A neural network approach to optimising treatments for depression using data from specialist and community psychiatric services in Australia, New Zealand and Japan

Aidan Cousins[1] · Lucas Nakano[2] · Emma Schofield[1] · Rasa Kabaila[3]

Abstract

This study investigated the application of a recurrent neural network for optimising pharmacological treatment for depression. A clinical dataset of 458 participants from specialist and community psychiatric services in Australia, New Zealand and Japan were extracted from an existing custom-built, web-based tool called *Psynary* . This data, which included baseline and self-completed reviews, was used to train and refine a novel algorithm which was a fully connected network feature extractor and long short-term memory algorithm was firstly trained in isolation and then integrated and annealed using slow learning rates due to the low dimensionality of the data. The accuracy of predicting depression remission before processing patient review data was 49.8%. After processing only 2 reviews, the accuracy was 76.5%. When considering a change in medication, the precision of changing medications was 97.4% and the recall was 71.4% . The medications with predicted best results were antipsychotics (88%) and selective serotonin reuptake inhibitors (87.9%). *This is the first study that has created an all-in-one algorithm for optimising treatments for all subtypes of depression.* Reducing treatment optimisation time for patients suffering with depression may lead to earlier remission and hence reduce the high levels of disability associated with the condition. Furthermore, in a setting where mental health conditions are increasing strain on mental health services, the utilisation of web-based tools for remote monitoring and machine/deep learning algorithms may assist clinicians in both specialist and primary care in extending specialist mental healthcare to a larger patient community.

Keywords Depression · Machine learning · Treatment optimisation · Mental health · LSTM

✉ Aidan Cousins
a.cousins@student.unsw.edu.au

Lucas Nakano
l-nakano-0q2@eagle.sophia.ac.jp

Emma Schofield
e.schofield@unsw.edu.au

Rasa Kabaila
Rasa.Kabaila@health.nsw.gov.au

1 Faculty of Medicine and Health, Rural Clinical School, University of New South Wales, Port Macquarie 2444, Australia

2 Sofia University, Tokyo, Japan

3 Mid North Coast Local Health District, Community Mental Health Services, Port Macquarie 2444, Australia

1 Introduction

1.1 Depression

According to the International Statistical Classification of Disease Revision 10 (ICD-10), depression is categorised under mood (affective) disorders where the central disturbance is a change in mood to depression or elation [1]. Depressive episodes occur when the patient suffers from decreased mood, energy, activity, self-esteem and self-confidence [1]. Patients exhibit lesser capacity for enjoyment and decreased interest in their avocations while also exhibiting ideas of guilt or worthlessness [1]. There are several subtypes of depression including major depressive disorder, bipolar disorder and melancholic depression [1].

Major depressive disorder (MDD) is described as a short- or long-term impairment causing significant reductions in quality of life and psychosocial functioning by affecting areas of life including mood, affect, motivation and cognition [2]. According to the ICD-10, bipolar disorder is characterised by two or more episodes where the patient's mood and activity levels are significantly altered resulting in occasions of elevation of mood and increased energy or decreased energy and activity [1]. Hypomania is a period of persistent mild elevation of mood, increased energy and activity, while mania is mood elevated higher than what would be expected in the patient's circumstances [1]. Melancholic depression has been a contentious issue within psychiatric circles, with some viewing the disorder as a dimension of a severe expression of clinical depression ("melancholia"), while others believe it is a separate type of depressive disorder [3–5]. Regardless of its classification, melancholic depression is a severe form of depression with prominent neurovegetative symptoms [3].

Depression is a major public health issue and cause of disability [6]. An analysis of data gathered in the Global Burden of Disease study (held from 1990 to 2017) showed that the incidence of depression had increased worldwide from 172 to 258 million or by 49.86% [7]. The study further analysed age-standardised incidence rates (ASR) and estimated annual percentage changes (EAPC) of the 195 countries included in the population study. ASR was found to be significantly increased in 29 countries, slightly increased in 132 countries, slightly decreased in 9 countries and significantly decreased in 9 countries [7]. Interestingly, the number of people with depression increased in all five socio-demographic index (SDI) levels (low, low-middle, middle, high-middle and high); however, the ASR only increased in the high-SDI region [7]. Geographically, the number of people with depression increased in all geographical locations. 93.7% of patients with depression in 2017 were found to have MDD [7].

A systematic review conducted on the same Global Burden of Disease data by Ferrari et al. showed that depressive disorders were the second leading cause of years lived with a disability (YLDs) in 2010. MDD accounted for 8.2% of global YLDs [6]. Depressive disorders were also a leading cause of disability-adjusted life-years (DALYs) with MDD accounting for 2.5% of global DALYs. Furthermore, MDD was found to be the cause of 16 million suicide DALYs.

In Australia, the *National Survey of Mental Health and Wellbeing* completed in 2007 estimated that 45% of the Australian population aged between 16 and 85 years would experience a mental disorder during their lifetime and 1 in 5 had experienced a mental disorder in the last year [8]. Since 1998, there have been many attempts to quantify the prevalence of depression with a 2019 study showing

depressive symptoms in 7.4% and 13.2% of Australian males and females, respectively [9–12]. The economic impact of mental health illness in Australia was estimated to be $10.6 billion AUD in the 2018–2019 financial year with studies projecting that health service costs will increase by 45% between 2006 and 2026, and that the cumulative cost of mental illness over the next 30 years will exceed $2.63 trillion AUD [8, 13–15].

In New Zealand, the *2018 New Zealand Mental Health Monitor (NZMHM)* provides one of the most recent reviews of the mental health of the New Zealand general population [16]. The NZMHM found that 32% of New Zealanders had an experience with mental distress and an additional 32% of New Zealanders lived with someone with a lifetime experience of mental distress [16]. Furthermore, 49% of the population were aware of a close friend who experiences mental distress [16]. During the recent COVID-19 pandemic in 2021, Gasteiger et al. completed a cross-sectional study with a cohort size of 681 adults older than 18 in New Zealand [17]. While the sample of the New Zealand population was 89% female and older than the median age of the general population (40 and 37.4 years, respectively), they found that 64% of participants reported symptoms of depression, 53% reported symptoms of anxiety, 31% reported moderate-to-severe symptoms of depression, and 24% reported moderate-to-severe symptoms of anxiety [17]. Although outdated, the 2005 *Depression Service Plan* released by the Midcentral District Health Board in New Zealand estimated the cost of depression in New Zealand at $750 million per year [18]. More recent estimates are correlated to population-adjusting Australian figures.

Similar to Australia and New Zealand, Japan has a high mental health burden [19, 20]. Community-based mental health surveys play an important role in estimating the prevalence of mental disorders because most people do not seek treatment even if they experience psychological distress equivalent to diagnosable mental disorders [20]. *The World Mental Health Japan Survey Second* conducted from 2013 to 2015 with a sample size of 2450, randomly selected residents between the ages of 20 and 75 [20, 21]. They found a lifetime prevalence of any mood disorder of 4.57% and 7.21% for men and women, respectively [20, 21]. They also found a 12-month prevalence of any mood disorder of 2.24% and 3.26% for men and women, respectively [20, 21]. Other studies have estimated the economic impact of depression to be between 1.29 billion and 3 trillion yen (or $16.46 billion AUD to $38 billion AUD) yearly [22, 23].

1.2 Diagnosis, treatment and management

Currently, there are no biological markers for depression; therefore, diagnosis is based only on symptomatology [24].

A wide range of screening tests that vary in length, style, administration and psychometric evaluation are currently used [25].

The *Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Report*, released in 2004, was an important milestone in optimising management of patients with depression [26, 27]. Four thousand patients from 41 different primary and psychiatric care sites were evaluated using a novel 4-level treatment paradigm. The remission rate, using *The Quick Inventory of Depressive Symptomatology*, was 32.9%, 30.6%, 13.6% and 14.7% for the levels 1, 2, 3 and 4, respectively [27]. Although a cumulative remission rate of 65.8% was achieved, crucially, 34.2% did not improve with medical interventions for depression [27]. For non-remission patients, it is critical to identify the earliest point at which to stop further medication trials because the longer the time to remission, the less chance a patient has of reaching remission [27].Optimising an individual patient's treatment plan so that they have the highest likelihood for remission is a needed advancement in psychiatry.

1.3 Psynary

Psynary (www.psynary.com) is a web-based tool used to support clinicians, organisations and patients in the diagnosis and treatment of mood and anxiety disorders in accordance with the ICD-10 [28]. The Psynary system was designed to collect de-identified data in a format that could be represented numerically with no requirement for language processing. It is anonymous, which avoids privacy and data protection issues that can be associated with online platforms [29].

The *Optimisation of Treatment for Mood and Anxiety Disorders Study 1* (OptiMA1) involved two parallel studies in New Zealand and Japan for the purpose of validating key outcome measures developed for the Psynary platform [28]. These outcome measures included the *R8 Depression* score and *R8 Anxiety* score [28]. The *R8 Depression* score was designed to fully capture the wide range of symptom domains seen in depression, across the full range of illness severity, and to be sensitive to treatment effects [28]. Participants (n=270) were recruited from patients registered on Psynary by the public community mental health clinic at Nelson Marlborough District Health Board in New Zealand (n=62) and by the private clinic serving the English-speaking population at the American Clinic Tokyo in Japan (n=208) [28]. Patients with probable mood or anxiety disorders who registered to Psynary between 24th March 2016 and 25th October 2018 were invited to complete either an online or written consent process prior to participating in the study [28]. Inclusion criteria included completing Psynary in the English language, being over

18 years of age for NZ, or 20 years of age for Japan, and having an ICD-10 diagnosis of a current depressive episode (unipolar or bipolar) or anxiety disorder (ICD-10 F31.3, F31.4, F31.81, F32.1, F32.2, F33.1, F33.2, F40-F43) confirmed by the treating clinician at their initial appointment [28].

An early analysis of the cohort (n=131) suggested a similar doubling of remission rates in response to treatment optimisation, but over a shorter 90-day time period compared to the STAR*D trial (Fig. 1) [29]. This validation study found that if patients had less than a 20% reduction in *R8 Depression score* after 6 days of treatment, the negative predictive value for non-remission was 98% [29] (Table 1) .

From OptiMA1, Psynary appears to be beneficial in monitoring response to treatment and guiding timing of medication optimisation [28]. These findings suggested that the Psynary database could potentially be utilised to develop predictive models of response to treatment and guide treatment selection [28]. Since then, OptiMA2 has been conducted, a qualitative study to establish Nurse Practitioner-Psynary-assisted care pathway in Port Macquarie, New South Wales, Australia. This is now being followed up by OptiMA3 that will collect naturalistic clinical outcomes from that pathway. Both these studies also include ethics approval to analyse the naturalistic clinical outcomes from all participants. These have been included in this current study to examine the feasibility of using the Psynary database to conduct machine learning approaches to develop predictive algorithms to guide treatment selection.

1.4 Applying predictive tools in mental health

Unlike other medical specialties that heavily rely on quantitative biomarkers to assist in the diagnosis of diseases, treatment planning and measurement of outcomes, mental health still predominantly relies on clinical measures. Mental health team members use patient interviews,
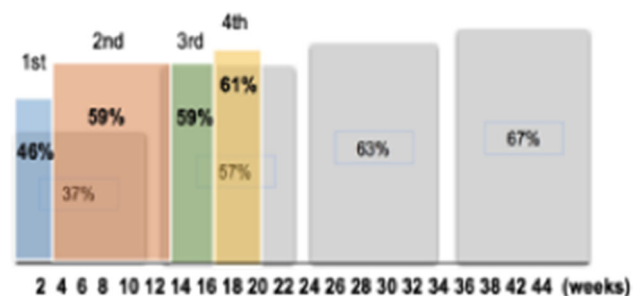

Fig. 1 OptiMA1 Data vs STAR*D Trial Remission Percentages over a 48-week period [29]. The OptiMA1 study results are shown in colour, while the STAR*D trial is shown in grey (Colour figure online)

Table 1 Day 4–6 *R8 Depression* percentage reduction [29]

Group	R8 reduction<20%	R8 reduction>20%	Total (*n*)
Total (*n*)	101	64	165
Non-remission (*n*, %)	99 (98%)	38 (59%)	137 (83%)
Remission (*n*, %)	2 (2%)	26 (41%)	28 (17%)

questionnaires and patient reports to evaluate signs and symptoms [30, 31]. The experience and subjectivity of the clinician are heavily leveraged to make inferences about symptomatology from this data, which is difficult to imitate using supervised deep learning (DL) models. Supervised DL models require a training set containing "true" labels to optimise model parameters before the model can be used to predict the diagnostic outcome of new subjects. Therefore, the quality of expert-provided diagnostic labels used for training sets the upper-bound for the predictive performance of the model [31–33].

1.5 Related studies and study rationale

Despite the difficulties, there has been a large amount of research in the usage of ML/DL techniques in mental health. A number of studies have focussed on scraping social media posts to predict depression for example, using multinomial naive Bayes (with an accuracy of 76.69%) and event-driven tendency warning models (best recall rate of 0.668 and *F*-measure of 0.624) [34, 35]. Other studies have focussed on analysing electronic health records. Nemesure et al. used a sample of 4184 students who underwent a general health and psychiatric assessment for the diagnosis of MDD and Generalised Anxiety Disorder. A high-level XGBoost model was used to produce an area under the receiver operating characteristic (AUROC) of 0.67, a sensitivity of 0.55 and specificity of 0.7 for major depression [36]. Further studies have used ML/DL techniques to classify, diagnose and grade depression using clinical data, externally validated screening tests and biomarkers [37–44].

