Neural Computing and Applications
https://doi.org/10.1007/s00521-021-06426-4

ORIGINAL ARTICLE

Deep learning for prediction of depressive symptoms in a large textual dataset

Md Zia Uddin[1] · Kim Kristoffer Dysthe[2] · Asbjørn Følstad[1] · Petter Bae Brandtzaeg[1,2]

Received: 18 March 2021 / Accepted: 17 August 2021

Abstract
Depression is a common illness worldwide with potentially severe implications. Early identification of depressive symptoms is a crucial first step towards assessment, intervention, and relapse prevention. With an increase in data sets with relevance for depression, and the advancement of machine learning, there is a potential to develop intelligent systems to detect symptoms of depression in written material. This work proposes an efficient approach using Long Short-Term Memory (LSTM)-based Recurrent Neural Network (RNN) to identify texts describing self-perceived symptoms of depression. The approach is applied on a large dataset from a public online information channel for young people in Norway. The dataset consists of youth's own text-based questions on this information channel. Features are then provided from a one-hot process on robust features extracted from the reflection of possible symptoms of depression pre-defined by medical and psychological experts. The features are better than conventional approaches, which are mostly based on the word frequencies (i.e., some topmost frequent words are chosen as features from the whole text dataset and applied to model the underlying events in any text message) rather than symptoms. Then, a deep learning approach is applied (i.e., RNN) to train the time-sequential features discriminating texts describing depression symptoms from posts with no such descriptions (non-depression posts). Finally, the trained RNN is used to automatically predict depression posts. The system is compared against conventional approaches where it achieved superior performance than others. The linear discriminant space clearly reveals the robustness of the features by generating better clustering than other traditional features. Besides, since the features are based on the possible symptoms of depression, the system may generate meaningful explanations of the decision from machine learning models using an explainable Artificial Intelligence (XAI) algorithm called Local Interpretable Model-Agnostic Explanations (LIME). The proposed depression symptom feature-based approach shows superior performance compared to the traditional general word frequency-based approaches where frequency of the features gets more importance than the specific symptoms of depression. Although the proposed approach is applied on a Norwegian dataset, a similar robust approach can be applied on other depression datasets developed in other languages with proper annotations and symptom-based feature extraction. Thus, the depression prediction approach can be adopted to contribute to develop better mental health care technologies such as intelligent chatbots.

Keywords Prediction · Depression · LSTM · RNN · Text · AI · XAI

1 Introduction

Depression, or depressive disorder, is a common disease. According to the World Health Organization (WHO), the number of people with depression was estimated at more than 300 million affected worldwide [1]. Depression may severely impact well-being and functioning at work, school, and family, and can even lead to self-harm. Adolescent depression is associated with mood disorders and severe mental illness in adult life [2, 3]. Nearly 0.8 million

✉ Md Zia Uddin
zia.uddin@sintef.no

Kim Kristoffer Dysthe
k.k.dysthe@medisin.uio.no

Asbjørn Følstad
asf@sintef.no

1 SINTEF Digital, Oslo, Norway

2 University of Oslo, Oslo, Norway

Published online: 27 August 2021

people die from suicide each year and suicide is the fourth leading cause of death in 15–19-year-olds, according to WHO [1]. Amongst the top major diseases causing disability or incapability, five are mental illnesses—depression being the most prominent of these [4]. Hence, the disease burden due to depression is vast. The prevalence of depression in the adult population is approximately 5% across cultures, and 20% in its milder forms (i.e., partial symptoms, mild depression, and probable depression) [5]. Among adults, those most at risk are within the middle-aged population. Also, the world-wide occurrence of depression is increasing, with a rise of 18% between 2005 and 2015. However, early professional intervention can improve mental symptoms (e.g., absence of self-confidence and rumination) and resolve somatic problems (e.g., gastrointestinal problems and sleeping disorders) in most of the cases [6, 7].

Early detection of depressive symptoms followed by assessment and treatment can considerably improve chances for curbing symptoms and the underlying disease; mitigate negative implications for well-being and health as well as personal, economic, and social life [7–10]. However, detection of depressive symptoms is challenging and resource demanding. Current approaches are mainly based on clinical interviews and questionnaire surveys by hospitals or agencies [11], where psychological evaluation tables are utilized to make predictions on mental disorder. This approach is mostly based on one-to-one questionnaires and can roughly diagnose the psychological disorder for depression.

An alternative approach to interview or questionnaire-based predictions of depression is the analysis of informal texts provided by users. Previous studies in clinical psychology have shown that the relationship between the user of a language (e.g., speaker or writer) and their text is meaningful and has potential for the future [12]. A recent study by Havigerová et al. indicate a potential for text-based detection of persons at risk for depression, using a sample of informal text written about a holiday [12]. Hence, online records and data are increasingly seen as a valuable data source in supporting health care with decision support. The approach to identify depression symptoms from informal texts is promising, as it allows for benefitting from recent advances in natural language processing and Artificial Intelligence (AI). AI applied for natural language processing employs linguistics and computing techniques to help machines to understand underlying phenomena such as sentiments or emotions from texts. In that case, the core intent is to analyse opinions, ideas, and thoughts via the assignment of polarities either negative or positive.

Previous work has found that automatic analysis of depression symptoms from texts can be applied in, for example, sentiment retrieval from suicide notes and detecting insulting or depressive words or sentences in conversations or blog posts [13–18]. However, there is still substantial untapped potential in research on extracting depressive symptoms from texts. Key challenges include portraying significant cues of depression from texts. Also, there is a substantial hurdle in detecting depression symptoms from short texts.

