not impute missing values. Instead, missing values were considered as categorical values and incorporated into our model to increase stability and robustness. If a patient had fewer than 4 observations, zero pre-padding (a common process in RNN where zeros are added at the start of the sequence if the length of a sample is shorter than the given standard length) was deployed for null observations. In our case, if a patient had only 1 observation, we padded the first 3 observations with zeros.

In the second step, we predicted the values of cognitive scores for all four available medications. This way, our model produced a contrafactual scenario, whereby we predicted an MMSE score for each medication. We then sorted the predicted scores and selected the medication that generated the highest score to derive the

NNTtoC. Consequently, the patients could be separated into two groups depending on whether they were prescribed NNTtoC or not. Finally, we evaluated our model and plotted the average change of MMSE given medication prescription across the time of disease. Due to the similar profile of AChEIs, we also conducted analysis where we excluded patients prescribed with memantine and tested our model using AChEIs only.

Analysis using MoCA score

Next, we tested our model on patient records with MoCA scores. However, MoCA scores were highly under-reported in SHFT. In this case, we retrained the LSTM on all patients' data from SHFT and externally validated it on OHFT. On average, each patient contributed to 1.64 observations; thus, we selected 2 continuous observations from each patient for model training.

Neural network performance analysis

Additional analyses were carried out to further explore the model robustness and evaluate the recommendation patterns suggested by the neural network. Firstly, we examined the model performance on a random observation rather than focusing on the cognitive score at treatment initiation. Secondly, we tried to address the question of whether the neural network was providing personalised recommendations, or whether it was instead giving better recommendations that generally worked well. To do this, for each patient, we assigned NNTtoC using the advice for another randomly selected patient. Thirdly, we compared the LSTM model with several other models, namely a more interpretable ridge regression, a traditional random forest machine learning model and a one-dimensional CNN. We pre-processed the data using the same method in our main analysis. For ridge regression and random forest, we flatten the longitudinal features before training [47]. Grid search was used to find the best combination of hyperparameters for every model. Fourthly, due to the imbalance in prescribed medications in the real world, we oversampled the samples in the minority classes by randomly duplicating patients taking minority drugs until each minority class have the same number of patients (patients taking donepezil). We then retrained the model and evaluated the performance. Fifthly, we tried to retrain the model using multitask learning [50], where we fed both MMSE and MoCA scales into one single model. The model then yielded both scores as outputs. Since no patient had both MMSE and MoCA tested on the same visit, i.e. one label was always missing, we included another label indicating which score was missing and only used the non-missing scores to calculate the model loss. Sixthly, instead of using 4 as the sequence length, we retrained the model using only 3 or 2 observations per

patient to check the impact of a shorter sequence length. Finally, we adopted permutation feature importance to measure the importance of predictors [51]. Permutation feature importance measures the decrease in a model's performance when a single predictor is randomly shuffled. Implementations of the model performance analyses are detailed in Additional file 1: Tables S9, S11, S13 and S15.

Data cleaning and data preparation were carried out in R [52]. The neural network model was trained in TensorFlow [53]. The validation and other analyses were carried out using Python [54].

Results

Our model predicted cognitive scores post-drug initiation for each drug. The treatment that resulted in the smallest decline in cognitive scores between prescription and the next visit was selected as the NNToC. Note that every patient in the study received one of the four possible drugs approved for cognitive symptoms in dementia. Thus, the patients could be separated into two groups depending on whether they were prescribed NNToC or not. We then use patients' mean cognitive score decline to evaluate the performance of our model.

Internal validation using MMSE score

The rest 1343 [39.99%] patients from SHFT, who were not randomly selected for model training, were used for internal validation. Overall, 285 [21.22%] patients were

prescribed NNToC, i.e. they received the medication that generated the highest predicted MMSE scores for their individual case, and 1058 [78.78%] did not get NNToC. We finally plotted the change of MMSE scores in relation to time, as shown in Fig. 1A. The cognitive performance of patients who were not prescribed NNToC declined significantly. In contrast, the patients receiving NNToC declined at a slower rate. Specifically, over the 2 years, their mean [SD] MMSE score declined 0.60 [0.26] points, compared to 2.80 [0.28] who did not receive NNToC (a mean difference of 2.20), which is statistically significant ($P = 0.02$).

External validation using MMSE score

In terms of external validation, we used the model to evaluate the medication prescription patterns of the OHFT data. We identified 1772 patients with MMSE scores, among which 222 [12.53%] were prescribed NNToC and 1550 [87.47%] were not. The demographics of patients are summarised in Table 2. No significant difference was found between these two cohorts in terms of age, MMSE score and ethnicity; however, there were fewer male ($P = 0.003$) and married patients ($P = 0.04$) who were prescribed NNToC. We also reported the ratio of medications among patients who received NNToC in Additional file 1: Table S4 to show the model was not always prescribing one drug that generally worked well, e.g. memantine.