In 2018, Gao, Calhoun and Sui completed a review of 66 studies focussed on MDD that have used magnetic resonance imaging to either classify MDD from controls or other mood disorders or investigated treatment outcome predictors for individual patients [45]. The 66 studies investigated by Gao, Calhoun and Sui could be further classified into 9 groups of studies that were focussed on diagnosis/classification of only MDD [46–66], only bipolar disorder [67], diagnosis/classification of MDD compared to bipolar disorder [68–82], diagnosis/classification of MDD compared to Generalised Anxiety Disorder [83], diagnosis/classification of Schizophrenia and mood disorders (including MDD and Generalised Anxiety Disorder) [84, 85], brain abnormalities in patients who have been

diagnosed with a particular mood disorder [86–94], analysing therapeutic responses to MDD [95–100], using neurobiological markers/neuroimaging for diagnosis/classification [101–109] and predicting responses to electro-convulsive therapy [110, 111].

In 2021, de Nijs et al. were able to create individualised models to predict 3- and 6-year symptomatic and global outcomes of patients with schizophrenia-spectrum disorders (mainly with established illness, but variable illness duration) based on patient-reportable data with a study size of 523 schizophrenia-spectrum patients [112]. Also in 2021, Taliaz et al. used genetic, clinical and demographic data from patients with solely MDD in the *STAR*D Report* to create an ML algorithm that generated an accurate predictor of response to three antidepressant medications with an average balanced accuracy of 72.3% and 70.1% across the medications in validation and test sets, respectively [113]. They then obtained data from the *Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS)* of patients treated with citalopram (a selective serotonin reuptake inhibitor) [113]. This external validation yielded accuracy of 60.5% and 61.3% for the *STAR*D* and *PGRN-AMPS*, respectively [113].

The aim of this study is to demonstrate how the clinical data collected by Psynary can be used to optimise the treatment of depression in the clinical setting. We hypothesised that ML/DL techniques could be used on the data collected in the Psynary database to support the optimisation of the treatment and management of depression. Although there have been previous studies that used clinical data to build ML/DL algorithms to diagnose depression and a quite recent study that predicted the response of three medications in treating MDD, *we are unaware of any published study that has used clinical data to build a single, all-inclusive ML/DL algorithm to optimise the treatment of all subtypes of depression.*

2 Methodology

2.1 Data collection and preprocessing

The Psynary database included details from patients being treated for depression (*in any of its forms*) at associated community and specialist psychiatric facilities in either

American Clinic Tokyo, Tokyo, Japan, Tokyo Mental Health, Tokyo, Japan, Kawai Clinic, Otago, New Zealand, or Port Macquarie Base Hospital Mental Health Service, Port Macquarie, Australia, between the dates of 24/03/2016 and 01/06/2021. Other inclusion criteria were 18 years or older, had capacity to consent, fluency in English or Japanese (as Psynary is only available in those two languages), patients who had alcohol-related comorbidities with a primary diagnosis of a mood or anxiety disorder and exclusion criteria were clients presenting with psychotic symptoms, significant comorbid alcohol and drug misuse where these were the primary diagnoses, terminal/life threatening physical comorbidity and cognitive impairment or intellectual disability. These were assessed on the basis of past medical history, liaison with GP and, with permission, collateral history from a relative or carer. All patients consented to their de-identified data being included in the Psynary database for the purposes of ongoing research, and the OptiMA1, 2 and 3 studies were approved by the relevant local ethics committees. OPTIMA patients were instructed to create an online Psynary account and complete a baseline evaluation with the in-person support of a psychiatrist or nurse practitioner and then complete weekly reviews either by themselves remotely or with the help of a nurse practitioner with expertise in mental health. The Psynary database then logged this baseline information which included; past history such as family history, past episodes of depression and other disorders, age of onset, hospital admissions, past deliberate self-harm and past attempted suicide; deliberate self-harm questions including suicidal thoughts intensity, suicidal planning and suicidal attempts and other domains such as alcohol intake, current psychiatric medication (name and dose) previous treatment changes and whether they had electroconvulsive therapy sessions. The Psynary system incorporates the Hypomania Checklist 16 item (HCL-16) [114], Generalised Anxiety Disorder-7 (GAD-7) [115] and Patient Health Questionnaire-9 (PHQ-9) [25] scores and generates ICD-10 diagnoses for all major mood and anxiety disorders. It also includes the proprietary main outcome measures of this study, the *R8 Depression* and *R8 Anxiety* scores. All of these metrics were used in building the algorithm reported in this article except for the HCL-16.

When a patient completed a review, the Psynary database then logged; current psychiatric medication list, current alcohol intake, current deliberate self-harm responses, recent electroconvulsive therapy treatment, *R8 Depression, R8 Anxiety*, PHQ-9, GAD-7 current scores. All of these metrics were used in the review aspect of the model except for the HCL-16.

On the 1st of June 2021, anonymised clinical data from Psynary was exported and preprocessing undertaken. All patient demographic categorical data was one-hot encoded and empty data values were handled by denoting 0 to "No" and -1 to "No Answer" depending on the question field. All data preprocessing and algorithm creation was built on Jupyter lab version 3.0.14, python version 3.8.5 and using an AMD Ryzen Threadripper 2950X 16-Core Processor 3.50 GHz, 32.0 GB of installed RAM and an NVIDIA RTX 2080 GPU.

The review data for each patient was stored in chronological order and filtered and preprocessed in function of the medications taken during the period corresponding to that review. Medications were categorised into one of the following drug classes: analgesics, antidepressants, antihistamines, antipsychotics, anxiolytics, benzodiazepines, hypnotics, mood stabilisers, opioid antagonists and stimulants. Antidepressants were further divided into their mechanism of action: mono-amine oxidase inhibitors (MAOIs), noradrenaline reuptake inhibitors (NaRIs), selective serotonin reuptake inhibitors (SSRIs), serotonin–norepinephrine reuptake inhibitors (SNRIs), tricyclic antidepressants (TCAs) and atypical antidepressant medications. Groups with low representation in reviews were discarded, as were patients without reviews for better-represented medications. If, for a given review, a respondent had taken medications of more than one group, that review appears as duplicates in each corresponding array. As is expected from this type of data, review length frequency decreased exponentially (Fig. 2). The average number of reviews per patient was 11.5, with 95% of patients having 39 reviews or fewer.

The objective of our analysis was to propose treatment by predicting the effectiveness of psychiatric medications for each individual patient. To better fit the available data, we formulated our problem as a regression of the best *R8 Depression* score the patient would achieve while taking any given medication [28]. While training the model, the sequence of reviews provided to the model was truncated randomly, for each medication. This truncation used a geometric distribution to allow for early prediction of remission.

In addition to the regression of *R8 Depression* score prediction, the model was trained to predict which medications will be prescribed by clinicians. We found this additional objective to both help prevent overfitting by increasing the ratio of training data to model capacity and improve performance of the model when used as a recommendation system. To do this, the medications prescribed by clinicians in all reviews are given as a multi-hot vector. For medication optimization, predicted *R8 Depression* scores S are corrected by coefficients calculated from the prescription probabilities P scaled with a learned factor β, the purpose of which is to control the relative importance of the prescription probabilities. The
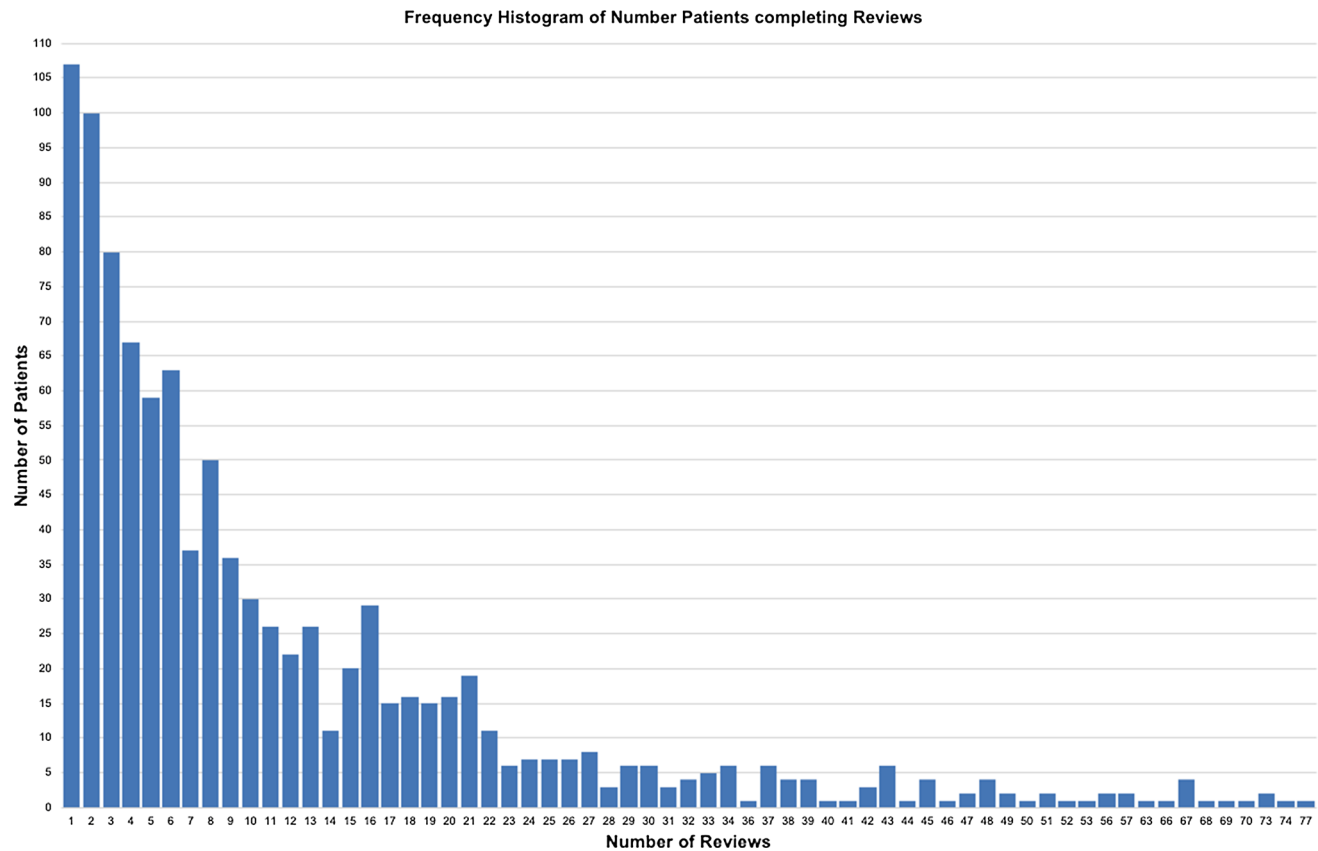
Fig. 2 Frequency histogram illustrating the number of reviews participants completed

algorithm then proposes the medication with the lowest scaled score (Eq. 1).

$$M = \text{Arg min}\left(\frac{S}{\text{Softmax}(\beta P)}\right) \quad (1)$$

2.2 Implementation of long short-term memory model

Neural network models have become the de-facto standard in most machine learning applications. There are several choices of architecture for sequence modelling, most notably recurrent neural networks (RNN), 1-dimensional convolutional networks and attention networks [116]. The advantages of the latter two have largely to do with better gradient propagation and performance with long sequence lengths [116]. The Psynary review data, having short sequence lengths and unidirectional chronological structure, appears to be best suited for the simple-to-implement RNN structure. For this particular application, a long short-term memory (LSTM) network was chosen due to its ready availability in deep learning libraries and proven effectiveness [117]. The architecture of the model in this study separately processes each sequence of reviews corresponding to each medication using the same LSTM

network. We suspect that the patterns learned by networks trained in each medication separately are largely similar, and training an individual network for each medication group would severely reduce sample sizes due to subdivision of the dataset. Using a larger, single recurrent network was empirically found to produce better and more consistent results. Fully connected networks filled in the rest of the picture—they were used for feature extraction from both the baseline questionnaire and reviews, and at the end of the network for the final *R8 Depression* score and doctor prescription predictions.

Training was performed in three steps. The fully connected network (FCN) review feature extractor and LSTM modules were first trained in isolation on the *R8 Depression* score prediction task. This recurrent model was then integrated into the baseline questionnaire feature extractor and final FCN layers with frozen weights, and the remainder of the model trained. Finally, the entire model was annealed using low learning rates. Due to the low dimensionality of the data, small size of the model and its recurrent structure, GPU acceleration was not used. Figure 3 shows a diagrammatic illustration of the designed algorithm.