To contribute towards solving these challenges, we aim to develop an automatic algorithm for detecting depression symptoms in texts, using a text-based sample of young people seeking advice about self-perceived depressive symptoms. We believe our automatic detection approach, describing the problems of the users in natural language, can be a substantial contribution to this research field. Hence, the current study focuses on how symptoms of depression are manifested through text in natural language using AI.

To visualize sample data of different groups in different applications, Linear Discriminant Analysis (LDA) is a good tool for data visualization based on discriminations [19–22]. It works on grouping of samples of similar classes. It tries to find the directions where the classes are best separated by considering minimizing the within-class scatter while maximizing the between-class scatter. LDA has already been used in various practical applications such as facial emotion recognition and human activity recognition. LDA projects the sample data of different classes onto a lower-dimensional vector space. Thus, the ratios of the between-class scatter and the within-class scatter is maximized to achieve highest discrimination.

Deep neural network has been contributing a lot recently in enormous fields of research, especially in pattern recognition and AI [23–34]. Though it is more robust than typical neural networks, it however consists of two major disadvantages. The first disadvantage is overfitting problem most of the time. The last one is taking much time for modelling the underlying data. The first successful deep learning algorithm was deep belief network that consisted of Restricted Boltzmann Machines (RBMs) that made the training quite faster than other previous learning approaches. Later, convolutional neural networks (CNN) was proposed and got popular especially in image processing fields. It showed better discriminative power compared to other approaches. CNN also extracts features alongside training the data. It has some convolutional stacks to generate a progressive hierarchy of abstract features via convolution, pooling, tangent squashing, rectifier, and normalization [24]. CNN is mostly applied for image and video pattern analysis rather than temporal information decoding. Hence, it has not been adopted for time-sequential data analysis. Recurrent Neural Networks (RNNs) is however a better choice than CNN since it consists of better discriminative power over others in case of sequential data and pattern analysis [30]. Since the basic RNNs usually consist of vanishing gradient problem due to long-term dependencies when it handles high-dimensional and time-sequential data, Long Short-Term Memory (LSTM) was introduced in RNN to overcome it. Hence, this work utilizes the advantage of LSTM-based RNN to model different emotional states in text data.

Among different approaches to analyse physical and mental states of human being from different data sources, machine learning has been very widely used [35–41]. Since machine learning models are progressively being employed to do significant predictions in crucial contexts day by day, the demand of transparency rises in such contexts from the various stakeholders in AI industry [42]. The high risk in this regard is making and applying the AI decisions that are unjustifiable and lacks explanations of the models' behaviour. Hence, explanations of the output of a model are vital. For example, specialists in precision medicine fields need further information from the machine learning models than simple prediction for supporting their diagnosis. Such necessities may also arise in other fields as well, such as medical emergencies. Hence, focusing merely on the performances of the AI models, gradually makes the systems towards unacceptance in some cases. Therefore, current research has highlighted the importance of explainable Artificial Intelligence (XAI) for establishing trust in machine learning-based decisions through the explanations of the black-box models. Popular state-of-the-art explanation algorithms include Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), and layer-wise relevance propagation (LRP). From which, LIME is very light-weight and yet tries to generate quick and satisfactory post-hoc explanations. Hence, this work adopts LIME to see the explanations (i.e., importance of the features) once the decision is provided by the model.

1.1 Contribution

This work focuses on processing text data, features, and depression symptoms text recognition with the target of chatbot as a smart application. Figure 1 shows a schematic setup of a text-based depression symptoms text detection system in a smart application where a user provides a query in text format and a server processes the text to apply feature extraction and deep learning. Based on the results, the server can suggest further advices to the user. Figure 2 shows the basic architecture of the proposed system consisting of training and testing procedure for the classification of texts describing symptoms of depression. In the training part, text data from all the users is obtained and then the features are trained using RNN. In the testing part, features from a sample test are applied to the trained model to take the decision whether the user describes depression or not. LDA is applied to show the robustness of the proposed features compared to other traditional ones. Finally, we apply one of the most popular algorithms (i.e., LIME) for post-hoc, local, and meaningful explanations of the machine learning decision regarding the existence of a potential depression or not, in the text. The contribution of the paper can be summarized as bellow:

- A large dataset of text is obtained from a public Norwegian online information channel: *ung.no*.
- Novel features are extracted representing the possible symptoms of depression defined by the experts from medical and psychology domains.
- RNN is applied based on LSTM, attention, and dense layers for modelling the emotional states.
- The machine learning decisions are explained using a state-of-the-art XAI approach, LIME to see the importance of the features.

2 Data collection and processing

To reliably detect symptoms of mental health issues, the collection of data for the detection model is crucial. For instance, data from social media such as Facebook status updates does not seem to be sufficiently detailed to develop reliable models to decode emotional states from data [43]. For this work, we obtained a large text-based dataset from a public Norwegian information website: *ung.no*. At *ung.no,* youth have the opportunity to post questions anonymously in Norwegian about their various challenges and problems in their everyday life. In response, corresponding professional experts (e.g. doctors, psychologist, nurses etc.) provide answers and offer advise. These questions and answers are published online and publicly available for everyone. Prior to submitting a question on *ung.no*, young people pre-define and categorize the topic of their post. We focused herein on the category "Mental health and emotions". Even if the texts are relatively short, they typically describe the activating factors leading to the mental state and the ensuing symptoms and behaviour. First, a proportion of the texts describe depressive conditions already diagnosed by a health professional. Second, many of the texts describe the narrative and the ensuing symptoms, either asking if it could represent depression or suggesting depression as a possible diagnosis. We believe these texts to present self-perceived depressive symptoms. Previous research suggest that self-perceived mental states correspond well with later clinical diagnoses [44–46]. Last, some of the texts describe the narratives and the succeeding mental states without mentioning a possible depression.