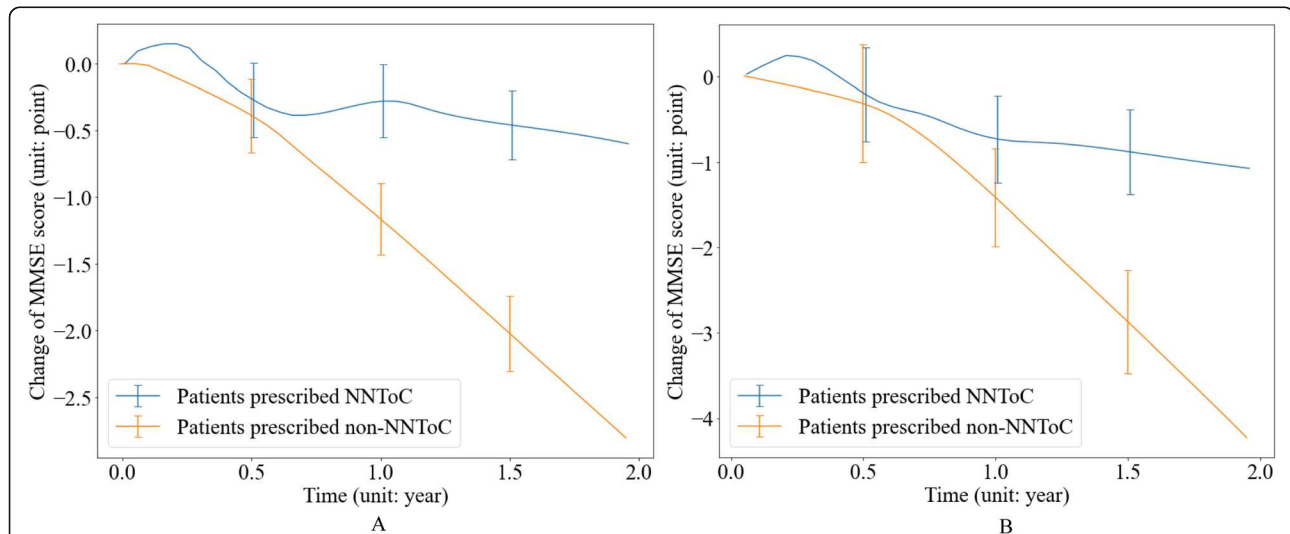


Fig. 1 Change of Mini-Mental State Examination (MMSE) score over time. Patients were grouped by whether they were prescribed neural network treatment of choice (NNToC) according to the Long Short Term Memory (LSTM) model. The X-axis represents the duration of taking a particular medication, where $x=0$ means the treatment initiation time. Data shown are mean values, with error bars indicating standard deviation. **A** Internal validation on Southern Health NHS Foundation Trust (SHFT). Two hundred eighty-five [21.22%] patients were prescribed NNToC and reported a smaller MMSE reduction after 2 years compared to the 1058 [78.78%] patients who were not (0.60 [0.26] vs 2.80 [0.28], respectively; $P = 0.02$). **B** External validation on Oxford Health NHS Foundation Trust (OHFT). Two hundred twenty-two [12.53%] patients were prescribed NNToC and reported a smaller MMSE reduction after 2 years compared to the 1550 [87.47%] patients who were not (1.01 [0.49] vs 4.23 [0.60], respectively; $P = 0.01$). A detailed quantitative report of score changes is provided in Additional file 1: Table S5

Table 2 Demographics of patients who (i) were prescribed neural network treatment of choice (NNToC) and (ii) were not prescribed NNToC and significance and external validation on Oxford Health NHS Foundation Trust (OHFT) at treatment initiation. SD standard deviation, MMSE Mini-Mental State Examination, χ^2 chi-square test, *t* Welch's *t*-test; *P* *P*-value

	Patients prescribed NNToC (<i>n</i> = 222, 12.53%)	Patients prescribed non-NNToC (<i>n</i> = 1550, 87.47%)	Significance
Gender (%male)	28.38%	39.10%	$\chi^2(1) = 9.15, P = 0.003$
Age, mean [SD]	81.12 [7.01]	82.05 [6.73]	$t(282.50) = -1.86$ $P = 0.06$
MMSE score, mean [SD]	20.84 [4.89]	21.06 [5.20]	$t(297.27) = -0.62$ $P = 0.53$
Ethnicity (%Caucasian)	81.08%	81.74%	$\chi^2(1) = 0.02, P = 0.88$
Marital status (%married)	34.23%	41.81%	$\chi^2(1) = 4.30, P = 0.04$

The change of MMSE score over time is shown in Fig. 1B. We found the trajectories of the cognitive performance of patients from OHFT and SHFT were similar. At the point of initial prescription, all cognitive trajectories had a period of stabilisation lasting approximately 6 months, which indicated treatment effect. After 6 months, the results show that the cognitive performance of patients who were not prescribed NNToC started to decline significantly. The patients not prescribed NNToC generally declined more than the ones taking NNToC in terms of cognitive performance. Specifically, the MMSE score declines of patients taking NNToC and non-NNToC were 0.70 [0.52] and 1.27 [0.59] after 1 year of treatment initiation ($P = 0.28$) and were 1.01 [0.49] and 4.23 [0.60] (a mean difference of 3.22) after 2 years of treatment initiation ($P = 0.01$). A detailed quantified report of MMSE scores on OHFT at selected time points is presented in Additional file 1: Table S5.

For further validation, we excluded patients prescribed with memantine and tested our model on AChEIs only due to the similar profile of these drugs. In this case, the NNToC only include donepezil, rivastigmine and galantamine. We identified 1660 patients who were prescribed AChEIs among which 212 [12.77%] were prescribed NNToC and 1448 [87.23%] were not. The MMSE score declines of patients taking NNToC and non-NNToC were 1.11 [0.47] and 4.29 [0.60] (a mean difference of 3.18) after 2 years of treatment initiation ($P = 0.02$). The MMSE score change and detailed quantified report are presented in Additional file 1: Fig. S3 and Table S6.

External validation using MoCA score

We identified 547 patients with MoCA scores in SHFT for model training and externally validated our model on 1127 patients from OHFT on the observations when a medication was prescribed for the first time, among which 160 [14.20%] were prescribed NNToC and 967 [85.80%] were not. The cognitive performance represented by MoCA scores is shown in Fig. 2. The MoCA score declines of patients who

were and were not prescribed NNToC were 1.03 [0.19] and 1.56 [0.21] (a mean difference of 0.53) at 1.75 years ($P = 0.09$).

Neural network performance analysis

Firstly, the model was tested on a random observation rather than the observation at treatment initiation, as shown in Additional file 1: Fig. S4 and Table S7. After 2 years of treatment initiation, the MMSE score declines of patients who did and did not receive NNToC were 1.92 [0.41] and 4.55 [0.49] (a mean difference of 2.63), respectively ($P = 0.04$). Secondly, we assigned NNToC using the advice for another randomly selected patient. We then recalculated the MMSE scores and the cognitive performance result is reported in Additional file 1: Fig. S5 and Table S8. We observed that the MMSE score change of patients following NNToC dropped significantly 3.26 [0.65] points at 2 years. Thirdly, performance comparison with other models is shown in Additional file 1: Fig. S6 and Table S10. At 2 years, all other models performed similarly, but all worse than the LSTM. However, in the first 6 months, patients who were prescribed NNToC based on the ridge regression had a higher MMSE drop than patients who were not. Fourthly, when the model was trained on the oversampled data, the performance remained almost the same (Additional file 1: Fig. S7 and Table S12). Fifthly, if multitask learning was applied (Additional file 1: Fig. S8 and Table S14), the performance of the model became slightly better when evaluated on MoCA (patients prescribed NNToC had a smaller drop of 0.93 [0.21] points), but worse when evaluated on MMSE (patients prescribed NNToC had a bigger drop of 1.98 [0.70] points). Sixthly, as the sequence length decreased, the difference between the NNToC and non-NNToC groups became less sound at 2 years (Additional file 1: Fig. S9 and Table S16). Particularly, when the model was trained on only 2 observations per patient, the MMSE score of patients who were prescribed NNToC dropped 2.08 [0.65] points, but still smaller than patients who were not at 3.03 [0.61] points ($P = 0.06$).