Fig. 3 Block diagram of created algorithm
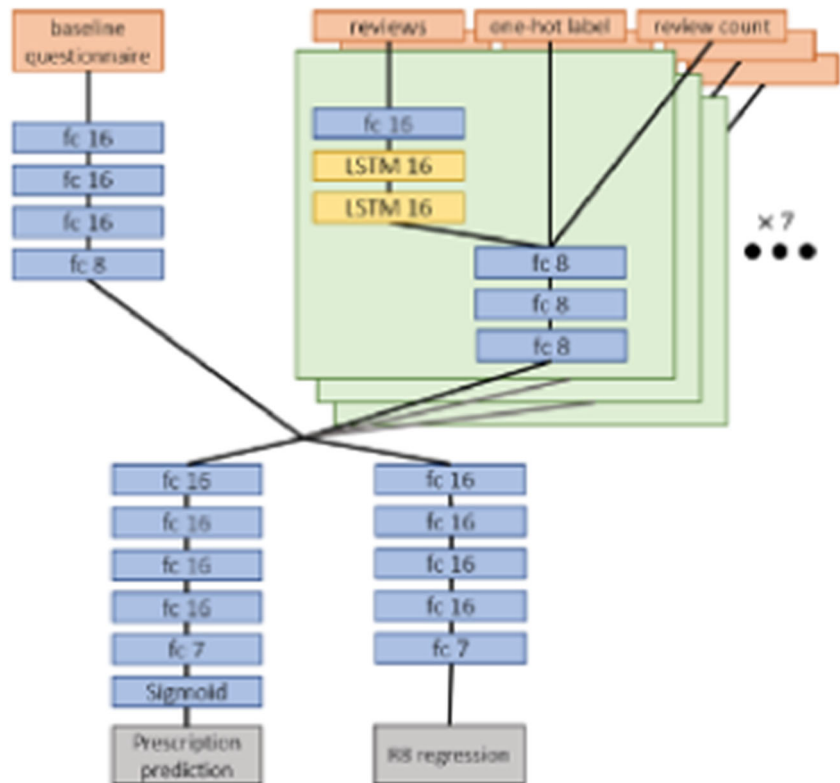
2.3 Optimisation test of model

To further validate the significance of the model, an optimisation test was performed. The model's outputs were taken and tested by posing the questions "After seeing the patient review data, should the medication regimen change or should it continue?" and "If the medication should change, what should the medication be changed to?".

Because of the extra data dimensionality of patient reviews, instead of sensitivity, specificity, positive predictive and negative predictive values being calculated, the sensitivity, positive predictive value, false positive rate and correct medication change accuracy were calculated (Eqs. 2, 3, 4 and 5).

$$\text{Sensitivity} = \frac{\text{Number of Correct Change Decisions}}{\text{Total Number of Required Changes}} \quad (2)$$

$$\text{Positive Predictive Value} = \frac{\text{Number of Correct Change Decisions}}{\text{Total Number of Change Decisions}} \quad (3)$$

$$\text{False Positive Rate} = \frac{\text{Number of Incorrect Change Decisions}}{\text{Total Number of Change Decisions}} \quad (4)$$

$$\text{Correct Medication Change Accuracy} = \frac{\text{Number of Correct Medication Choices}}{\text{Total Number of Medication Choices}} \quad (5)$$

3 Results

This study utilised data collected from the OptiMA1, 2 and 3 [28, 118] studies, contained in the Psynary database. In total, data from 458 participants was included in this study (Table 2). The large majority of participants were from Japan (85.6%), followed by New Zealand (9.6%) and then Australia (4.8%). The sex distribution slightly favoured females (57%) and the median age and age range of the participants were 32.5 and 18 to 73 years, respectively (Fig. 4). Most of the participants were employed full-time (52.4%), spoke English (98.9%) and did not have the

Table 2 Variables and their descriptions used in the proposed algorithm

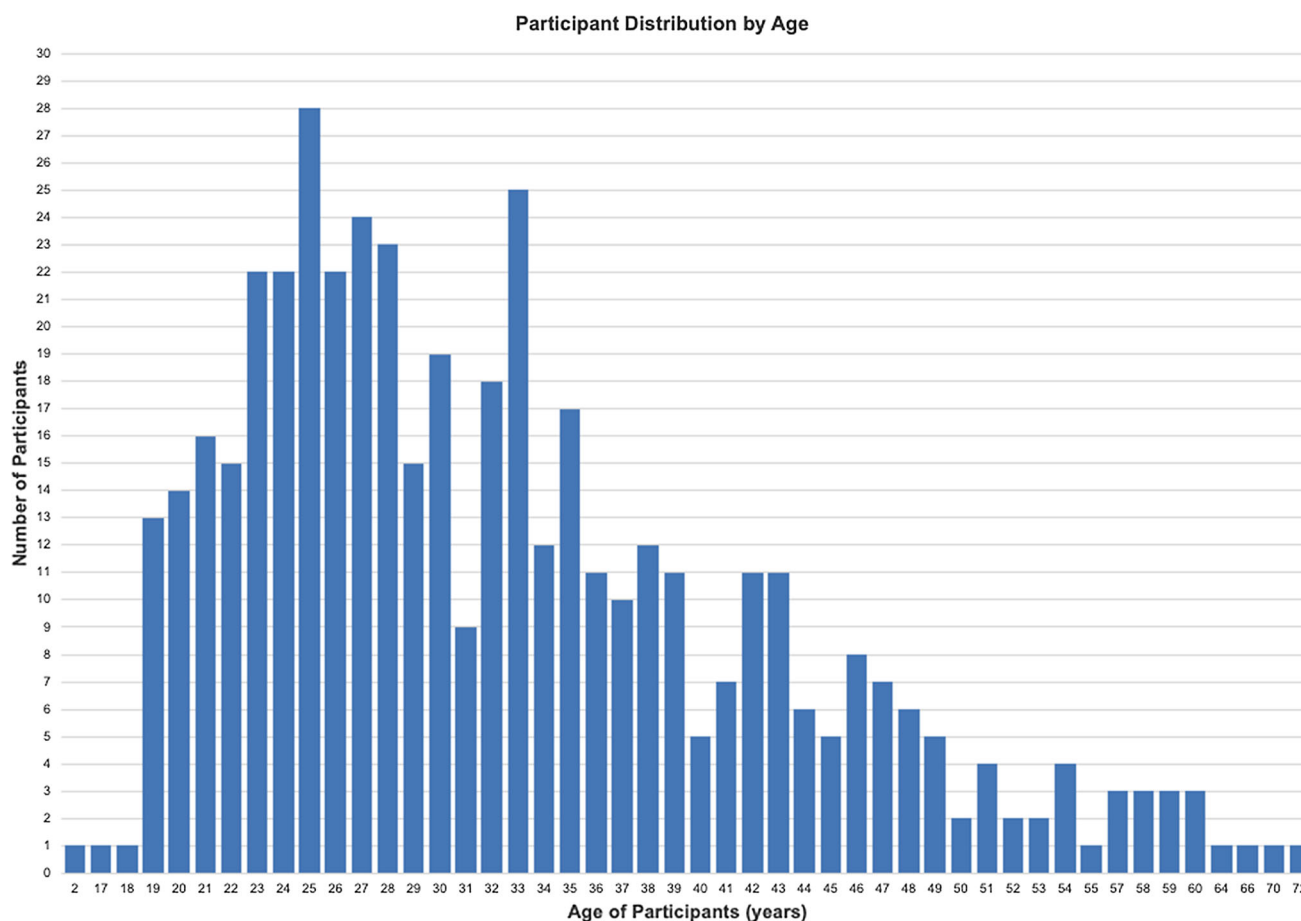| Symbol	Description
P	Prescription probabilities vector (network output)
S	Predicted R8 scores vector (network output)
β	Scalar balancing coefficient
M	Scaled medication scores

Participant Distribution by Age

Fig. 4 Frequency histogram illustrating patient age distribution (in years)

support of a carer (85.2%). The patient's age distribution showed two anomalous data points (ages 2 and 17) which were deleted and considered missing data.

The participants' psychiatric characteristics were also interpreted (Table 3). The average *R8 Depression* raw score was 34.28 out of a maximum of 84 with depression remission considered at a raw score less than 14 [28]. The PHQ-9 score is grouped based on severity into none (0–4), mild (5–9), moderate (10–14), moderately severe (15–19) and severe (20–27) [25]. The participants' average PHQ-9 score was 16.54 (considered within moderately severe depression) with 59% of the participants considered either moderately severe or severe. The GAD-7 score is grouped into mild (0–4), moderate (5–9) or severe (10–15) anxiety [115]. A total of 354 participants (77.3%) recorded scores reflecting severe anxiety. Furthermore, 326 (71%) participants were considered to have a subtype of unipolar depression and 84 participants (18.3%) were considered to have a diagnosis of a subtype of bipolar disorder . The remaining 3 participants with a provided diagnosis (0.7%) were considered to have hypomania (Table 4) .

The participants in this study were treated by two psychiatrists. An analysis of the frequency of medication classes prescribed showed SSRIs being the most prescribed medication class, followed by antipsychotics and mood stabilisers (Fig. 5). SSRIs are likely to be the most prescribed medication class due to the existing treatment guidelines in Japan [119], New Zealand and Australia [120] recommending them as first-line medications.

The model in this study, trained with the Psynary dataset, without reviews, had an accuracy of predicting *R8 Depression*-defined remission of 49.8% (seen in Fig. 6). After the model had processed 2 reviews, the model had a prediction accuracy of 76.5%.

The accuracy of predicting the individual medication classes that led to the best *R8 Depression scores* were then examined to understand the strengths and weaknesses of the algorithm. Without including reviews, it was found that the most accurate group was SSRIs (78.9%), followed by antipsychotics (53.6%), mood stabilisers (36.8%), benzodiazepines (35.7%), atypical antidepressants (29.4%), NaRIs (16.7%) and finally SNRIs (10.5%). However, when the model was trained with the reviews, antipsychotics

Table 3 Participant demographic characteristics

Demographic characteristics	n (%)
Total participants	458
Males	197 (43%)
Females	261 (57%)
Country of residence	
Australia	22 (4.8%)
New Zealand	44 (9.6%)
Japan	392 (85.6%)
Work Status	
Employed full-time	240 (52.4%)
Employed part-time	45 (9.8%)
Unpaid domestic work	26 (5.7%)
Permanently sick/disabled	4 (0.9%)
Retired	3 (0.7%)
Self-employed full-time	16 (3.5%)
Self-employed part-time	18 (3.9%)
Student	76 (16.6%)
Unemployed	30 (6.6%)
Language	
English	453 (98.9%)
Japanese	5 (1.1%)

Table 4 Participant psychiatric characteristics

Psychiatric characteristics	n (%)
Total participants	458
R8 Depression score	
Mean	34.28
Standard deviation	13.03
Maximum	77
Minimum	0
PHQ-9 score	
Mean	16.54
Standard deviation	5.57
Max	27
Min	1
0–4	2 (0.4%)
5–9	54 (11.8%)
10–14	95 (20.7%)
15–19	133 (29%)
20–27	138 (30%)
Not measured	36 (7.9%)
GAD-7 score	
Mean	9.04
<5	7 (1.5%)
5–10	61 (13.3%)
11–15	354 (77.3%)
Not measured	36 (7.9%)
ICD-10 diagnosis	
F30: hypomania	3 (0.7%)
F31.31: bipolar disorder, current episode depressed, mild	3 (0.7%)
F31.32: bipolar disorder, current episode depressed, moderate	7 (1.5%)
F31.4: bipolar affective disorder, current episode severe depression without psychotic symptoms	30 (6.6%)
F31.7: bipolar affective disorder, currently in remission	1 (0.2%)
F31.81: bipolar II disorder	43 (9.4%)
F32.0: mild depressive episode	23 (5%)
F32.1: moderate depressive episode	31 (6.8%)
F32.2: severe depressive episode without psychotic symptoms	95 (20.7%)
F33.0: recurrent depressive disorder, current episode mild	24 (5.2%)
F33.1: recurrent depressive disorder, current episode moderate	43 (9.4%)
F33.2: recurrent depressive disorder, current episode severe without psychotic symptoms	110 (24%)
Not provided	45 (9.3%)

were predicted to be the medication with the best *R8 Depression* scores (88%) followed by SSRIs (87.9%), mood stabilisers (69.1%), benzodiazepines (53.5%), atypical antidepressants (39.2%), SNRIs (27.7%) and NaRIs (25.9%) (Fig. 7). These findings are not surprising as they mirror the frequency of patients prescribed a particular medication class (Fig. 6).

Analysis of training and test loss per epoch graph showed no significant changes between training and test loss (Fig. 8).

We also optimised the β coefficient (learned factor) to balance the trade-off between *R8 Depression* scores and prescribed medication prediction. The best β coefficient was found to be 0.6465 (Fig. 9).