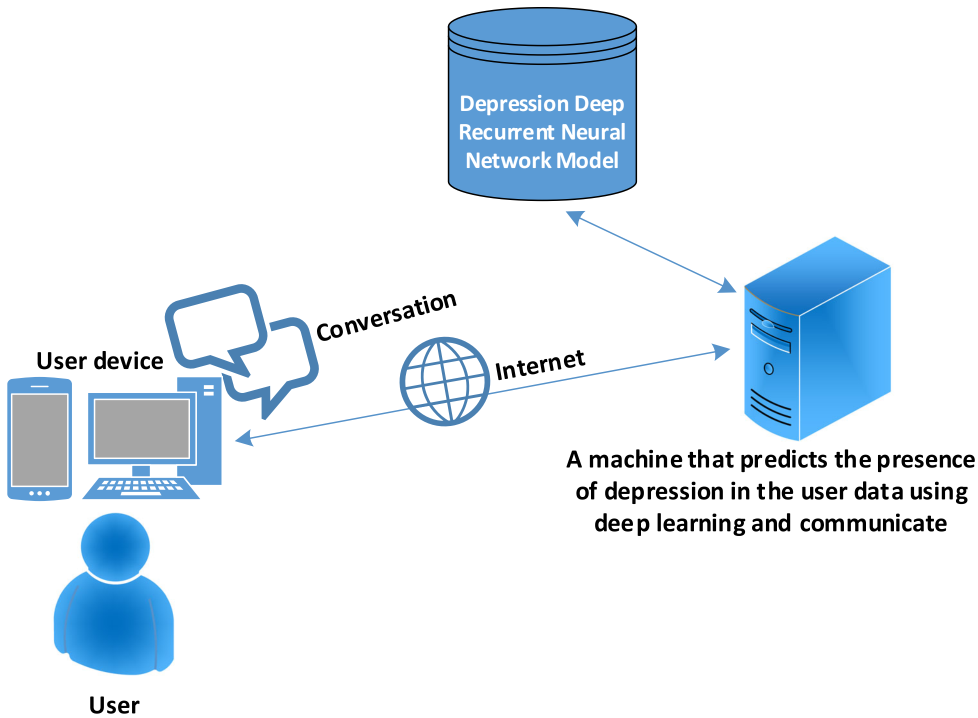
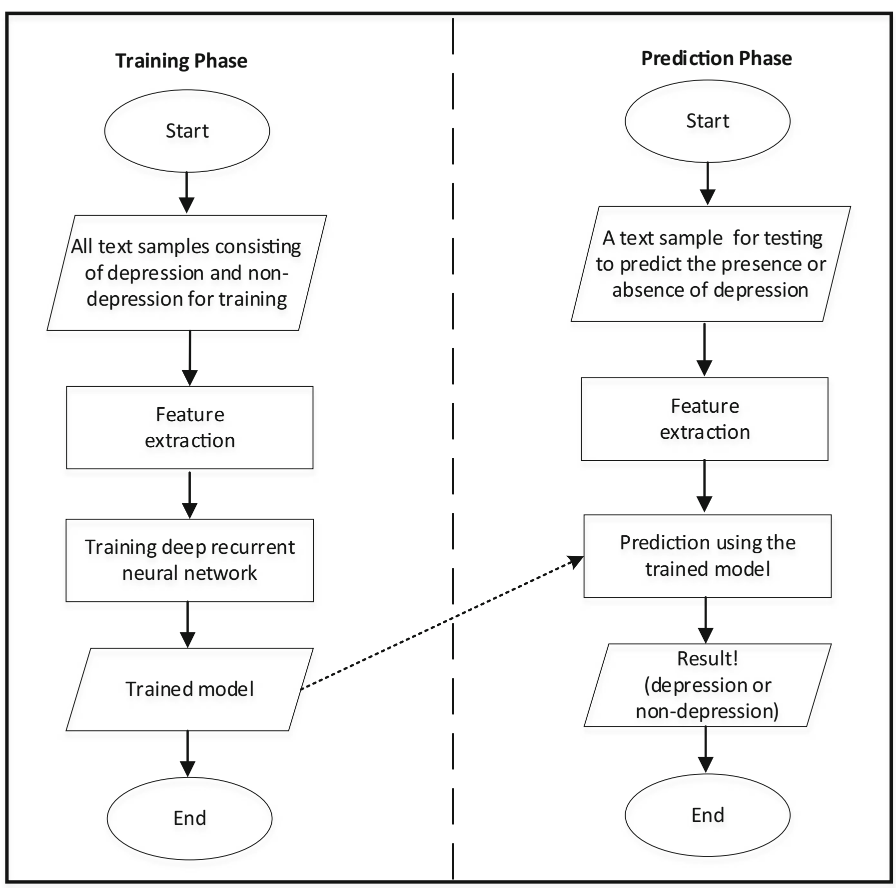


Fig. 1 A schematic setup for classifying texts containing symptoms of depression

Fig. 2 Flowcharts of the proposed depression prediction system from text



The staff interpret the texts as describing symptoms of depression. Accordingly, the data is classified into categories, depression being one of them. Then, a trained GP went through the posts, confirming descriptions of depressive symptoms. A list of sentences and words are summarized analysing the messages in the database where they may indicate the person having depression. A medical practitioner validated the sentences and words. Table 1 shows some important features from "Appendix" representing the possible sentences and/or words may occur in the queries by the youth having depression.