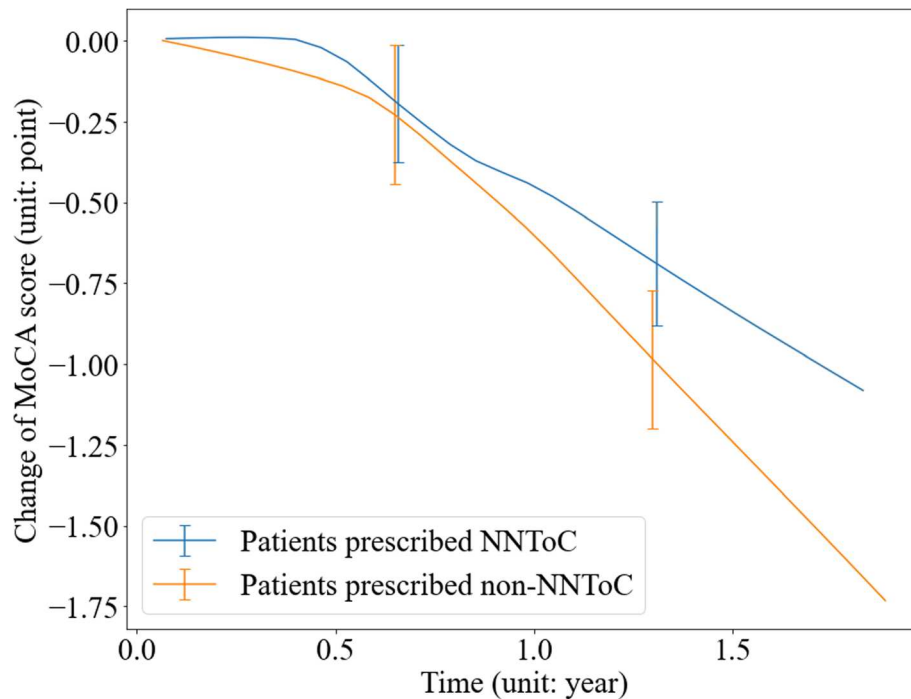


Fig. 2 Change of Montreal Cognitive Assessment (MoCA) score over time. Patients were grouped by whether they were prescribed neural network treatment of choice (NNTToC) according to the Long Short Term Memory (LSTM) model. The X-axis represents the duration of taking a particular medication, where $x = 0$ means the treatment initiation time. Data shown are mean values, with error bars indicating standard deviation. External validation on Oxford Health NHS Foundation Trust (OHFT). One hundred sixty [14.20%] patients were prescribed NNTToC and reported a smaller MoCA reduction after 1.75 years compared to the 967 [85.80%] patients who were not (1.03 [0.19] vs 1.56 [0.21], respectively; $P = 0.09$). A detailed quantitative report of score changes is provided in Additional file 1: Table S5

Finally, we present the permutation feature importance analysis of the LSTM model in Additional file 1: Fig. S10 and Table S17. For the MMSE model, the cognitive score was the absolute dominant predictor (165.52% increase in MAE, 3 times more important than the second important predictor), followed by duration, medication (around 60% increase) and age (36.65% increase). Sex and marital status were much less important (around 5% increase) and ethnicity was not an important predictor (0.10% increase). In terms of the MoCA model, we saw a similar ranking and pattern, but medication was slightly more important than duration.

Discussion

This paper expands the existing knowledge base on dementia treatment and demonstrates that some individual-to-individual variability exists regarding how patients react to each particular agent used, allowing the identification of individualised treatments that can prolong the positive effects of medication prescription beyond the first 6 months. The model caught the recommendations that were more effective beyond this initial period and the weaker cognitive decline provided by NNTToC became more significant later in the disease. Patients prescribed NNTToC showed a significantly

slower rate of decline and outperformed controls (i.e. patients prescribed a medication different from NNTToC) in terms of cognitive performance after 2 years of treatment initiation. In the internal validation on the MMSE cognitive score, the decline was slower by a magnitude of 5 for NNTToC relative to the controls. In the external validation sample, we saw a similar, fourfold slower rate of decline in the NNTToC group relative to controls. These results demonstrate the value offered by AI-based recommendations in terms of personalised prescription. Currently, the pharmacological treatment with AChEIs is prescribed mainly on the custom of each particular physician [4–6], as it was assumed by default that all AChEIs would have comparable results. In this situation, an informed personalised prescription of AChEIs only has the potential to be more effective. The only approved cognitive drug whose prescription is guided by the particulars of each patient is memantine, given it is most effective in moderate to severe dementia [55], and may better manage behavioural problems [56]. However, our results also show that dementia-based prescription continues having the same overall effect across patients when memantine is taken into account.

Our results about non-personalised prescriptions are consistent with previous studies that after the initial

positive effect (typically 3–6 months in duration), patients continue to decline in their cognitive performance [33, 48, 49, 57, 58]. This is important as there is still a debate about whether or not to continue these drugs over a long-term period, considering the marginal cognitive benefit afforded by them when prescribed in a non-personalised manner (0.91 MMSE points at 6 months for Alzheimer's disease/vascular dementia patients) [59]. These concerns have induced some regulators, e.g. in France, to defund their prescription [60]. We hope that a personalised approach to the prescription of medications in dementia may lead to a proper re-evaluation of the cost-benefit ratio for these drugs.

The neural network was able to generate recommendations even when patients had already been on at least one other anti-dementia drug according to the neural network performance analysis. Notably, the MoCA-based external validation model pointed to a longer stabilisation period of roughly 8 months and a gentler cognitive decline relative to the MMSE-based dataset. This might be because MoCA is capturing earlier cognitive decline through its larger reliance on executive function cognitive domains and it may be less sensitive than MMSE in capturing posterior impairment [61].