We then completed an optimisation assessment of our model. Due to the extra dimension of patient reviews, an error matrix was not developed. Instead, we evaluated the model by assessing if asked to change medication, whether the model continued with the current medication or changed and then if it changed whether the model selected the correct medication (Fig. 10). After just 2 reviews, the precision of changing medications (Change Recommendation Precision) was 97.4%, the recall (Change Recommendation Recall) was 71.4%, the accuracy of changing medications to the correct medication (Correct Change
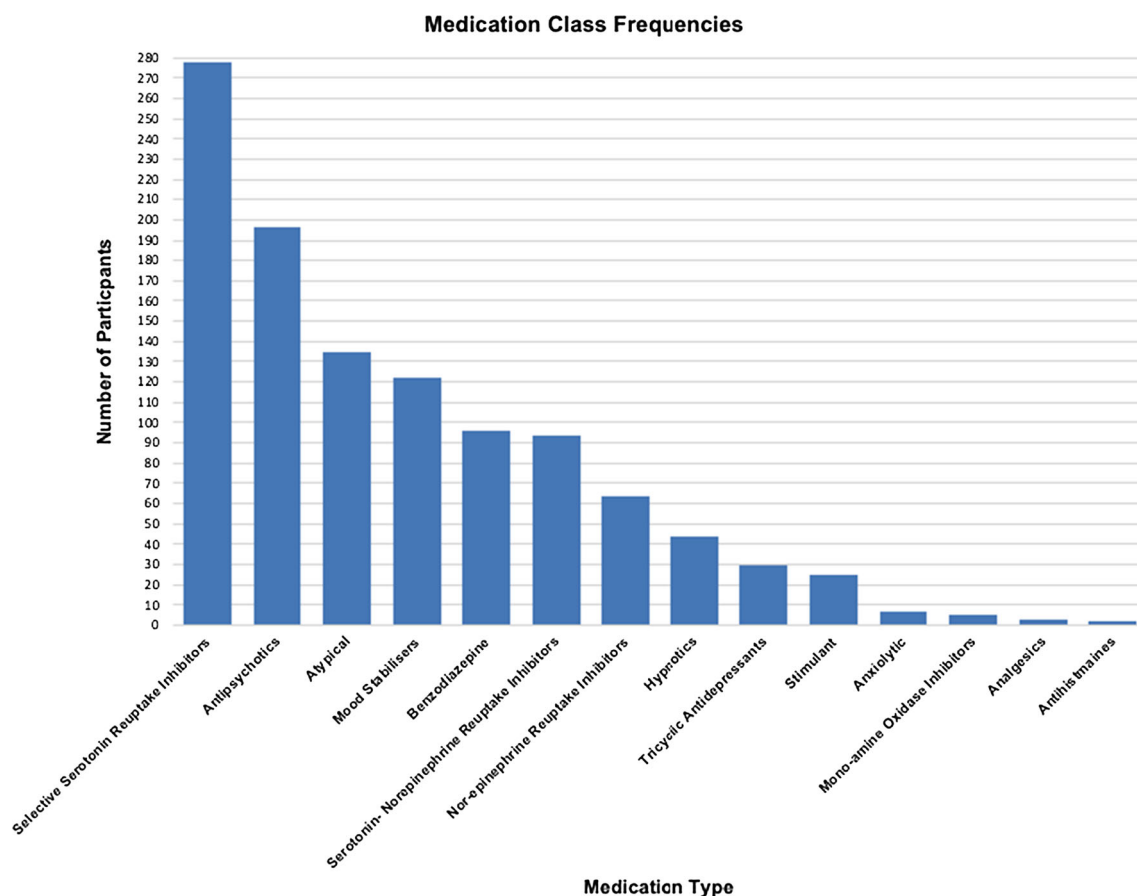
Medication Class Frequencies

Fig. 5 Frequency histogram illustrating the number of patients prescribed a particular medication class

Recommendation) was 54.3%, and the false positive rate (Incorrect Change Recommendation) was 2.6% (Fig. 10).

Finally, we reviewed the mean absolute error by number of reviews. Our model can predict *R8 Depression* scores before reviews with a mean absolute error of 14.6 ± 9.5. As the number of reviews increases, accuracy increases and deviation decreases significantly, though with some instability, likely due to the decreasing sample count (as shown in Fig. 11).

4 Discussion

In this study, we describe the development of a DL algorithm that predicts reaching remission from depression in response to psychiatric medications. The capabilities of the algorithm are shown by its ability to initially predict remission at an accuracy of 49.8% before processing reviews (Fig. 6). After only 2 reviews, the model had an accuracy of 76.5%. The medications predicted to have the best *R8 Depression* score results were antipsychotics (88%), followed by SSRIs (87.9%) and mood stabilisers (69.1%) (Fig. 7). This is of significant clinical importance

as current treatment protocols for all depression subtypes are largely reliant on trial-and-error of treatment guidelines. Using the model created in this study as an adjunct to normal care could reduce the iterations of trial-and-error.

Our model was designed around the restrictions imposed by our available data. Several design decisions were made that took into account the specific biases imposed by the naturalistic dataset. An analysis of a single medication was made difficult by the large variety of prescribed medication, which subdivided the data into very small subsets. To ameliorate this and to reduce dataset entropy, the data was categorised by medication class, with antidepressants divided into groups based on mechanism of action. Furthermore, many patients were given more than one medication at a time, which made unitary classification incompatible with the desired application of the model. Because the best indicator of remission was the *R8 Depression* score, regression of this value is a very similar task to binary prediction of remission, with the added benefit of giving a measure of expected improvement.

The biggest performance improvement came from the addition of the auxiliary prediction of prescribed medications. This served as a heuristic technique that directed
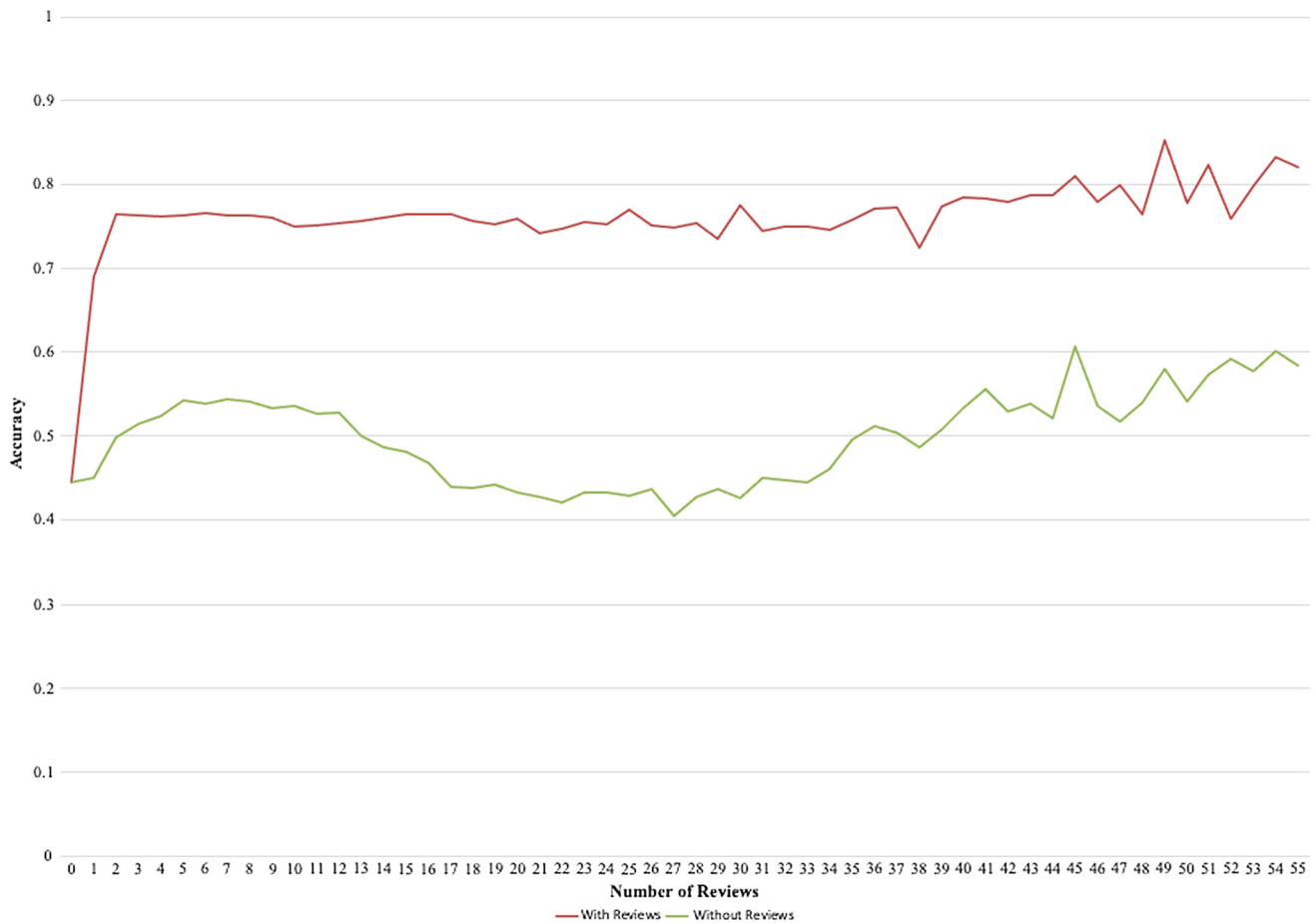
Fig. 6 Accuracy of the developed model at predicting a patient's *R8 Depression* score, with and without reviews

medication selection to that which a doctor is likely to prescribe during the treatment process. Put simply, it acts as a tiebreaker, allowing the model to accurately select between medication types it expects to perform well, without adding penalties to the *R8 Depression* regression.

Training a model on a naturalistic dataset like Psynary comes with several drawbacks. The most notable of these are bias caused by the skewed distributions of the data, information gaps caused by the selectiveness of prescribed medications and several layers of survivorship bias. Our dataset comes from patients treated in Australia, New Zealand, and Japan and from clinics with different population profiles. The clinic in Tokyo focuses on primary care of mostly foreign residents and make up the majority of the dataset (85.6%), while patients from Oceania are largely secondary care patients. There is also asymmetric representation of medication types. As there were only two psychiatrists prescribing medications for the participants in this study, the dataset is heavily reliant on the two clinicians' experience which could be a potential source of survivorship bias.

Our choices for model design also came with drawbacks. Most notably, the aggregation of medication types, the use of regression over explicit recommendation and the implementation of a recurrent model were decisions that compromised the model in specific ways. The prediction of a medication class is less useful for a clinician than predicting a specific medication, as patient response can vary significantly to medications of the same action type. Throughout a treatment, a patient may be prescribed several different medications of the class, which, as of now, our model cannot distinguish in any meaningful way. A possible solution for future work may be to use a finer category system, possible from learned clusters.

As our model is not an explicit recommender system, multi-medication recommendation becomes difficult, with significantly lower accuracy. The main reason for choosing *R8 Depression* regression was the strong biases in the data. A binary recommender system would mostly replicate the biases of the prescribing clinicians, rather than learning from the performance of medications on patients. This was another reason that medications were aggregated by class, eliminating the bias coming from preference to a specific
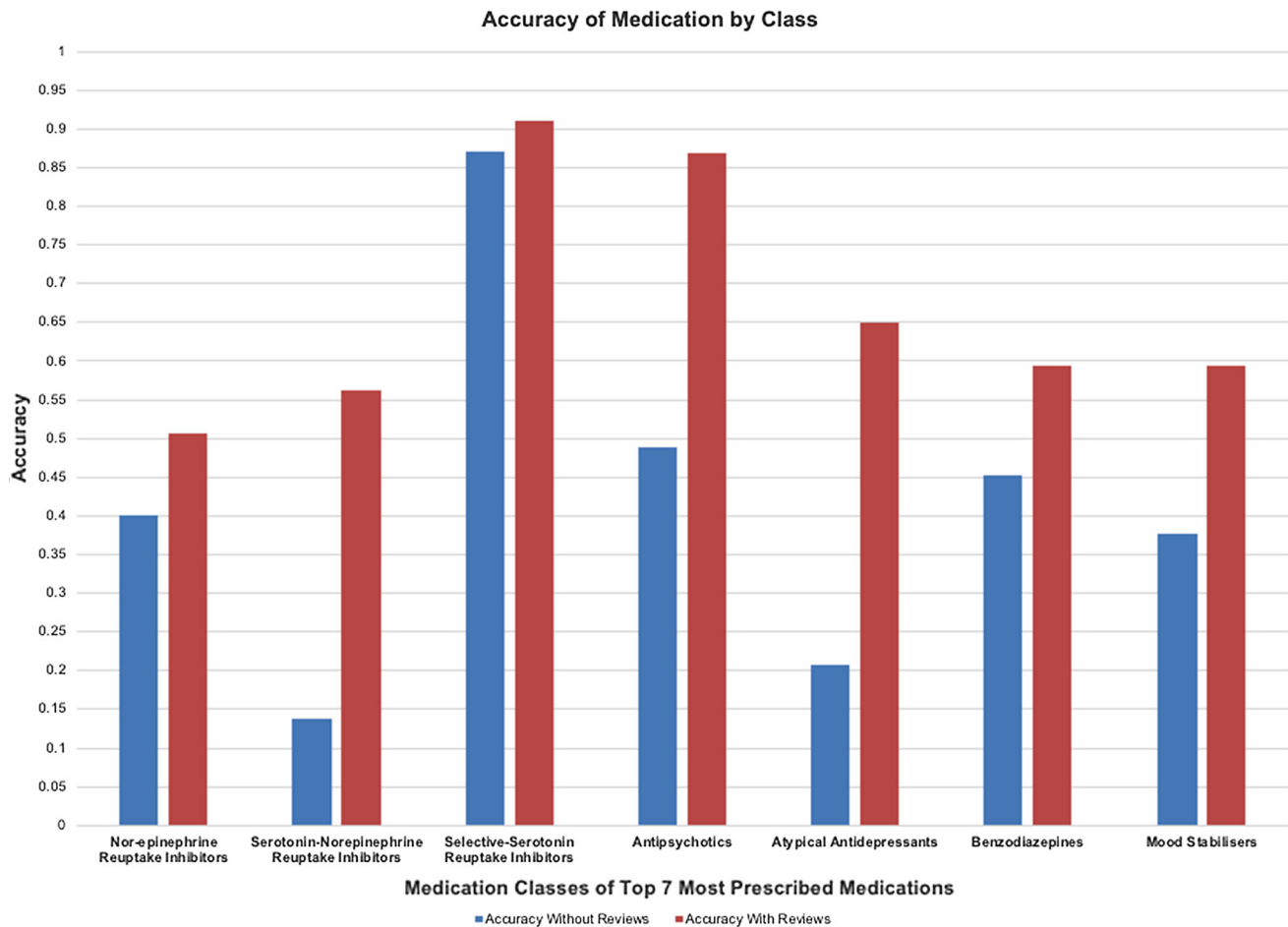
Fig. 7 Accuracy of model predicting by medication class with and without reviews

brand over another. Future work should focus on this specific point, possibly with a parallel network that uses the activation values of the regression network to produce specific recommendations.