The sentences and words are used to obtain features for each message of the dataset. Five translated and paraphrased examples of depression texts derived from a Norwegian text dataset at *ung.no* are shown in Fig. 3. The Norwegian dataset consists of 277,552 free-text posts in different categories including depression texts. From that dataset, we utilized 11,807 and 21,470 posts of different length for our two different experiments in this work. For feature extraction process to model depression and non-depression machine learning model, we augment all the feature rows of "Appendix" first. Then, all rows in the "Appendix" are tokenized word by word and stemmed for feature extraction process. The stemmed words from the list of symptoms are represented as F = "*all hat meg alt er jæv … noe mer å lev for*". To extract features from a text input, one-hot process is applied on the stemmed words of the input text based on each word of F (i.e., 1 if a word from F is present and 0 otherwise). Thus, the features for the texts represents binary patterns to be applied with machine learning model of depression prediction. The collection of 189 unique words extracted from the list of possible symptoms is shown in Fig. 4 where the words in Norwegian are in alphabetical order in Fig. 4a and the corresponding translated words are in Fig. 4b. Unique extractions of stemmed words are listed to illustrate the diversity of possible words associated with symptoms of depression.

Table 1 Some important sentences and words used for depression in Norwegian and English (translated)

Symptoms (Norwegian)	Symptoms (Translated)
Ikke apetitt	No appetite
Avslutte livet	End life
Bryr meg ikke om noe	I do not care about anything
Ikke mer energi	No more energy
Suicid	Suicide
Gråt	Crying
Selvmordstanker	Suicidal thoughts
Ende livet mitt	End my life
Ta livet av meg	Take my life
Helt tom	Completely empty
Tristhet	Sadness
Alltid trett	Always tired
Umotivert	Unmotivated
Ingenting føles	Nothing feels
Ingenting interesserer meg	Nothing interests me
Ikke lenger konsentrasjon	No longer concentration
Har ikke matlyst	Have no appetite
Tenke negativt	Think negatively

The symptoms presented in "Appendix" are obtained with the help of Norwegian professionals (e.g., medical doctors and psychologists). However, the way of expressing the emotions in Norwegian texts may be linguistically different from other languages. Therefore, professionals in those languages can contribute to building dataset and features in modelling depression and non-depression. To be noted, the English texts are shown in Table 1, Figs. 3, 4, and "Appendix" only for the readability of international readers and researchers. Otherwise, whole approaches from input text to emotional state modelling via feature processing, is done based on the Norwegian language.

The main reason to go for using one-hot on the robust depression features rather than traditional ones such as typical one-hot and Term Frequency—Inverse Document Frequency (TF-IDF) [47] that are related to typical word frequencies rather than word importance is, the features describing depression symptoms are much more important than just word frequencies to predict depression in the text. Figure 5 shows the algorithm for one-hot feature extraction based on the unique feature words in the list of depression symptoms. Thus, the one-hot binary features based on the depression symptoms for the ith text in the dataset can be represented as L_i.

3 Linear discriminant analysis (LDA) for visualization

To visualize different features, we adopt linear discriminant analysis (LDA) here. LDA is basically an eigenvalue decomposition problem trying to maximize the inter-class scatterings of the samples whereas minimizing the inner-class scatterings of them. The formulas for the inter-class scattering, M_B and inner-class scattering matrix, M_W are shown as follows:

$$M_B = \sum_{i=1}^{c} N_i \left(\widetilde{m_i} - m_j\right)\left(\widetilde{m_i} - m_j\right)^T \quad (1)$$

1. I've been so busy lately, and not felt any good. My doctor did not seem to bother. Schoolwork can be quite enjoyable, but now I have lost motivation to do anything; spend my days thinking about what I need to do, but it's challenging to get started. I'm not happy anyway, so I often think it would be better not to live. I kind of have no feelings. What's happening to me?
2. I feel mentally exhausted and struggle to get through everyday life. I cry every day. Getting up in the morning feels like a struggle. I know I have to get through it, but it troubles me a lot, both mentally and physically. If I have to do something after work, I must sleep to be able to manage. Motivation is gone. Work is no longer enjoyable. Some days I sleep a lot, other days, nothing. What shall I do?
3. I can't be alone anymore without crying, and I cannot find joy in the things I usually found joyful. After work, I have no energy left for my work-out sessions. I see no point in living, except that I will hurt my family if I took my own life. I have so many negative thoughts about myself. Struggling to get psychological help even though I know I need to. Money is scarce, and I can't leave work to visit a psychologist. I think the economy plays only a small role here. I don't know what's causing it. How can I get better?
4. My friends leave me, and I feel so lonely. Easy things, like going to the store or picking up mail or just brushing my teeth, are tough to do. During the last month, I have lost weight. In periods I eat almost nothing. I feel empty inside. I usually eat and throw it up, and I don't like things I found pleasant before anymore. At school, I can no longer concentrate. Please, I can't live like this. I have considered taking my own life; take an overdose of something. Life is not worth living.
5. I have experienced bad things growing up. Living with my dad, who had drug problems, then moved to stay with my mom, who struggles with mental disorders. I have lost all motivation for school or work. I feel tired, sleep all day, and I eat almost nothing. I have dropped out of school two years in a row, lost all hope for the future. I have never gone to a psychologist to talk about my problems. I get very easily stressed. I hate to be among people. I just sit inside thinking every day. What is this?