In terms of the pattern of recommendations made by the model (ratio of medications), we found that the NNToC highly corresponded to the current guidelines on the prescription of these drugs [4–6]. AChEIs were generally selected as NNToC for patients with mild to moderate severity, whereas memantine was suggested for patients with moderate to severe disease. In the UK, the NICE guidelines do not differentiate AChEIs and only suggest treatment to be started with the drug with the lowest acquisition cost if prescribing an AChEI, which consequently makes donepezil being prescribed more than the other drugs in the real world. However, interestingly, the neural network further avoided suggesting donepezil to men as well as those scoring low on MMSE and tended to instead recommend galantamine, rivastigmine and memantine for these patients. Across disease severity, galantamine, rivastigmine and memantine tended to be NNToC more often. These findings warrant further analysis and validation, with the aim of further refining the current guidelines.

Some potentially important differences between the two clinical cohorts were evident. The patients from the SHFT sample had a lower degree of cognitive decline compared to the OHFT group. One possible explanation is that only 24.38% (432/1772) patients from OHFT were re-examined after a period of 5 months, whereas this proportion for SHFT was 44.58% (1497/3358). According to our previous study [7], patients who were always on one drug benefitted more than patients switching at least once to a different drug. Thus, patients from OHFT

may have switched to alternative medications in a timely manner on a follow-up visit, whereas patients from SHFT were able to remain on an agent for a longer time. Another reason might be that patients from OHFT were on average 1.45 years older than those from SHFT. Previous reports have indicated that older age is associated with a faster decline and therefore could account for the larger MMSE score decline in OHFT patients [62, 63].

By comparing the decline of patient cognitive performance in the NNToC and non-NNToC groups, the biggest difference is seen by using the LSTM model trained with 4 observations. Compared to other models, the LSTM explored and captured a higher degree of nonlinearity both in the spatial (predictors) and temporal (observations) space. Although our explanatory analysis using permutation feature shows medication is a strong predictor for the LSTM model (more than 50% increase in MAE), in this case as a general rule for feature importance analysis, it has limited ability to identify the predictor that is most decisive as the predictors cannot be explained individually. It is also worth noting that permutation feature importance only provides a reliable interpretation of the specific model fitted on the training data, but limited insights into the data itself. Therefore, caution is needed when making clinical decisions based on these feature importance results.

A limitation of the study is the defined set of predictors used to build the recommender system. Current predictors include selected demographic variables, cognitive scale scores and medication information that were available through the UK CRIS system. Despite ethnicity being included as a predictor, the distribution was highly skewed towards white-Caucasian (close to 80% in the MMSE samples). Thus, the study may lack the power to detect the potential effect of ethnicity. Although the proposed model showed good performance, the inclusion of a more comprehensive set of predictors (clinical, diagnostic, genetic, omics information) may further improve the capabilities of personalised medication prescription. Neuroimaging data also can contribute volumetric or functional data that may further improve the accuracy of the predictions [64]. Multimodal dataset linkage and big data harmonisation may also offer novel ways to optimise the symptomatic treatment of dementia [65].

Furthermore, it should be noted that widely used assessments, such as MMSE and MoCA, are not necessarily the most accurate. In some studies, MMSE has been shown to be a weaker predictor of dementia severity as compared to other cognitive and functional assessments [66, 67]. Some research suggested that a battery of assessments would be preferable instead of a stand-alone single-administration assessment [68, 69] and this approach may improve the accuracy of our model in the future.

On the other hand, it is known that patients respond differently to drugs depending on the aetiology of the cognitive impairment and the type of dementia, e.g. there is no evidence for efficacy in vascular dementia [70]. In our study, we did not conduct subgroup analyses as 71.74% of the included population were diagnosed with Alzheimer's disease and the other subtypes of dementia were not large enough to draw meaningful conclusion: only 4.68% had vascular dementia, 4.02% Parkinson's disease dementia and 0.62% dementia in other diseases (the remaining 18.94% had a diagnosis of unspecified dementia). As a potential future development, we will aim to expand the set of predictors and use individual patient data from randomised control trials combined with observational data to enrich the neural network further. In addition, an online version of the model to be used as a clinical decision support making tool is currently under development.

Finally, in our data, the clinician's choice of medication was not randomised and may have been influenced by comorbidities and other unknown factors [71]. The conclusion on the effect of a drug in a given person could be compromised by reverse causation, i.e. a particular drug was prescribed more often in subjects that had a priori a worse or better prognosis. We plan to incorporate such factors as predictors to enrich our model in the future. A randomised controlled trial will also be required to test prospectively the utility of the proposed model to establish its real-world efficacy.

Despite the limitations, the findings from external validation have proved the significance of our work and revealed the potential of deep neural networks in the medical science domain. To the best of our knowledge, this is the first personalised prescription model for the pharmacological treatment of cognitive deficits in dementia. This finding suggests the proposed model, and AI-based prescription recommendation in general has the potential to benefit clinical outcomes and can be used by clinical staff as a clinical decision support tool.

Conclusions

In conclusion, AI-based recommendations produced personalised treatment in dementia, and it was possible to identify at the individual patient level the most effective drug to reduce cognitive impairment over two years. Real-world patients whose prescribed medications were the best fit according to our model had better cognitive performance after 2 years. These results should be replicated in longer-term prospective studies.

Abbreviations

AChEI: Acetylcholinesterase inhibitor; AI: Artificial intelligence; CNN: Convolutional neural network; HER: Electronic health record; ICD-10: International Classification of Disease 10th revision; LSTM: Long Short Term Memory; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; NHS: National Health Service; NLP: Natural language processing; NNToc: Neural network treatment of choice; OHFT: Oxford

Health NHS Foundation Trust; RNN: Recurrent Neural Network; SD: Standard deviation; SHFT: Southern Health NHS Foundation Trust

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-022-02250-2>.