The use of recurrent models in neural networks may be on the decline. Many sequence modelling objectives are now better solved using convolutional or attention networks, such as in the state-of-the-art audio generator WaveNet [121] and for translation tasks in transformer architecture [117]. These options have several advantages over recurrent networks, such as shorter gradient propagation distances, better performance with long sequences and easier training. Our choice of using an LSTM caused, in particular, difficulty in training the network, as the parallel fully connected and recurrent portions did not train at equal rates. Using a different sequence modelling architecture would, however, have cost added complexity and development time that we felt was best allocated to other parts of the model.

Finally, the training target is the best *R8 Depression* score reached during treatment in the presented model. Because of this, the model does not handle a relapse of

depression once a participant has reached *R8 Depression*-defined remission. In clinical practice, it is common for patients to relapse in depression and is evident in this study where 177 participants (38.6%) have been diagnosed with a form of recurrent depression (Table 2). Future studies will include updating this current model to reflect this clinical situation.

To evaluate the performance of the model under the intended conditions of treatment optimization, the model was evaluated in the binary decision problem of continuing the current medication or changing medication. We evaluated the model's accuracy both in the choice of changing medication and in the precision in recommending a new medication. The results are presented as a function of the number of reviews in Fig. 11.

After just 2 patient reviews, the model shows high recall (71.4%) in determining if a change in medication is needed. In the scenario of such a recommendation, the precision of the medication choice is similarly high (97.4%). To further validate the model, we then asked the model to select a different medication. The accuracy of changing to
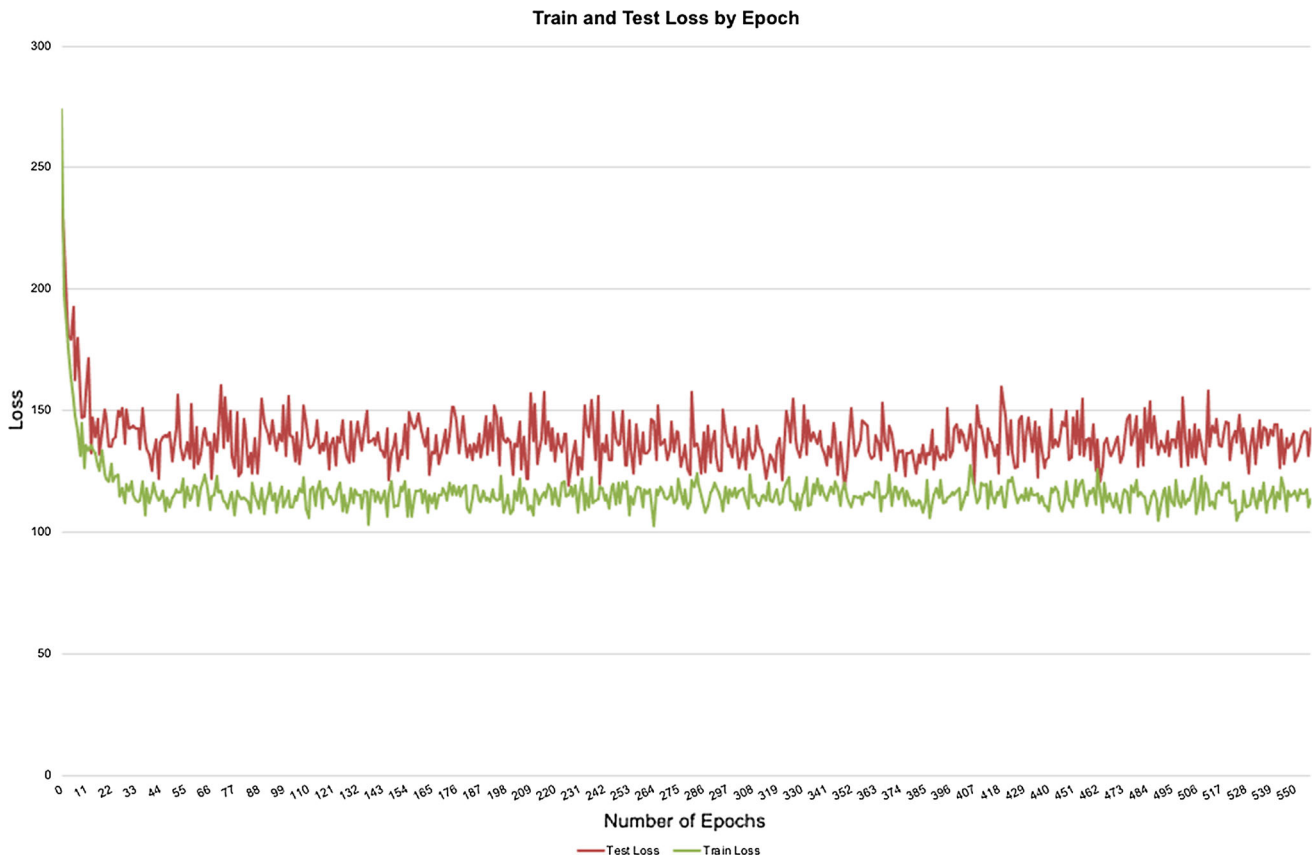
Train and Test Loss by Epoch

Fig. 8 The developed model's train/test loss per epoch

the correct medication was 54.3%. As illustrated in Fig. 10, these values are quite consistent over 40 patient reviews.

Taliaz et al. is the only published study similar to this one. Using genetic, clinical and demographic data from 1679 *STAR*D* participants, the team generated a hybrid, multi-step DL binary-response algorithm to predict whether a participant was either a "responder" or "nonresponder" for citalopram (SSRI), sertraline (SSRI) and venlafaxine (SNRI). This yielded an average balanced accuracy of 70.1% for the final test set, compared to a 46.8% initial response rate for participants with MDD. While they used a wide variety of data to arguably create a more visible clinical picture of the participant, our model was able to perform with an initial accuracy of 49.8% and accuracy of 76.5% after 2 patient reviews while accommodating for several different types of medication classes, subclasses and subtypes of depression.

Taliaz et al. also completed additional statistical calculations of the models performance. The sensitivity, specificity, positive predictive value and negative predictive value were 68.7%, 71.4%, 71.7% and 69%, respectively [113]. In comparison after only 2 reviews, our recall rate (also known as sensitivity) was 71.4%, similar to Taliaz et al., results and our precision rate (also known as positive predictive value) was 97.4% which far exceeded Taliaz et al., results.

From the 66 studies analysed by Gao, Calhoun and Sui, only 5 focussed on predicting the therapeutic response and all 5 solely focussed on MDD instead of all subtypes of depression [45]. In 2015, Korgaonkar et al. investigated objective brain volumetric measures of patients with MDD to reliably predict symptomatic remission with their initial antidepressant medication [96]. Their study found two decision trees that had high probability prediction scores of non-remission and were replicated [96]. These were 1) left middle frontal volume less than 14.8 mL and right angular gyrus volume greater than 6.3 mL which discerned 55% of non-remitters with an 85% accuracy; and 2) fractional anisotropy values in the left cingulum bundle less than 0.63, right superior fronto-occipital fasciculus less than 0.5 which discerned 15% of non-remitters with 84% accuracy [96]. Also in 2015, Williams et al. investigated whether amygdala activation stimulated by emotion was a general or differential predictor of response to escitalopram (SSRI), sertraline (SSRI) and venlafaxine (SNRI) using MRI [97]. Their model classified responders vs non-responders with an overall accuracy, cross-validation accuracy, sensitivity and specificity of 75%, 75%, 77% and 72%, respectively
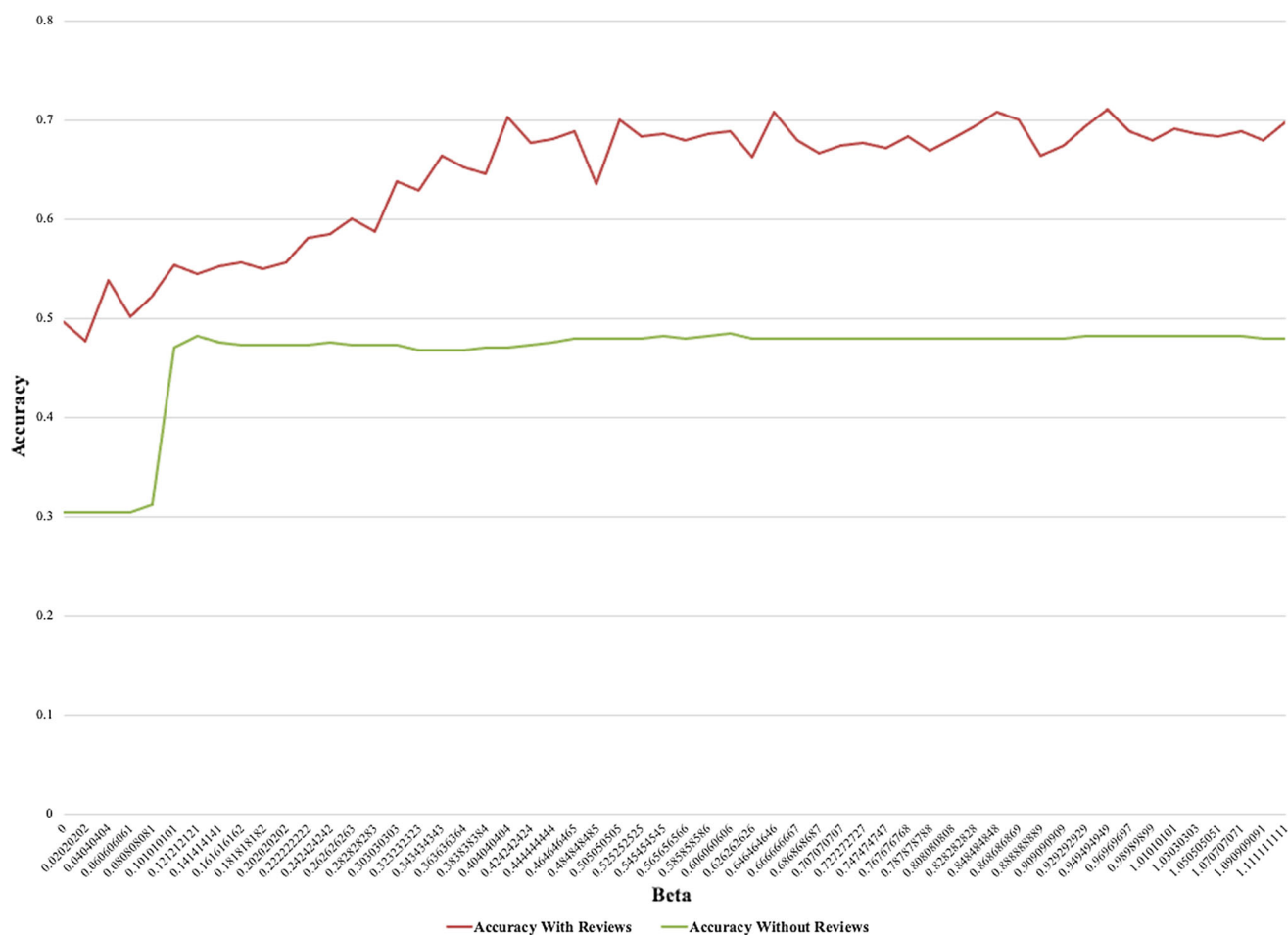
Fig. 9 Optimisation of the developed model's β coefficient by accuracy with and without reviews

[97]. Korgaonkar et al. examined whether diffusion tensor imaging measures of the anterior cingulate and limbic white matter are useful prognostic biomarkers for MDD [98]. They found that when adding age to a model that looked at the stria terminalis fractional anisotropy and the cingulate fractional anisotropy, their model had an overall accuracy of 74%, sensitivity of 74% and specificity of 75% [98]. Gong et al. investigated the diagnostic and prognostic potential of pre-treatment structural neuroanatomy using support vector machine (SVM) in patients with non-refractory depressive disorder or refractory depressive disorder (two subtypes of MDD) [99]. Sixty-one patients were prescribed either an SSRI, TCA or SNRI. The diagnostic accuracy, sensitivity and specificity when applying SVM to both grey and white matter images were 65.22% [99]. The prognostic accuracy based on both grey and white matter images resulted in an accuracy, sensitivity and specificity of 69.57% [99]. Finally, Costafreda et al. interestingly investigated the functional neuroanatomy of showing sad faces of different intensities to patients with acute MDD before cognitive behavioural therapy to predict clinical

response [100]. They found that prediction of remission from MDD was significant at the lowest and highest intensities of sadness. Both situations had a sensitivity of 71% and specificity of 86% [100]. Our optimisation assessment illustrates that our model is highly comparable to these 5 studies when comparing our sensitivity rate (71.4%). What is very important from our results and cannot be compared to these 5 studies are our precision rate of 97.4%, false positive rate of 2.6% and correct change accuracy of 54.3%. Additionally, when solely comparing model accuracies, our model has an accuracy of 76.5% after 2 patient reviews and performs better than all studies apart from Korgaonkar et al.