Fig. 3 Five translated and paraphrased examples of depression posts derived from the Norwegian dataset used in the work

$$M_W = \sum_{i=1}^{c} \sum_{m_k \in C_i} (m_k - \widetilde{m_i})(m_k - \widetilde{m_i})^T \tag{2}$$

where c is the total number of classes, N_i the number samples in class C_i, m_k the feature vectors from class C, m_i the mean of class i, and m_j the mean of all feature vectors. The LDA feature space representing the optimal discrimination matrix can be found by maximizing the ratio of the determinant of M_B and M_W as

$$Q_{opt} = \frac{|Q^T M_B Q|}{|Q^T M_W Q|} \tag{3}$$

where Q basically represents the set of discriminant vectors. Thus, the discriminant ratio of inner as well as inter-class samples of different classes can be found by solving an eigenvalue problem as

$$M_B Q = \Lambda M_W Q \tag{4}$$

where Λ is the eigenvalue matrix in the singular value decomposition process. Figures 6, 7, 8, and 9 show the feature visualizations using 3-D plots of typical one-hot in LDA, typical TF-IDF in LDA, and proposed features in PCA, and proposed features in LDA features spaces, respectively. In the figures, the proposed features (i.e., Fig. 9) shows superior clustering of the samples of same class and better separation among the samples of different classes compare to the two other approaches, indicating the robustness of the proposed features in this regard. However, the traditional PCA projection on the features Thus, the text feature matrix F is projected to the LDA feature space Q_{opt} as

$$U = L Q_{\text{opt}}^T \tag{5}$$

aldri, all, alltid, alt, apetitt, av, avslutt, bar, bli, blitt, bort, brydd, bryr, bunn, burd, de, demotiver, denn, depp, depremer, depresjon, deprimer, deprisjon, det, distanser, distenser, dø, død, dødd, dør, eget, en, end, energi, energiløs, energinivå, er, et, for, forferd, forter, fra, få, får, føl, gjør, glad, gled, god, grin, grusomt, gråt, gå, går, har, hat, hatt, hel, helt, hull, håp, håpløs, håpløst, ikk, ill, indr, ing, ingenting, initiativ, inn, inni, interess, interesser, jeg, jæv, kan, klar, konsentrasjon, konsentrer, konstant, langt, lavt, lei, leng, lev, lik, likegyld, lit, liv, lykk, lyst, lås, mat, matlyst, med, meg, mening, meningsløs, mennesk, mer, min, mist, mitt, morsomt, motivasjon, mørk, mørkest, mørkt, ned, nedenfor, nedfor, nedstemt, negativ, nervøs, noe, noen, nok, nokk, nytteløst, og, om, oppgitt, overskudd, person, psykisk, på, rikt, savn, seg, selv, selvmord, selvmordstank, ser, sinnyskt, skyen, skyv, slit, sliten, slutt, smak, som, sosial, sov, sovn, sted, steng, stengt, suicid, suisid, så, søvn, ta, tank, tap, tapp, tar, tenk, tid, til, ting, tom, trett, trist, tro, trøtt, tår, ubetyd, ubruk, ukonsentrer, umotiver, uro, usosial, ut, utbrent, utslitt, uuthold, vansk, var, vekk, veld, venn, verden, verdiløs, vil, vill, vond, vær, å, ønsk

(a)

never, all, always, all, appetite, of, quit, bar, become, become, gone, bother, care, bottom, burd, the, demotives, this, depress, depress, depression, depress, depression, the, distances, distances, die, dead, dead, die, own, one, than, energy, energyless, energy level, is, one, for, horror, forts, from, get, get, feel, make, happy, rejoiced, good, laugh, cruel, cry, go, go, have, hate, had, had, whole, whole, hole, hope, hopeless, hopeless, not, ill, indr, ing, nothing, initiative, in, inside, interest, interests, I, damn, can, clear, concentration, concentrate, constant, long, low, bored, long, live, equal, indifferent, trust, life, happiness, desire, lock, food, appetite, with, me, meaning, meaningless, human, more, my, lost, my, funny, motivation, dark, darkest, dark, down, below, down, downhearted, negative, nervous, something, someone, enough, enough, useless, and, if, given, surplus, person, mental, on, rich, miss, oneself, suicide, suicidal idea, look, insane, cloud, push, toil, tired, end, taste, as, social, slept, sleep, place , close, closed, suicid, suisid, so, sleep, take, tank, loss, tap, take, think, time, to, things, empty, tired, sad, believe, tired, tear, insignificant, unused, unconcentrated, unmotivated , unrest, unsocial, out, burnt out, worn out, impatience, difficult, was, away, well, friend, world, worthless, will, wild, hurt, be, to, desire

(b)

Fig. 4 Unique words extracted from the stemmed words of possible symptoms reported in APENDIX A: **a** Norwegian words in alphabetical order and **b** translated in English

4 Deep recurrent neural network (RNN) for modelling emotional states

Emotional states can be represented as time-sequential words in text data while conversating with others. Hence, a machine learning model capable of encoding time-sequential data is quite suitable for such kind of work. Hence, Recurrent Neural Networks (RNNs) is adopted in this work. RNN can be considered as most popular deep learning approaches used to model time-sequential information [22]. RNNs basically consists of recurrent connections between history to present state and hidden states. That is a quite important role of the memory in neural networks. The usual RNN algorithms very often face a vanishing gradient problem, a limitation of processing long-term data which is mostly known as Long-Term Dependencies. To overcome the problem, Long Short-Term Memory (LSTM) was developed [23]. Figure 10 shows a sample deep neural network consists of 50 LSTM units.