Additional file 1: Figure S1. An overview of the Long Short Term Memory (LSTM) model. **Figure S2.** Training loss of the Mini Mental State Examination (MMSE) Long Short Term Memory (LSTM) model. **Figure S3.** Change of Mini Mental State Examination (MMSE) score over time when the recommendation drugs were 3 acetylcholinesterase inhibitors (AChEIs; donepezil, rivastigmine and galantamine) only. **Figure S4.** Change of Mini Mental State Examination (MMSE) score over time when a medication was prescribed on a randomly selected visit. **Figure S5.** Change of Mini Mental State Examination (MMSE) score over time when recommendations were randomly shuffled. **Figure S6.** Change of Mini Mental State Examination (MMSE) score over time when ridge regression, random forest and one-dimensional Convolutional Neural Network (1D CNN) were used for drug recommendation, compared to the Long Short Term Memory (LSTM) model. **Figure S7.** Change of Mini Mental State Examination (MMSE) score over time when the Long Short Term Memory (LSTM) model was trained on the oversampled data. **Figure S8.** Change of Mini Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) score over time when recommendations were given by a single Long Short Term Memory (LSTM) model developed using multitask learning. **Figure S9.** Change of Mini Mental State Examination (MMSE) score over time when the Long Short Term Memory (LSTM) model was trained with fewer observations per patient. **Figure S10.** Permutation feature importance of the Long Short Term Memory (LSTM) model. **Table S1.** Validation of the natural language processing (NLP) model performance on the UK Clinical Record Interactive Search (CRIS) data. **Table S2.** Validation of the natural language processing (NLP) performance on additional categories. **Table S3.** Implementation details of the Long Short Term Memory (LSTM) model. **Table S4.** The ratio of prescribed medications. **Table S5.** Quantified Mini Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) score changes when Long Short Term Memory (LSTM) model was used for drug recommendation. **Table S6.** Quantified Mini Mental State Examination (MMSE) score changes when the recommendation drugs were 3 acetylcholinesterase inhibitors (AChEIs; donepezil, rivastigmine and galantamine) only. **Table S7.** Quantified Mini Mental State Examination (MMSE) score changes when a medication was prescribed on a randomly selected visit. **Table S8.** Quantified Mini Mental State Examination (MMSE) score changes when recommendations were randomly shuffled. **Table S9.** Implementation details of the ridge regression, random forest and one-dimensional Convolutional Neural Network (1D CNN). **Table S10.** Quantified Mini Mental State Examination (MMSE) score changes when ridge regression, random forest and one-dimensional Convolutional Neural Network (1D CNN) were used for drug recommendation. **Table S11.** Implementation details of the Long Short Term Memory (LSTM) model for the oversampled data. **Table S12.** Quantified Mini Mental State Examination (MMSE) score changes when the Long Short Term Memory (LSTM) model was trained on the oversampled data. **Table S13.** Implementation details of the Long Short Term Memory (LSTM) model for multitask learning. **Table S14.** Quantified Mini Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) score changes when recommendations were given by a single Long Short Term Memory (LSTM) model developed using multitask learning. **Table S15.** Implementation details of the Long Short Term Memory (LSTM) model when the model was trained with fewer observations per patient. **Table S16.** Quantified Mini Mental State Examination (MMSE) score changes when the Long Short Term Memory (LSTM) model was trained with fewer observations per patient. **Table S17.** Permutation feature importance scales of the Long Short Term Memory (LSTM) model.

Acknowledgements

We would like to acknowledge the support from the Oxford CRIS Team, Tanya Smith (Head of Research Informatics), Adam Pili (CRIS Academic Support and Information Analyst), Suzanne Fisher (CRIS Academic Support

and Information Analyst) and Lulu Kane (CRIS Administrator). We also acknowledge and are grateful for the support from the Southern Health NHS Foundation Trust team, Dr Peter Phiri (Research & Development Manager/CRIS Academic Lead) and Maria Lane (Research facilitator and CRIS coordinator). Finally, our thanks go to Emma Cable from the Akrivia Health team. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Authors' contributions

QL was responsible for the machine learning analysis. NV provided statistical contributions. QL, NV and AK were responsible for data acquisition. QL, NV and ZL carried out data cleaning. QL, NV, IK and ANH were responsible for study conception. IK and AC provided interpretation. ANH supervised the project. All authors contributed to the manuscript. The authors read and approved the final manuscript.

Funding

The study was funded by the National Institute for Health Research's (NIHR) Oxford Health Biomedical Research Centre (BRC-1215-20005), by the MRC Pathfinder Grant (MC-PC-17215) and EPSRC-NIHR HTC Partnership Award 'Plus': NewMind - Partnership with the MindTech HTC (EP/N026977/1). The work was technically supported by the UK Clinical Record Interactive Search (UK CRIS) platform using data and systems of the NIHR Oxford Health Biomedical Research Centre (BRC-1215-20005). Ivan Koychev declares support for the work on this project through NIHR Oxford Health Biomedical Research Centre and MRC Dementias Platform UK Project (MR/L023784/2). Andrea Cipriani is supported by the NIHR Oxford Cognitive Health Clinical Research Facility, by an NIHR Research Professorship (grant RP-2017-08-ST2-006) and by the NIHR Oxford and Thames Valley Applied Research Collaboration.

Availability of data and materials

The data that support the findings of this study (textual documents and structured information) are available by application through the UK CRIS network (<https://crisnetwork.co/>) via the Akrivia Health (<https://akriviahealth.com/>, Oxford, UK) Data Research Platform. Owing to the sensitivity of clinical information, access is dependent on receiving research approvals from the NHS trust oversight bodies at Oxford and Southern Health NHS Foundation Trusts.

Declarations

Ethics approval and consent to participate

We assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human subjects/patients were approved by the local CRIS oversight committee and are covered by approval for the CRIS database granted by the Oxfordshire and Southern Health Research Ethics Committee. Individual patient consent is not required for this use of anonymised, routine data.

Consent for publication

Individual patient consent is not required for this use of anonymised, routine data.

Competing interests

Ivan Koychev is a paid medical advisor for Cognitivity, Mantrah and Sharp Therapeutics companies which specialise in developing digital solutions for monitoring cognition or supporting independence in patients. Andrea Cipriani has received research and consultancy fees from INCiPIT (Italian Network for Paediatric Trials), CARIPLO Foundation and Angelini Pharma, outside the submitted work. Alejo Nevado-Holgado works as a consultant in Akrivia Health (<https://akriviahealth.com/>) and has obtained funding from Johnson & Johnson, Ono Pharma and GSK. None of those funds is involved in this publication. The other authors declare no conflicts of interest.