With depression prevalence rates increasing and placing pressure on existing services, new management techniques need to be considered to ensure all patients are treated, and in remission, as soon as possible [122]. Currently, the majority of patients with depression present to primary care [123] and secondary mental health services predominantly accommodate severe and/or high risk presentations, leaving the majority of people with depression unable to access
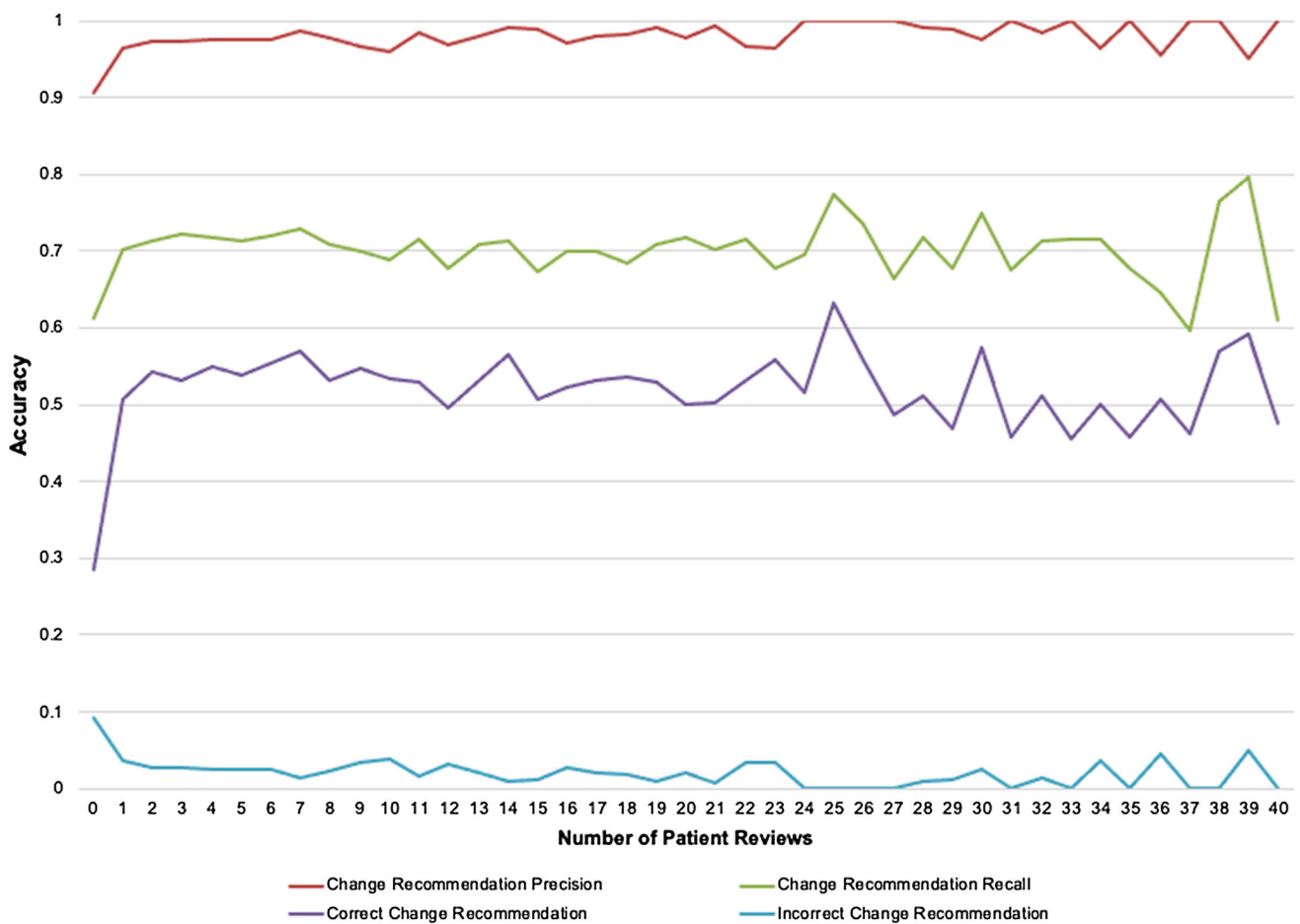
Fig. 10 Optimisation change measures in relation to number of patient reviews

specialist mental health services [123–126]. Introduction of telemedicine as an aspect of management for mental disorders covers two important factors of care; it improves patient access in areas where specialists are limited (including regional, rural and remote areas), and it improves disease control and relapse prevention [127]. Self-completed web-based systems ensure consistency of data collection and some patients are more likely to disclose relevant information on a self-completed assessment compared to clinician-interviewed settings [128, 129]. Furthermore, online systems provide the opportunity to collect results and track progress [128].

This study has illustrated the benefits of using Psynary, a web-based tool as an adjunct to normal care for depression. Unlike typical mental health care settings, the data is recorded in a quantitative format that lends itself to analysis. This study has also shown the clinical importance of the application of ML/DL algorithms to clinically collected data for the optimisation of depression treatment. To further validate the use of Psynary and the DL algorithm

created in this study, further work will involve diversifying the demographic of participants and clinicians, the locations of the clinics and implementing genomic data (single nucleotide polymorphisms) with a genomics team to create a multi-dimensional, robust model .

5 Conclusion

Depression has a major health and socio-economic impact, is continuing to increase in prevalence and mental health services are struggling with the increased demand . Using a combination of web-based tools and DL algorithms in a clinical setting, as outlined in this study, could lead to an increase in clinician accessibility and reduce time taken to reach optimal treatment protocols thereby reducing the prevalence of depression and its socio-economic impact on societies.
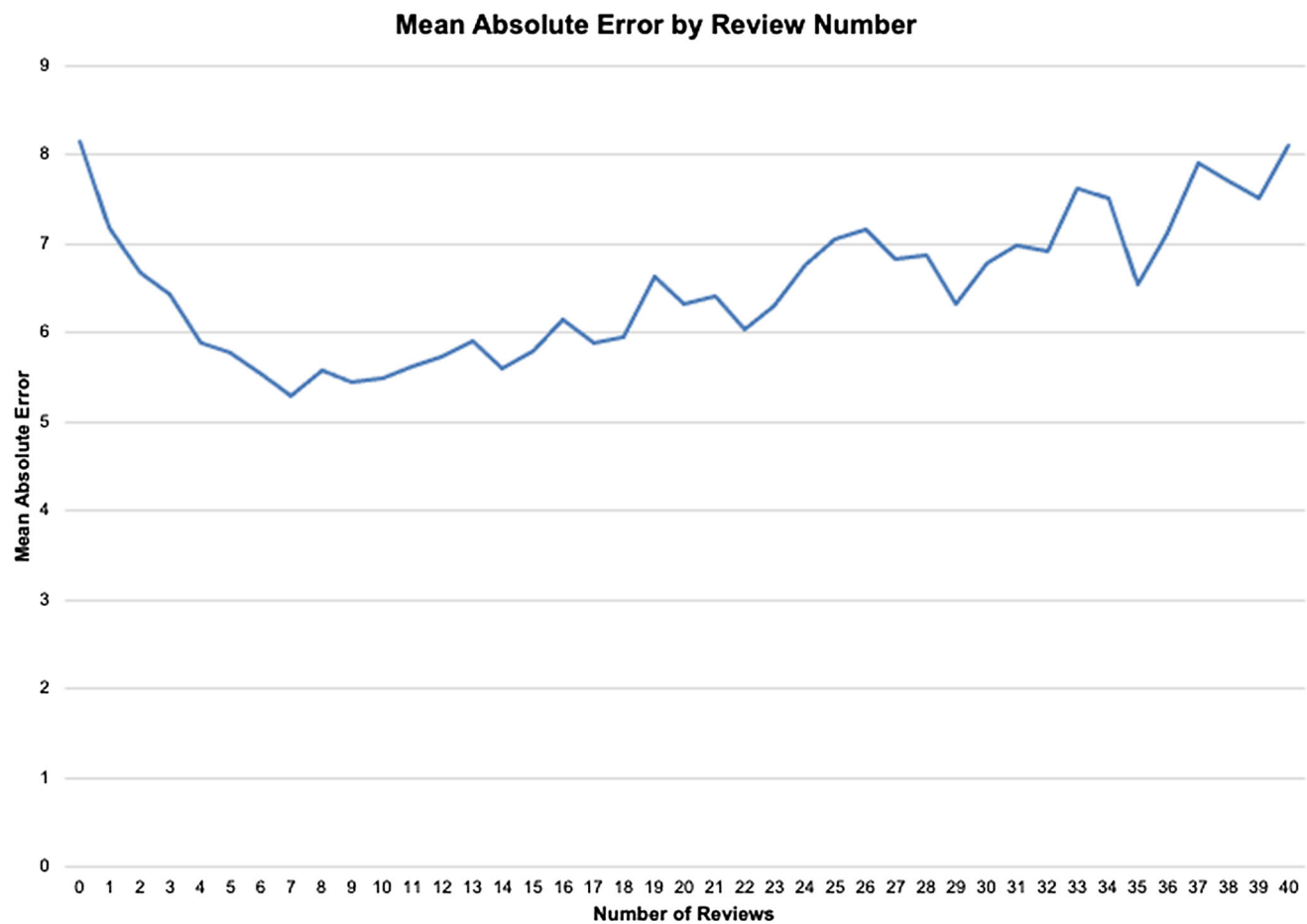
Mean Absolute Error by Review Number

Fig. 11 The developed model's mean absolute error per number of reviews

6 Data and material availability

All clinical data used in this article is fully anonymised from the point of capture. It is available on application to Dr Andrew Kissane for the purpose of validation, regulation and further research. The data and the web-based Psynary system remain the intellectual property of International Medical K.K.

Author's contribution Aidan Cousins and Lucas Nakano were involved in study design, implementation and analysis and drafting of the manuscript. Rasa Kabaila and Emma Schofield were involved in study design and expert review.

Code availability For review purposes, the code used is available on application to Dr Andrew Kissane.

Declarations

Conflict of interests The authors declare that they have no conflicts of interest.

Ethics approval This research was approved by the clinical research ethics committee of University of Otago (New Zealand, approval #: H16/014), Asai Hifuka Institutional Review Board (Japan, approval #: 承認番号: 20170724-1) and the North Coast NSW Human Research Ethics Committee (Australia, approval #:HREA271 2019/ ETH13489).

Consent to participate Written consent was obtained from all participants for use of de-identified data from Psynary for the purpose of analysing the naturalistic clinical outcomes.

Consent for publication All authors approve the final version of the manuscript and this submission for possible publication in *Neural Computing and Applications*.