Each LSTM memory block has a cell state as well as three gates, which are input, forget, and the output gates. The input gate F_t can be represented as

$$I_t = \beta(W_{LI}L_t + W_{HL}H_{t-1} + b_I) \quad (6)$$

where W is weight matrix, b bias vectors, and β a logistic function. The forget gate F can be expressed as

$$F_t = \beta(W_{LF}L_t + W_{HF}H_{t-1} + b_F). \quad (7)$$

The long-term memory is stored in a cell state vector S that is expressed as

$$S_t = F_tS_{t-1} + F_t \tanh(W_{LS}L_t + W_{HS}H_{t-1} + b_S). \quad (8)$$

```
Algorithm1: Feature_Extraction(Text)
1.Begin
2. k = 1
3. N = Number of Feature Sentences
4.  for i = 1 to N:
5.    Feat = Obtain Feature_Sentence
6.    Tokenized_Feat = Seperate Words from Feat
7.     For Each Word in Tokenized_Feat:
8.        If Word in Text Then
9.          L[k]=1
10.       Else
11.         L[k]=0
12.       End If
12.        k = k+1
13.   End For
14. Return L
6.End
```

Fig. 5 The algorithm of one-hot depression symptom feature extraction

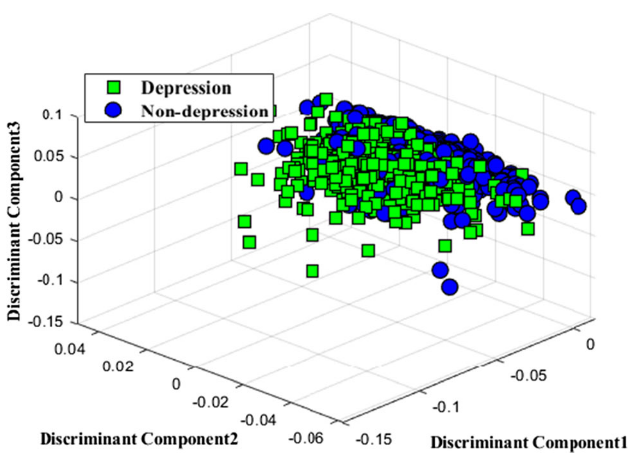


Fig. 6. 3-D plot after LDA on the traditional one-hot features of two emotional states

The output gate V produces the output for the unit and can be expressed as

$$V_t = \beta(W_{LV}L_t + W_{HV}H_{t-1} + b_V). \tag{9}$$

The hidden state H is expressed as

$$H_t = V_t \tanh(S_t). \tag{10}$$

We adopt an attention layer over the LSTM units before applying dense layer [48] as

$$A(att)_t = LSTM\left(H_t, A(att)_{t-1}\right) \tag{11}$$

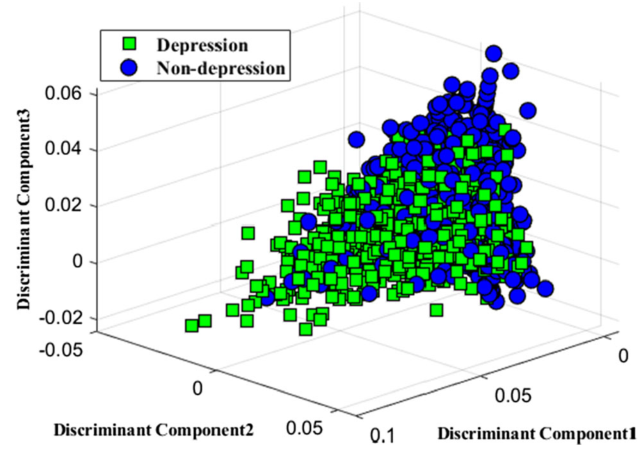


Fig. 7. 3-D plot after LDA on the traditional TF-IDF features of two emotional states

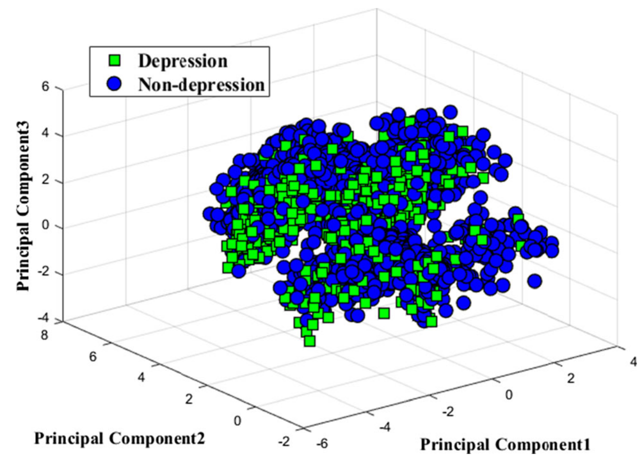


Fig. 8. 3-D plot after PCA on the proposed robust features of two emotional states

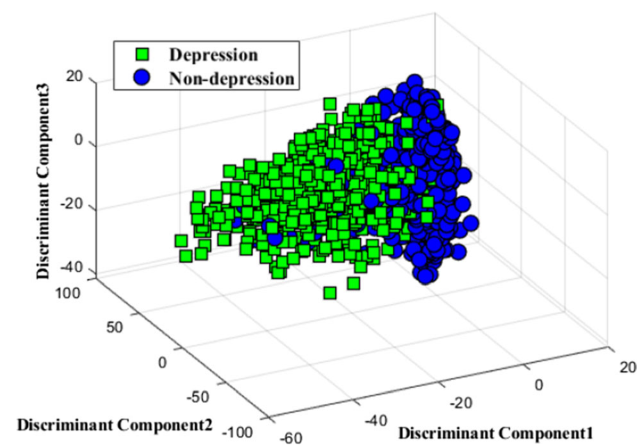


Fig. 9. 3-D plot after LDA on the proposed robust features of two emotional states