Author details

¹Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK. ²Department of Psychology, University of Sheffield, Sheffield, UK. ³Institute of Mathematics, University of Oxford, Oxford, UK. ⁴Oxford

Health NHS Foundation Trust, Warneford Hospital, Oxford, UK. ⁵Big Data Institute, University of Oxford, Oxford, UK. ⁶Akrivia Health, Oxford, UK.

Received: 11 August 2021 Accepted: 11 January 2022

Published online: 01 February 2022

References

- Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 2020;396(10248):413–46. [https://doi.org/10.1016/S0140-6736\(20\)30367-6](https://doi.org/10.1016/S0140-6736(20)30367-6).
- Dementia - World Health Organization 2 September 2021 [cited 2021 31 October]. Available from: <https://www.who.int/news-room/fact-sheets/detail/dementia>.
- Yunusa I, Alsahali S, Ranes A, Egualé T. Comparative value of cholinesterase inhibitors and memantine in persons with moderate-to-severe Alzheimer's disease in the United States: a cost-effectiveness analysis. *J Alzheimers Dis Rep*. 2021;5(1):705–13. <https://doi.org/10.3233/ADR-210307>.
- Rabins PV, Rovner BW, Rummans T, Schneider LS, Tariot PN. Guideline watch (October 2014): practice guideline for the treatment of patients with Alzheimer's disease and other dementias. *Focus (Am Psychiatr Publ)*. 2017; 15(1):110–28. <https://doi.org/10.1176/appi.focus.151106>.
- Pink J, O'Brien J, Robinson L, Longson D. Dementia: assessment, management and support: summary of updated NICE guidance. *BMJ*. 2018; 361:k2438. <https://doi.org/10.1136/bmj.k2438>.
- Moore A, Patterson C, Lee L, Vedel I, Bergman H. Fourth Canadian Consensus Conference on the Diagnosis and Treatment of Dementia: recommendations for family physicians. *Can Fam Physician*. 2014;60(5):433–8.
- Vaci N, Koychev I, Kim C-H, Kormilitzin A, Liu Q, Lucas C, et al. Real-world effectiveness, its predictors and onset of action of cholinesterase inhibitors and memantine in dementia: retrospective health record study. *Br J Psychiatry*. 2021;218(5):261–7. <https://doi.org/10.1192/bjp.2020.136>.
- Seibert M, Mühlbauer V, Holbrook J, Voigt-Radloff S, Brefka S, Dallmeier D, et al. Efficacy and safety of pharmacotherapy for Alzheimer's disease and for behavioural and psychological symptoms of dementia in older patients with moderate and severe functional impairments: a systematic review of controlled trials. *Alzheimers Res Ther*. 2021;13(1):1–20. <https://doi.org/10.1186/s13195-021-00867-8>.
- Wong-Lin K, McClean PL, McCombe N, Kaur D, Sanchez-Bornot JM, Gillespie P, et al. Shaping a data-driven era in dementia care pathway through computational neurology approaches. *BMC Med*. 2020;18(1):1–10. <https://doi.org/10.1186/s12916-020-01841-1>.
- Miah Y, Prima CNE, Seema SJ, Mahmud M, Kaiser MS. Performance comparison of machine learning techniques in identifying dementia from open access clinical datasets. *Adv Smart Soft Comput Springer*. 2021:79–89. https://doi.org/10.1007/978-981-15-6048-4_8.
- Sepehrband F, Barisano G, Sheikh-Bahaei N, Choupan J, Cabeen RP, Crawford MS, et al. Alteration of perivascular spaces in early cognitive decline: neuroimaging/optimal neuroimaging measures for early detection. *Alzheimers Dement*. 2020;16(S5):e045605. <https://doi.org/10.1002/alz.045605>.
- Forouzaneshad P, Abbaspour A, Li C, Fang C, Williams U, Cabrerizo M, et al. A Gaussian-based model for early detection of mild cognitive impairment using multimodal neuroimaging. *J Neurosci Methods*. 2020;333: 108544. <https://doi.org/10.1016/j.jneumeth.2019.108544>.
- Erdoğan O, Esmé M, Balci C, Rafatov S, Cankurtaran M, Yavuz BB, et al. Identification of genomic biomarkers with machine learning for early and differential diagnosis of late-onset Alzheimer's disease (LOAD). *Alzheimers Dement*. 2020;16(S2):e042558. <https://doi.org/10.1002/alz.042558>.
- Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities. *Br J Gen Pract*. 2018;68(668):143–4. <https://doi.org/10.3399/bjgp18X695213>.
- Kumar S, Oh I, Schindler S, Lai AM, Payne PR, Gupta A. Machine learning for modeling the progression of Alzheimer disease dementia using clinical data: a systematic literature review. *JAMIA Open*. 2021;4(3):ooab052.
- Amini S, Zhang L, Hao B, Gupta A, Song M, Karjadi C, et al. An artificial intelligence-assisted method for dementia detection using images from the clock drawing test. *J Alzheimers Dis*. 2021;(Preprint):1–9.
- Zhang Z, Jiang Y, Cao X, Yang X, Zhu C, Li Y, et al. Deep learning based gait analysis for contactless dementia detection system from video camera. *IEEE Int Symp Circuits Syst Proc*. 2021; IEEE.