References

1. World Health Organisation (2019) ICD-10 Version:2019. In: World Health Organisation. https://icd.who.int/browse10/2019/ en. Accessed 15 Apr 2021

2. Carvalho AF, Miskowiak KK, Hyphantis TN et al (2014) Cognitive dysfunction in depression–pathophysiology and novel targets. CNS & Neurol Disord-Drug Targets (Former Curr Drug Targets-CNS & Neurol Disord) 13:1819–1835

3. Parker G, Tavella G, Hadzi-Pavlovic D (2019) Identifying and differentiating melancholic depression in a non-clinical sample. J Affect Disord 243:194–200. https://doi.org/10.1016/j.jad.2018.09.024

4. Parker G, Bassett D, Outhred T et al (2017) Defining melancholia: a core mood disorder. Bipolar Disord 19:235–237. https://doi.org/10.1111/bdi.12501

5. Parker G (2017) Diagnosing melancholic depression: some personal observations. Australas Psychiatry 25:21–24. https://doi.org/10.1177/1039856216657696

6. Ferrari AJ, Charlson FJ, Norman RE et al (2013) Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 10:e1001547. https://doi.org/10.1371/journal.pmed.1001547

7. Liu Q, He H, Yang J et al (2020) Changes in the global burden of depression from 1990 to 2017: findings from the Global Burden of Disease study. J Psychiatr Res 126:134–140. https://doi.org/10.1016/j.jpsychires.2019.08.002

8. Australian Institute of Health and Welfare (2021) Mental Health Services in Australia. AIHW, Australian Government

9. Mohebbi M, Agustini B, Woods RL et al (2019) Prevalence of depressive symptoms and its associated factors among healthy community-dwelling older adults living in Australia and the United States. Int J Geriatr Psychiatry 34:1208–1216. https://doi.org/10.1002/gps.5119

10. Goldney RD, Eckert KA, Hawthorne G, Taylor AW (2010) Changes in the prevalence of major depression in an Australian community sample between 1998 and 2008. Aust N Z J Psychiatry 44:901–910. https://doi.org/10.3109/00048674.2010.490520

11. Brodaty H, Draper B, Saab D et al (2001) Psychosis, depression and behavioural disturbances in Sydney nursing home residents: prevalence and predictors. Int J Geriatr Psychiatry 16:504–512. https://doi.org/10.1002/gps.382

12. Pirkis J, Pfaff J, Williamson M et al (2009) The community prevalence of depression in older Australians. J Affect Disord 115:54–61. https://doi.org/10.1016/j.jad.2008.08.014

13. Doran CM, Kinchin I (2019) A review of the economic impact of mental illness. Aust Health Rev 43:43–48. https://doi.org/10.1071/AH16115

14. Smetanin P, Stiff D, Briante C et al (2011) The Life and Economic Impact of Major Mental Illnesses in Canada. Mental Health Commission of Canada, Ottawa, Canada

15. McCrone P, Dhanasiri S, Patel A et al (2008) Paying the price—the cost of mental health care in England to 2026. London, England

16. Hudson S, Trowland H, Russell L (2019) 2018 New Zealand Mental Health Monitor. Wellington, New Zealand

17. Gasteiger N, Vedhara K, Massey A et al (2021) Depression, anxiety and stress during the COVID-19 pandemic: results from a New Zealand cohort study on mental well-being. BMJ Open 11:e045325. https://doi.org/10.1136/bmjopen-2020-045325

18. MidCentral District Health Board (2005) Depression Service Plan. MidCentral District Health Board

19. Murray CJL, Lopez AD (1997) Alternative projections of mortality and disability by cause 1990–2020: global burden of disease study. Lancet 349:1498–1504. https://doi.org/10.1016/S0140-6736(96)07492-2

20. Nishi D, Ishikawa H, Kawakami N (2019) Prevalence of mental disorders and mental health service use in Japan. Psychiatry Clin Neurosci 73:458–465. https://doi.org/10.1111/pcn.12894

21. Ishikawa H, Tachimori H, Takeshima T et al (2018) Prevalence, treatment, and the correlates of common mental disorders in the mid 2010's in Japan: the results of the world mental health Japan 2nd survey. J Affect Disord 241:554–562. https://doi.org/10.1016/j.jad.2018.08.050

22. Sado M, Yamauchi K, Kawakami N et al (2011) Cost of depression among adults in Japan in 2005. Psychiatry Clin Neurosci 65:442–450. https://doi.org/10.1111/j.1440-1819.2011.02237.x

23. Okumura Y, Higuchi T (2011) Cost of depression among adults in Japan. Prim Care Companion CNS Disord. https://doi.org/10.4088/PCC.10m01082

24. Mokkink LB, Terwee CB, Patrick DL et al (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 19:539–549. https://doi.org/10.1007/s11136-010-9606-8

25. El-Den S, Chen TF, Gan Y-L et al (2018) The psychometric properties of depression screening tools in primary healthcare settings: a systematic review. J Affect Disord 225:503–522. https://doi.org/10.1016/j.jad.2017.08.060

26. Rush AJ, Fava M, Wisniewski SR et al (2004) Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. Control Clin Trials 25:119–142. https://doi.org/10.1016/s0197-2456(03)00112-0

27. Rush AJ, Trivedi MH, Wisniewski SR et al (2006) Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 163:1905–1917. https://doi.org/10.1176/ajp.2006.163.11.1905

28. Takao Y, Figueroa E, Berna KFJ et al (2021) Validation of a novel online depression symptom severity rating scale: the R8 Depression. Health Qual Life Outcomes 19:163. https://doi.org/10.1186/s12955-020-01654-z

29. Taniguchi M, Takao Y, Kissane A, Tranter R (2019) Measuring Early Treatment Response and Accelerating Treatment Optimization in the Treatment of Depression: Naturalistic Outcomes from OptiMA1 In: EPA European Congress of Psychiatry 2020

30. Graham S, Depp C, Lee EE et al (2019) Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep 21:116. https://doi.org/10.1007/s11920-019-1094-0

31. Su C, Xu Z, Pathak J, Wang F (2020) Deep learning in mental health outcome research: a scoping review. Transl Psychiatry 10:1–26. https://doi.org/10.1038/s41398-020-0780-3

32. Insel T, Cuthbert B, Garvey M et al (2010) Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 167:748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

33. Nelson B, McGorry PD, Wichers M et al (2017) Moving from static to dynamic models of the onset of mental disorder: a review. JAMA Psychiat 74:528–534. https://doi.org/10.1001/jamapsychiatry.2017.0001

34. Tung C, Lu W (2016) Analyzing depression tendency of web posts using an event-driven depression tendency warning model. Artif Intell Med 66:53–62. https://doi.org/10.1016/j.artmed.2015.10.003

35. Aldarwish MM, Ahmad HF (2017) Predicting depression levels using social media posts. In: 2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS). pp. 277–280

36. Nemesure MD, Heinz MV, Huang R, Jacobson NC (2021) Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Sci Rep 11:1980. https://doi.org/10.1038/s41598-021-81368-4

37. Dinga R, Marquand AF, Veltman DJ et al (2018) Predicting the naturalistic course of depression from a wide range of clinical,

psychological, and biological data: a machine learning approach. Transl Psychiatry 8:241. https://doi.org/10.1038/s41398-018-0289-1

38. Wahle F, Kowatsch T, Fleisch E et al (2016) Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 4:e111. https://doi.org/10.2196/mhealth.5960

39. Chekroud AM, Zotti RJ, Shehzad Z et al (2016) Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 3:243–250. https://doi.org/10.1016/S2215-0366(15)00471-X

40. Erguzel TT, Sayar GH, Tarhan N (2016) Artificial intelligence approach to classify unipolar and bipolar depressive disorders. Neural Comput Appl 27:1607–1616. https://doi.org/10.1007/s00521-015-1959-z

41. Chattopadhyay S (2017) A neuro-fuzzy approach for the diagnosis of depression. Appl Comput Inform 13:10–18. https://doi.org/10.1016/j.aci.2014.01.001

42. Vai B, Parenti L, Bollettini I et al (2020) Predicting differential diagnosis between bipolar and unipolar depression with multiple kernel learning on multimodal structural neuroimaging. Eur Neuropsychopharmacol 34:28–38. https://doi.org/10.1016/j.euroneuro.2020.03.008

43. Richter T, Fishbain B, Markus A et al (2020) Using machine learning-based analysis for behavioral differentiation between anxiety and depression. Sci Rep 10:16381. https://doi.org/10.1038/s41598-020-72289-9

44. Sharma A, Verbeke WJMI (2020) Improving diagnosis of depression with XGBOOST machine learning model and a large biomarkers Dutch dataset (n = 11,081). Front Big Data 3:15. https://doi.org/10.3389/fdata.2020.00015

45. Gao S, Calhoun VD, Sui J (2018) Machine learning in major depression: from classification to treatment outcome prediction. CNS Neurosci Ther 24:1037–1052. https://doi.org/10.1111/cns.13048

46. Patel MJ, Andreescu C, Price JC et al (2015) Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. Int J Geriatr Psychiatry 30:1056–1067. https://doi.org/10.1002/gps.4262

47. Jing B, Long Z, Liu H et al (2017) Identifying current and remitted major depressive disorder with the Hurst exponent: a comparative study on two automated anatomical labeling atlases. Oncotarget 8:90452–90464. https://doi.org/10.18632/oncotarget.19860

48. Yoshida K, Shimizu Y, Yoshimoto J et al (2017) Prediction of clinical depression scores and detection of changes in whole-brain using resting-state functional MRI data with partial least squares regression. PLoS ONE 12:e0179638. https://doi.org/10.1371/journal.pone.0179638

49. Zhong X, Shi H, Ming Q et al (2017) Whole-brain resting-state functional connectivity identified major depressive disorder: a multivariate pattern analysis in two independent samples. J Affect Disord 218:346–352. https://doi.org/10.1016/j.jad.2017.04.040

50. Wang X, Ren Y, Zhang W (2017) Depression disorder classification of fMRI data using sparse low-rank functional brain network and graph-based features. Comput Math Methods Med 2017:3609821. https://doi.org/10.1155/2017/3609821

51. Schnyer DM, Clasen PC, Gonzalez C, Beevers CG (2017) Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor MRI measures in individuals with major depressive disorder. Psychiatry Res Neuroimaging 264:1–9. https://doi.org/10.1016/j.pscychresns.2017.03.003

52. Sundermann B, Feder S, Wersching H et al (2017) Diagnostic classification of unipolar depression based on resting-state functional connectivity MRI: effects of generalization to a diverse sample. J Neural Transm 124:589–605. https://doi.org/10.1007/s00702-016-1673-8

53. Bhaumik R, Jenkins LM, Gowins JR et al (2017) Multivariate pattern analysis strategies in detection of remitted major depressive disorder using resting state functional connectivity. Neuroimage Clin 16:390–398. https://doi.org/10.1016/j.nicl.2016.02.018

54. Ramasubbu R, Brown MRG, Cortese F et al (2016) Accuracy of automated classification of major depressive disorder as a function of symptom severity. Neuroimage Clin 12:320–331. https://doi.org/10.1016/j.nicl.2016.07.012

55. Sacchet MD, Prasad G, Foland-Ross LC et al (2015) Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory. Front Psychiatry 6:21. https://doi.org/10.3389/fpsyt.2015.00021

56. Sato JR, Moll J, Green S et al (2015) Machine learning algorithm accurately detects fMRI signature of vulnerability to major depression. Psychiatry Res 233:289–291. https://doi.org/10.1016/j.pscychresns.2015.07.001

57. Johnston BA, Steele JD, Tolomeo S et al (2015) Structural MRI-Based predictions in patients with treatment-refractory depression (TRD). PLoS ONE 10:e0132958. https://doi.org/10.1371/journal.pone.0132958

58. Shimizu Y, Yoshimoto J, Toki S et al (2015) Toward probabilistic diagnosis and understanding of depression based on functional MRI data analysis with logistic group LASSO. PLoS ONE 10:e0123524. https://doi.org/10.1371/journal.pone.0123524

59. Rosa MJ, Portugal L, Hahn T et al (2015) Sparse network-based models for patient classification using fMRI. Neuroimage 105:493–506. https://doi.org/10.1016/j.neuroimage.2014.11.021

60. Cao L, Guo S, Xue Z et al (2014) Aberrant functional connectivity for diagnosis of major depressive disorder: a discriminant analysis. Psychiatry Clin Neurosci 68:110–119. https://doi.org/10.1111/pcn.12106

61. Zeng L-L, Shen H, Liu L, Hu D (2014) Unsupervised classification of major depression using functional connectivity MRI. Hum Brain Mapp 35:1630–1641. https://doi.org/10.1002/hbm.22278

62. Habes I, Krall SC, Johnston SJ et al (2013) Pattern classification of valence in depression. Neuroimage Clin 2:675–683. https://doi.org/10.1016/j.nicl.2013.05.001

63. Wei M, Qin J, Yan R et al (2013) Identifying major depressive disorder using Hurst exponent of resting-state brain networks. Psychiatry Res 214:306–312. https://doi.org/10.1016/j.pscychresns.2013.09.008

64. Mwangi B, Ebmeier KP, Matthews K, Steele JD (2012) Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. Brain 135:1508–1521. https://doi.org/10.1093/brain/aws084

65. Mourão-Miranda J, Hardoon DR, Hahn T et al (2011) Patient classification as an outlier detection problem: an application of the one-class support vector machine. Neuroimage 58:793–804. https://doi.org/10.1016/j.neuroimage.2011.06.042

66. Schmaal L, Marquand AF, Rhebergen D et al (2015) Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: a multivariate pattern recognition study. Biol Psychiatry 78:278–286. https://doi.org/10.1016/j.biopsych.2014.11.018

67. Frangou S, Dima D, Jogia J (2017) Towards person-centered neuroimaging markers for resilience and vulnerability in bipolar disorder. Neuroimage 145:230–237. https://doi.org/10.1016/j.neuroimage.2016.08.066

68. Gao S, Osuch EA, Wammes M et al (2017) Discriminating bipolar disorder from major depression based on kernel SVM

using functional independent components. In: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP). ieeexplore.ieee.org, pp. 1–6

69. Jie N-F, Zhu M-H, Ma X-Y et al (2015) Discriminating bipolar disorder from major depression based on SVM-FoBa: efficient feature selection with multimodal brain imaging data. IEEE Trans Auton Ment Dev 7:320–331. https://doi.org/10.1109/TAMD.2015.2440298

70. Rubin-Falcone H, Zanderigo F, Thapa-Chhetry B et al (2018) Pattern recognition of magnetic resonance imaging-based gray matter volume measurements classifies bipolar disorder and major depressive disorder. J Affect Disord 227:498–505. https://doi.org/10.1016/j.jad.2017.11.043