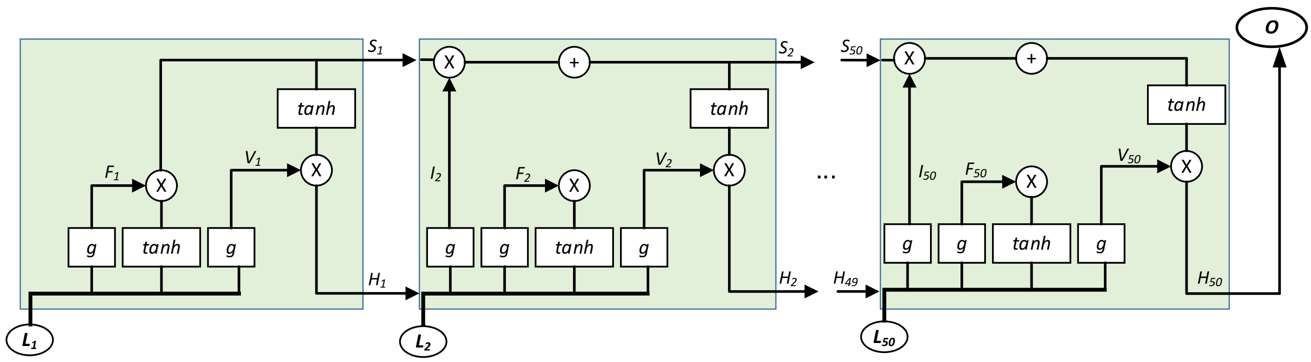


Fig. 10 A basic structure of LSTM-based RNN

The attention technique is basically used for emphasising important information in the current task rather than other useless information. Hence, it can be applied on top of the LSTM layers to improve the model's accuracy. Finally, the output can be determined using a *softmax* function as

$$O = \mathrm{softmax}(W_O A_O + b_O) \tag{12}$$

where W and b represent weights and bias, respectively. Figures 11 and 12 show the algorithms for training and prediction of depression or non-depression through RNN, respectively.

5 Experimental results and discussion

For experiments, two text datasets were obtained from the queries and answers from ung.no website. The dataset comprises of several categories including depression texts. The annotations of the messages were done with the help of professionals such as medical doctors and psychologists. All the experiments are done on a computer that has

Algorithm 2: Model_Training (Training_Texts)

1.Begin

2. Assign M= Number of Training Text Samples

3. For i:= 1 to M do

4. Obtain i^{th} Text, T = Training_Texts(i).

5. Obtain features L_i

6.Assign label of i^{th} Text to Y_i

7. End for

8.Obtain all training features, L and labels, Y

9.Train an RNN, R based on L and Y

10.End

Fig. 11 The algorithm of training features from all texts with RNN

Algorithm3: Model_Testing(Testing_Text)

1.Begin

2. Obtain features L

3. N = Weights after applying L on the trained RNN, R

4. W = Weights of the final 2 neurons in N

5. D = Arg_Max(W)

6. **If** D =0 **then** Decision = Depression

7. Else Decision = non-depression

6.End

Fig. 12 The algorithm of testing of a test text message with the trained RNN

Intel(R) Core(TM) i7-7700HQ CPU with the speed of 2.80 GHz and 2.81 GHz, memory of 32 GB, Windows® 10 operating system, and TensorFlow deep learning tool version 2.4.1.

5.1 First dataset and experiments

From the whole collection of texts of different categories, 11,807 of them were extracted for the first dataset and experiments that consisted of 1820 texts categorized as depression texts (describing symptoms of depression) and the other 9987 as non-depression texts (not describing symptoms of depression). Tables 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 represent the classification reports of tenfold used in the

Table 2 Classification report of fold-1 in the first dataset using proposed approach

State	Precision	Recall	F1-score	Support
Depression	0.96	0.98	0.97	189
Non-depression	1.00	0.99	0.99	992
Mean/Total	0.98	0.99	0.98	1181

Table 3 Classification report of fold-2 in the first dataset using proposed approach

State	Precision	Recall	F1-score	Support
Depression	0.98	0.97	0.97	184
Non-depression	0.99	1.00	0.99	997
Mean/Total	0.98	0.98	0.98	1181

Table 4 Classification report of fold-3 in the first dataset using proposed approach

State	Precision	Recall	F1-score	Support
Depression	0.97	0.95	0.96	187
Non-depression	0.99	0.99	0.99	994
Mean/Total	0.98	0.97	0.975	1181

Table 5 Classification report of fold-4 in the first dataset using proposed approach

State	Precision	Recall	F1-score	Support
Depression	0.97	0.97	0.97	162
Non-depression	1.00	1.00	1.00	1019
Mean/Total	0.98	0.98	0.98	1181

Table 6 Classification report of fold-5 in the first dataset using proposed approach

State	Precision	Recall	F1-score	Support
Depression	1.00	0.95	0.98	190
Non-depression	0.99	1.00	1.00	991
Mean/Total	0.99	0.97	0.99	1181

Table 7 Classification report of fold-6 in the first dataset using proposed approach

State	Precision	Recall	F1-score	Support
Depression	0.98	0.93	0.96	188
Non-depression	0.99	1.00	0.99	993
Mean/Total	0.98	0.96	0.97	1181

experiments where each fold consist of 90% data as training and rest as testing. Figures 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 show the confusion matrices of each fold. Figure 23 depicts the accuracy and loss for 100 epochs during the training of the ten different folds. The overall training of the folds looks good except a little negligible fluctuation. Figure 24 shows the attention-based LSTM model used in this work where there are 53,358 parameters represented by an LSTM layer with 50 memory units, an attention layer, and a dense layer for 2 different emotional states (i.e., depression and non-depression).