18. Mahajan P, Baths V. Acoustic and language based deep learning approaches for Alzheimer's dementia detection from spontaneous speech. *Front Aging Neurosci.* 2021;13. <https://doi.org/10.3389/fnagi.2021.623607>.
19. Ozdemir D, Cibulka J, Stepankova O, Holmerova I. Design and implementation framework of social assistive robotics for people with dementia—a scoping review. *Health Technol (Berl).* 2021;1–12.
20. Khan AA. Robotics in dementia: caregiving needs addressed by an assistive robot in combination with wearable sensors. *Alzheimers Dement.* 2020;16(S11):e037855. <https://doi.org/10.1002/alz.037855>.
21. Astell AJ, Bouranis N, Hoey J, Lindauer A, Mihailidis A, Nugent C, et al. Technology and dementia: the future is now. *Dement Geriatr Cogn Disord.* 2019;47(3):131–9. <https://doi.org/10.1159/000497800>.
22. Müllers P, Taubert M, Müller NG. Physical exercise as personalized medicine for dementia prevention? *Front Physiol.* 2019;10:672. <https://doi.org/10.3389/fphys.2019.00672>.
23. Evans SC, Garabedian C, Bray J. 'Now he sings'. The My Musical Memories Reminiscence programme: personalised interactive reminiscence sessions for people living with dementia. *Dementia (London).* 2019;18(3):1181–1198. <https://doi.org/10.1177/1471301217710531>.
24. Denis M. U.K. clinical record interactive search (CRIS). *Alzheimers Dement.* 2017;13(7).
25. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12(3):189–98. [https://doi.org/10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6).
26. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* 2005;53(4):695–9. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>.
27. Goodday S, Kormilitzin A, Vaci N, Liu Q, Cipriani A, Smith T, et al. Maximizing the use of social and behavioural information from secondary care mental health electronic health records. *J Biomed Inform.* 2020;107:103429. <https://doi.org/10.1016/j.jbi.2020.103429>.
28. Senior M, Burghart M, Yu R, Kormilitzin A, Liu Q, Vaci N, et al. Identifying predictors of suicide in severe mental illness: a feasibility study of a clinical prediction rule (Oxford Mental Illness and Suicide Tool or OxMIS). *Front Psychiatry.* 2020;11:268. <https://doi.org/10.3389/fpsy.2020.00268>.
29. Vaci N, Liu Q, Kormilitzin A, De Crescenzo F, Kurtulmus A, Harvey J, et al. Natural language processing for structuring clinical text data on depression using UK-CRIS. *Evid Based Ment Health.* 2020;23(1):21–6. <https://doi.org/10.1136/ebmental-2019-300134>.
30. Gehrmann S, Démoncourt F, Li Y, Carlson ET, Wu JT, Welt J, et al. Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. *PLoS ONE.* 2018;13(2):e0192360. <https://doi.org/10.1371/journal.pone.0192360>.
31. Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C. Age, neuropathology, and dementia. *N Engl J Med.* 2009;360(22):2302–9. <https://doi.org/10.1056/NEJMoa0806142>.
32. Zuliani G, Polastri M, Romagnoli T, Marabini L, Seripa D, Cervellati C, et al. Clinical and demographic parameters predict the progression from mild cognitive impairment to dementia in elderly patients. *Aging Clin Exp Res.* 2021;33(7):1895–902. <https://doi.org/10.1007/s40520-020-01697-8>.
33. Perera G, Khondoker M, Broadbent M, Breen G, Stewart R. Factors associated with response to acetylcholinesterase inhibition in dementia: a cohort study from a secondary mental health care case register in London. *PLoS One.* 2014;9(11):e109484. <https://doi.org/10.1371/journal.pone.0109484>.
34. Ruitenber A, Ott A, van Swieten JC, Hofman A, Breteler MM. Incidence of dementia: does gender make a difference? *Neurobiol Aging.* 2001;22(4):575–80. [https://doi.org/10.1016/S0197-4580\(01\)00231-7](https://doi.org/10.1016/S0197-4580(01)00231-7).
35. Mukadam N, Cooper C, Livingston G. A systematic review of ethnicity and pathways to care in dementia. *Int J Geriatr Psychiatry.* 2011;26(1):12–20. <https://doi.org/10.1002/gps.2484>.
36. Sundström A, Westerlund O, Kotyrló E. Marital status and risk of dementia: a nationwide population-based prospective study from Sweden. *BMJ Open.* 2016;6(1):e008565.
37. Zhang Z, Liu H, Choi S-wE. Marital loss and risk of dementia: do race and gender matter. *Soc Sci Med.* 2021;275:113808.
38. Mielke MM. Sex and gender differences in Alzheimer's disease dementia. *Psychiatr Times.* 2018;35(11):14.
39. Zhou Z, Wang P, Fang Y. Loneliness and the risk of dementia among older Chinese adults: gender differences. *Aging Ment Health.* 2018;22(4):519–25. <https://doi.org/10.1080/13607863.2016.1277976>.
40. Arevalo-Rodriguez I, Smailagic N, i Figuls MR, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev.* 2015;3. <https://doi.org/10.1002/14651858.CD010783.pub2>.
41. Davis DH, Creavin ST, Yip JL, Noel-Storr AH, Brayne C, Cullum S. Montreal Cognitive Assessment for the detection of dementia. *Cochrane Database Syst Rev.* 2021;7(7).
42. Pezzotti P, Scalmana S, Mastromattei A, Di Lallo D. The accuracy of the MMSE in detecting cognitive impairment when administered by general practitioners: a prospective observational study. *BMC Fam Pract.* 2008;9(1):1–11. <https://doi.org/10.1186/1471-2296-9-29>.
43. Roalf DR, Moberg PJ, Xie SX, Wolk DA, Moelter ST, Arnold SE. Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. *Alzheimers Dement.* 2013;9(5):529–37. <https://doi.org/10.1016/j.jalz.2012.10.001>.
44. Kumar S, Sharma R, Tsunoda T, Kumarevel T, Sharma A. Forecasting the spread of COVID-19 using LSTM network. *BMC Bioinforma.* 2021;22(6):1–9. <https://doi.org/10.1186/s12859-021-04224-2>.
45. Liu Q, Li R, Hu H, Gu D. Using unsupervised deep learning technique for monocular visual odometry. *IEEE Access.* 2019;7:18076–88. <https://doi.org/10.1109/ACCESS.2019.2896988>.
46. Guo A, Beheshti R, Khan YM, Langabeer JR, Foraker RE. Predicting cardiovascular health trajectories in time-series electronic health records with LSTM models. *BMC Med Inform Decis Mak.* 2021;21(1):1–10. <https://doi.org/10.1186/s12911-020-01345-1>.
47. Zhao J, Feng Q, Wu P, Lupu RA, Wilke RA, Wells QS, et al. Learning from longitudinal data in electronic health record and genetic data to improve cardiovascular event prediction. *Sci Rep.* 2019;9(1):1–10. <https://doi.org/10.1038/s41598-018-36745-x>.
48. Putt RVD, Dineen C, Janes D, Series H, McShane R. Effectiveness of acetylcholinesterase inhibitors: diagnosis and severity as predictors of response in routine practice. *Int J Geriatr Psychiatry.* 2006;21(8):755–60. <https://doi.org/10.1002/gps.1557>.
49. Wattmo C, Wallin ÅK, Minthon L. Functional response to cholinesterase inhibitor therapy in a naturalistic Alzheimer's disease cohort. *BMC Neurol.* 2012;12(1):1–10. <https://doi.org/10.1186/1471-2377-12-134>.
50. Harutyunyan H, Khachatryan H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. *Sci Data.* 2019;6(1):1–18. <https://doi.org/10.1038/s41597-019-0103-9>.
51. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics.* 2010;26(10):1340–7. <https://doi.org/10.1093/bioinformatics/btq134>.
52. Ramachandran KM, Tsokos CP. *Mathematical statistics with applications in R 3rd Edition*: Academic Press; 2020. <https://doi.org/10.1016/C2018-0-02285-9>.
53. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. *TensorFlow: large-scale machine learning on heterogeneous distributed systems*. arXiv preprint arXiv. 2016:160304467.
54. Chollet F. *Deep learning with Python*: Simon and Schuster; 2021. <https://www.manning.com/books/deep-learning-with-python-second-edition>
55. McShane R, Westby MJ, Roberts E, Minakaran N, Schneider L, Farrimond LE, et al. Memantine for dementia. *Cochrane Database Syst Rev.* 2019;3. <https://doi.org/10.1002/14651858.CD003154.pub6>.
56. Robinson DM, Keating GM. Memantine: a review of its use in Alzheimer's disease. *Drugs.* 2006;66(11):1515–34. <https://doi.org/10.2165/00003495-200666110-00015>.
57. Cui C-C, Sun Y, Wang X-Y, Zhang Y, Xing Y. The effect of anti-dementia drugs on Alzheimer disease-induced cognitive impairment: a network meta-analysis. *Medicine (Baltimore).* 2019;98(27):e16091.
58. Whitehead A, Perdomo C, Pratt RD, Birks J, Wilcock GK, Evans JG. Donepezil for the symptomatic treatment of patients with mild to moderate Alzheimer's disease: a meta-analysis of individual patient data from randomised controlled trials. *Int J Geriatr Psychiatry.* 2004;19(7):624–33. <https://doi.org/10.1002/gps.1133>.
59. Knight R, Khondoker M, Magill N, Stewart R, Landau S. A systematic review and meta-analysis of the effectiveness of acetylcholinesterase inhibitors and memantine in treating the cognitive symptoms of dementia. *Dement Geriatr Cogn Disord.* 2018;45(3-4):131–51. <https://doi.org/10.1159/000486546>.
60. Hazif-Thomas C, Lefebvre des Noëttes V. Issues of the delisting of anti-Alzheimer's drugs in France: between the law and ethics. *Geriatr Psychol Neuropsychiatr Vieil.* 2020;18(1):97–102. <https://doi.org/10.1684/pnv.2020.0843>.