71. Deng F, Wang Y, Huang H et al (2018) Abnormal segments of right uncinate fasciculus and left anterior thalamic radiation in major and bipolar depression. Prog Neuropsychopharmacol Biol Psychiatry 81:340–349. https://doi.org/10.1016/j.pnpbp.2017.09.006

72. Li M, Das T, Deng W et al (2017) Clinical utility of a short resting-state MRI scan in differentiating bipolar from unipolar depression. Acta Psychiatr Scand 136:288–299. https://doi.org/10.1111/acps.12752

73. He H, Sui J, Du Y et al (2017) Co-altered functional networks and brain structure in unmedicated patients with bipolar and major depressive disorders. Brain Struct Funct 222:4051–4064. https://doi.org/10.1007/s00429-017-1451-x

74. Bürger C, Redlich R, Grotegerd D et al (2017) Differential abnormal pattern of anterior cingulate gyrus activation in unipolar and bipolar depression: an fMRI and pattern classification approach. Neuropsychopharmacology 42:1399–1408. https://doi.org/10.1038/npp.2017.36

75. Rive MM, Redlich R, Schmaal L, Marquand AF (2016) Distinguishing medication-free subjects with unipolar disorder from subjects with bipolar disorder: state matters. Bipolar Disord. https://doi.org/10.1111/bdi.12446

76. Sacchet MD, Livermore EE, Iglesias JE et al (2015) Subcortical volumes differentiate major depressive disorder, bipolar disorder, and remitted major depressive disorder. J Psychiatr Res 68:91–98. https://doi.org/10.1016/j.jpsychires.2015.06.002

77. Fung G, Deng Y, Zhao Q et al (2015) Distinguishing bipolar and major depressive disorders by brain structural morphometry: a pilot study. BMC Psychiatry 15:298. https://doi.org/10.1186/s12888-015-0685-5

78. Redlich R, Almeida JR, Grotegerd D et al (2014) Brain morphometric biomarkers distinguishing unipolar and bipolar depression: a voxel-based morphometry-pattern classification approach. JAMA Psychiatry 71:1222–1230. https://doi.org/10.1001/jamapsychiatry.2014.1100

79. MacMaster FP, Carrey N, Langevin LM et al (2014) Disorder-specific volumetric brain difference in adolescent major depressive disorder and bipolar depression. Brain Imaging Behav 8:119–127. https://doi.org/10.1007/s11682-013-9264-x

80. Serpa MH, Ou Y, Schaufelberger MS et al (2014) Neuroanatomical classification in a population-based sample of psychotic major depression and bipolar I disorder with 1 year of diagnostic stability. Biomed Res Int 2014:706157. https://doi.org/10.1155/2014/706157

81. Grotegerd D, Stuhrmann A, Kugel H (2014) Amygdala excitability to subliminally presented emotional faces distinguishes unipolar and bipolar depression: an fMRI and pattern classification study. Hum Brain Mapp. https://doi.org/10.1002/hbm.22380

82. Zeng L-L, Shen H, Liu L et al (2012) Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. Brain 135:1498–1507. https://doi.org/10.1093/brain/aws059

83. Hilbert K, Lueken U, Muehlhan M, Beesdo-Baum K (2017) Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: a multimodal machine learning study. Brain Behav 7:e00633. https://doi.org/10.1002/brb3.633

84. Koutsouleris N, Meisenzahl EM, Borgwardt S et al (2015) Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. Brain 138:2059–2073. https://doi.org/10.1093/brain/awv111

85. Yu Y, Shen H, Zeng L-L et al (2013) Convergent and divergent functional connectivity patterns in schizophrenia and depression. PLoS ONE 8:e68250. https://doi.org/10.1371/journal.pone.0068250

86. Yang W, Chen Q, Liu P et al (2016) Abnormal brain activation during directed forgetting of negative memory in depressed patients. J Affect Disord 190:880–888. https://doi.org/10.1016/j.jad.2015.05.034

87. Johnston BA, Tolomeo S, Gradin V, Christmas D (2015) Failure of hippocampal deactivation during loss events in treatment-resistant depression. Brain. https://doi.org/10.1093/brain/awv177

88. Guo H, Cheng C, Cao X et al (2014) Resting-state functional connectivity abnormalities in first-onset unmedicated depression. Neural Regen Res 9:153–163. https://doi.org/10.4103/1673-5374.125344

89. Modinos G, Mechelli A, Pettersson-Yeo W et al (2013) Pattern classification of brain activation during emotional processing in subclinical depression: psychosis proneness as potential confounding factor. PeerJ 1:e42. https://doi.org/10.7717/peerj.42

90. Ma Z, Li R, Yu J et al (2013) Alterations in regional homogeneity of spontaneous brain activity in late-life subthreshold depression. PLoS ONE 8:e53148. https://doi.org/10.1371/journal.pone.0053148

91. Grotegerd D, Suslow T, Bauer J et al (2013) Discriminating unipolar and bipolar depression by means of fMRI and pattern classification: a pilot study. Eur Arch Psychiatry Clin Neurosci 263:119–131. https://doi.org/10.1007/s00406-012-0329-4

92. Fang P, Zeng L-L, Shen H et al (2012) Increased cortical-limbic anatomical network connectivity in major depression revealed by diffusion tensor imaging. PLoS ONE 7:e45972. https://doi.org/10.1371/journal.pone.0045972

93. Lord A, Horn D, Breakspear M, Walter M (2012) Changes in community structure of resting state functional connectivity in unipolar depression. PLoS ONE 7:e41282. https://doi.org/10.1371/journal.pone.0041282

94. Lythe KE, Moll J, Gethin JA et al (2015) Self-blame–selective hyperconnectivity between anterior temporal and subgenual cortices and prediction of recurrent depressive episodes. JAMA Psychiatry 72:1119–1126. https://doi.org/10.1001/jamapsychiatry.2015.1813

95. Liu F, Guo W, Yu D et al (2012) Classification of different therapeutic responses of major depressive disorder with multivariate pattern analysis method based on structural MR scans. PLoS ONE 7:e40968. https://doi.org/10.1371/journal.pone.0040968

96. Korgaonkar MS, Rekshan W, Gordon E et al (2015) Magnetic resonance imaging measures of brain structure to predict antidepressant treatment outcome in major depressive disorder. EBioMedicine 2:37–45. https://doi.org/10.1016/j.ebiom.2014.12.002

97. Williams LM, Korgaonkar MS, Song YC et al (2015) Amygdala reactivity to emotional faces in the prediction of general and medication-specific responses to antidepressant treatment in the randomized iSPOT-D trial. Neuropsychopharmacology 40:2398–2408. https://doi.org/10.1038/npp.2015.89

98. Korgaonkar MS, Williams LM, Song YJ et al (2014) Diffusion tensor imaging predictors of treatment outcomes in major depressive disorder. Br J Psychiatry 205:321–328. https://doi.org/10.1192/bjp.bp.113.140376

99. Gong Q, Wu Q, Scarpazza C et al (2011) Prognostic prediction of therapeutic response in depression using high-field MR imaging. Neuroimage 55:1497–1503. https://doi.org/10.1016/j.neuroimage.2010.11.079

100. Costafreda SG, Khanna A, Mourao-Miranda J, Fu CHY (2009) Neural correlates of sad faces predict clinical remission to cognitive behavioural therapy in depression. NeuroReport 20:637–641. https://doi.org/10.1097/WNR.0b013e3283294159

101. Drysdale AT, Grosenick L, Downar J et al (2017) Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat Med 23:28–38. https://doi.org/10.1038/nm.4246

102. Sankar A, Zhang T, Gaonkar B et al (2016) Diagnostic potential of structural neuroimaging for depression from a multi-ethnic community sample. BJPsych Open 2:247–254. https://doi.org/10.1192/bjpo.bp.115.002493

103. Foland-Ross LC, Sacchet MD, Prasad G et al (2015) Cortical thickness predicts the first onset of major depression in adolescence. Int J Dev Neurosci 46:125–131. https://doi.org/10.1016/j.ijdevneu.2015.07.007

104. Rondina JM, Hahn T, de Oliveira L et al (2014) SCoRS—A Method Based on Stability for Feature Selection and Mapping in Neuroimaging. IEEE Trans Med Imaging 33:85–98. https://doi.org/10.1109/TMI.2013.2281398

105. Hahn T, Marquand AF, Ehlis A-C et al (2011) Integrating neurobiological markers of depression. Arch Gen Psychiatry 68:361–368. https://doi.org/10.1001/archgenpsychiatry.2010.178

106. Nouretdinov I, Costafreda SG, Gammerman A et al (2011) Machine learning classification with confidence: application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression. Neuroimage 56:809–813. https://doi.org/10.1016/j.neuroimage.2010.05.023

107. Costafreda SG, Chu C, Ashburner J, Fu CHY (2009) Prognostic and diagnostic potential of the structural neuroanatomy of depression. PLoS ONE 4:e6353. https://doi.org/10.1371/journal.pone.0006353

108. Fu CHY, Mourao-Miranda J, Costafreda SG et al (2008) Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. Biol Psychiatry 63:656–662. https://doi.org/10.1016/j.biopsych.2007.08.020

109. Van Waarde JA, Scholte HS, Van Oudheusden LJB et al (2015) A functional MRI marker may predict the outcome of electroconvulsive therapy in severe and treatment-resistant depression. Mol Psychiatry 20:609–614

110. Jiang R, Abbott CC, Jiang T et al (2018) SMRI biomarkers predict electroconvulsive treatment outcomes: accuracy with independent data sets. Neuropsychopharmacology 43:1078–1087. https://doi.org/10.1038/npp.2017.165

111. Redlich R, Opel N, Grotegerd D et al (2016) Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. JAMA Psychiatry 73:557–564. https://doi.org/10.1001/jamapsychiatry.2016.0316

112. de Nijs J, Burger TJ, Janssen RJ et al (2021) Individualized prediction of three- and six-year outcomes of psychosis in a longitudinal multicenter study: a machine learning approach. Npj Schizophr 7:1–11. https://doi.org/10.1038/s41537-021-00162-3

113. Taliaz D, Spinrad A, Barzilay R et al (2021) Optimizing prediction of response to antidepressant medications using machine learning and integrated genetic, clinical, and demographic data. Transl Psychiatry 11:1–9. https://doi.org/10.1038/s41398-021-01488-3

114. Smith DJ, Griffiths E, Kelly M et al (2011) Unrecognised bipolar disorder in primary care patients with depression. Br J Psychiatry 199:49–56. https://doi.org/10.1192/bjp.bp.110.083840

115. Spitzer RL, Kroenke K, Williams JBW, Löwe B (2006) A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 166:1092–1097. https://doi.org/10.1001/archinte.166.10.1092

116. Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is All you Need. In: Guyon I, Luxburg UV, Bengio S, et al (eds) Advances in Neural Information Processing Systems. Curran Associates, Inc

117. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

118. Kabaila R (2021) Implementation of an innovative nurse led service to support treatment for depression in primary care (OptiMA2). Aust Nurs Midwifery J 27:22–23

119. Higuchi T (2010) Major depressive disorder treatment guidelines in Japan. J Clin Psychiatry 71(Suppl E1):e05. https://doi.org/10.4088/JCP.9058se1c.05gry

120. Malhi GS, Bell E, Bassett D et al (2021) The 2020 Royal Australian and New Zealand college of psychiatrists clinical practice guidelines for mood disorders. Aust N Z J Psychiatry 55:7–117. https://doi.org/10.1177/0004867420979353

121. van den Oord A, Dieleman S, Zen H et al (2016) WaveNet: a generative model for raw audio. arXiv [cs.SD]

122. Krog MD, Nielsen MG, Le JV et al (2018) Barriers and facilitators to using a web-based tool for diagnosis and monitoring of patients with depression: a qualitative study among Danish general practitioners. BMC Health Serv Res 18:503. https://doi.org/10.1186/s12913-018-3309-1

123. Cuijpers P, Quero S, Dowrick C, Arroll B (2019) Psychological treatment of depression in primary care: recent developments. Curr Psychiatry Rep 21:129. https://doi.org/10.1007/s11920-019-1117-x

124. Bedirhan Üstün T, Sartorius N (1995) Mental illness in general health care: an international study. Geneva, Switzerland

125. Bijl RV, Ravelli A (2000) Psychiatric morbidity, service use, and need for care in the general population: results of The Netherlands mental health survey and incidence study. Am J Public Health 90:602–607. https://doi.org/10.2105/ajph.90.4.602

126. Bortolotti B, Menchetti M, Bellini F et al (2008) Psychological interventions for major depression in primary care: a meta-analytic review of randomized controlled trials. Gen Hosp Psychiatry 30:293–302. https://doi.org/10.1016/j.genhosppsych.2008.04.001

127. García-Lizana F, Muñoz-Mayorga I (2010) Telemedicine for depression: a systematic review. Perspect Psychiatr Care 46:119–126. https://doi.org/10.1111/j.1744-6163.2010.00247.x

128. Wade AG (2010) Use of the internet to assist in the treatment of depression and anxiety: a systematic review. Prim Care Companion J Clin Psychiatry. https://doi.org/10.4088/PCC.09r00876blu

129. Berger M, Wagner TH, Baker LC (2005) Internet use and stigmatized illness. Soc Sci Med 61:1821–1827. https://doi.org/10.1016/j.socscimed.2005.03.025

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