61. Biundo R, Weis L, Bostantjopoulou S, Stefanova E, Falup-Pecurariu C, Kramberger M, et al. MMSE and MoCA in Parkinson's disease and dementia with Lewy bodies: a multicenter 1-year follow-up study. *J Neural Transm*. 2016;123(4):431–8. <https://doi.org/10.1007/s00702-016-1517-6>.
62. Dufouil C, Clayton D, Brayne C, Chi L-Y, Denning TR, Paykel E, et al. Population norms for the MMSE in the very old: estimates based on longitudinal data. *Neurology*. 2000;55(11):1609–13. <https://doi.org/10.1212/WNL.55.11.1609>.
63. Perera G, Mueller C, Stewart R. Factors associated with slow progression of cognitive impairment following first dementia diagnosis. *Int J Geriatr Psychiatry*. 2021;36(2):271–85. <https://doi.org/10.1002/gps.5420>.
64. Bae J, Stocks J, Heywood A, Jung Y, Jenkins L, Hill V, et al. Transfer learning for predicting conversion from mild cognitive impairment to dementia of Alzheimer's type based on a three-dimensional convolutional neural network. *Neurobiol Aging*. 2021;99:53–64. <https://doi.org/10.1016/j.neurobiolaging.2020.12.005>.
65. De Crescenzo F, Garriga C, Tomlinson A, Coupland C, Efthimiou O, Fazel S, et al. Real-world effect of antidepressants for depressive disorder in primary care: protocol of a population-based cohort study. *Evid Based Ment Health*. 2020;23(3):122–6. <https://doi.org/10.1136/ebmental-2020-300149>.
66. Mateos-Álvarez R, Ramos-Ríos R, López-Moríñigo J. Comparative analysis between the MMSE and the RUDAS for dementia screening in low educated people in a Spanish psychogeriatric clinic. *Eur J Psychiatry*. 2017;31(3):119–26. <https://doi.org/10.1016/j.ejpsy.2017.06.003>.
67. Mitchell AJ. The Mini-Mental State Examination (MMSE): update on its diagnostic accuracy and clinical utility for cognitive disorders. *Cogn Screen Instrum*: Springer. 2017:37–48. https://doi.org/10.1007/978-3-319-44775-9_3.
68. Arevalo-Rodríguez I, Smailagic N, Roqué-Figuls M, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev*. 2021;7(7). <https://doi.org/10.1002/14651858.CD010783.pub3>.
69. Piersma D, Fuermaier A, de Waard D, De Deyn P, Davidse R, De Groot J, et al. The MMSE should not be the sole indicator of fitness to drive in mild Alzheimer's dementia. *Acta Neurol Belg*. 2018;118(4):637–42. <https://doi.org/10.1007/s13760-018-1036-3>.
70. O'Brien JT, Holmes C, Jones M, Jones R, Livingston G, McKeith I, et al. Clinical practice with anti-dementia drugs: a revised (third) consensus statement from the British Association for Psychopharmacology. *J Psychopharmacol*. 2017;31(2):147–68. <https://doi.org/10.1177/0269881116680924>.
71. Levine SZ, Yoshida K, Goldberg Y, Samara M, Cipriani A, Efthimiou O, et al. Linking the Mini-Mental State Examination, the Alzheimer's Disease Assessment Scale–Cognitive Subscale and the Severe Impairment Battery: evidence from individual participant data from five randomised clinical trials of donepezil. *Evid Based Ment Health*. 2021;24(2):56–61. <https://doi.org/10.1136/ebmental-2020-300184>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