Table 8 Classification report of fold-7 in the first dataset using proposed approach

State	Precision	Recall	F1-score	Support
Depression	0.96	1.00	0.98	172
Non-depression	1.00	0.99	1.00	1009
Mean/Total	0.98	0.99	0.99	1181

Table 9 Classification report of fold-8 in the first dataset using proposed approach

State	Precision	Recall	F1-score	Support
Depression	0.98	0.99	0.99	179
Non-depression	1.00	1.00	1.00	1001
Mean/Total	0.99	0.99	0.99	1181

Table 10 Classification report of fold-9 in the first dataset using proposed approach

State	Precision	Recall	F1-score	Support
Depression	1.00	0.97	0.98	187
Non-depression	0.99	1.00	1.00	993
Mean/Total	0.99	0.98	0.99	1181

Table 11 Classification report of fold-10 in the first dataset using proposed approach

State	Precision	Recall	F1-score	Support
Depression	0.98	0.97	0.97	182
Non-depression	0.99	1.00	0.99	998
Mean/Total	0.98	0.98	0.98	1181

5.2 Comparison with traditional approaches

We compared the proposed approach with traditional approaches where the proposed one showed superior results than others. We first applied traditional machine learning approaches using different features (i.e., typical one-hot, TF-IDF, and proposed features) with other conventional machine learning algorithms (i.e., logistic

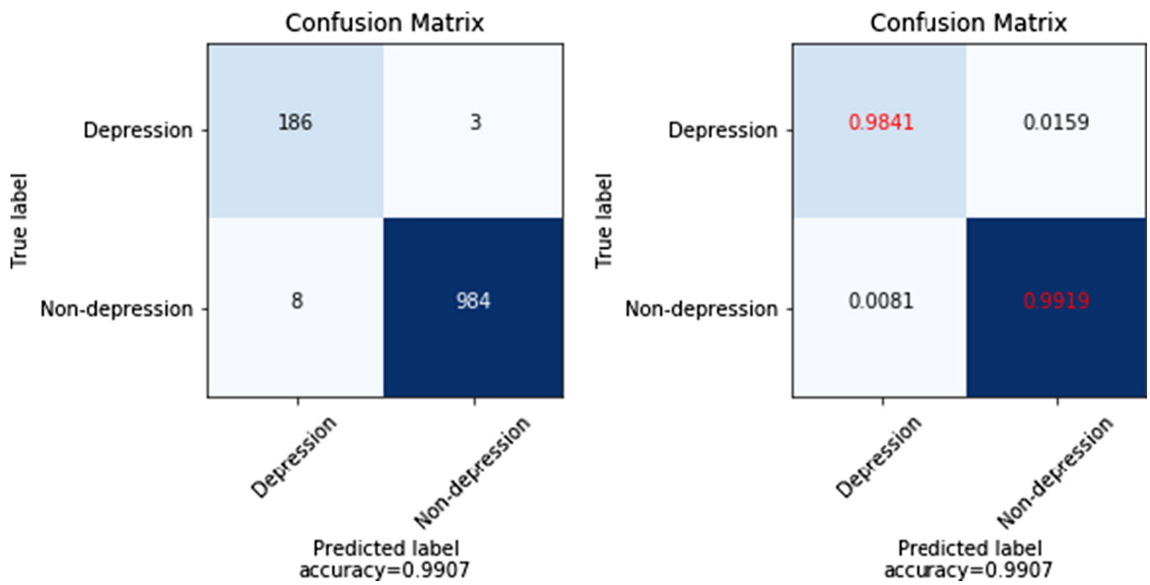


Fig. 13 Confusion matrix of fold-1 in the first dataset using proposed approach

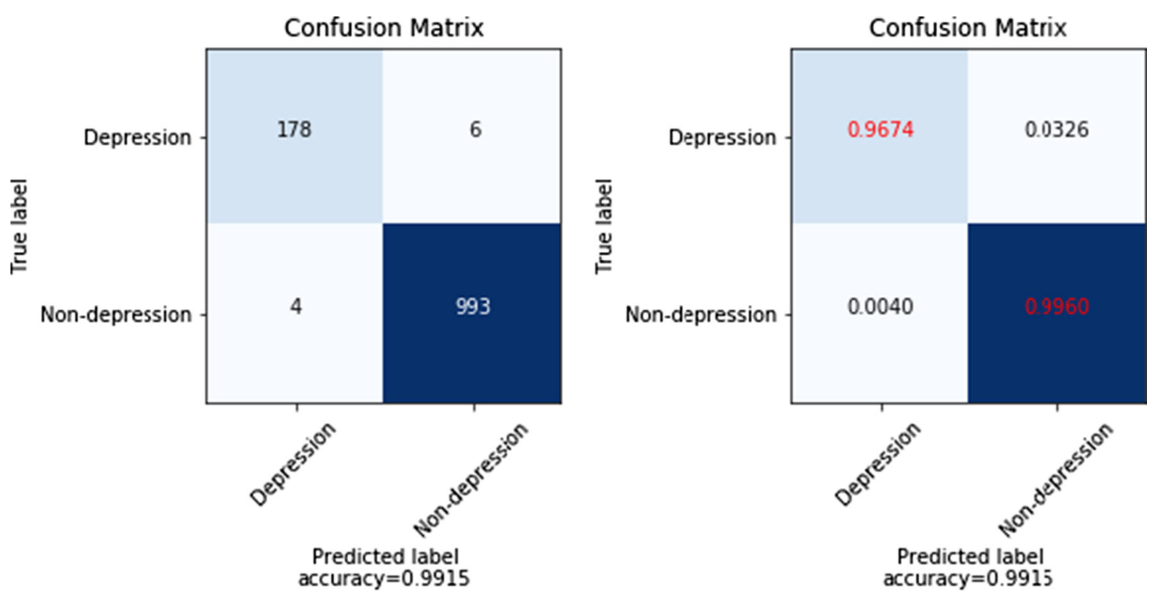


Fig. 14 Confusion matrix of fold-2 in the first dataset using proposed approach

regression, decision trees, support vector machines (SVM), typical large artificial neural network (ANN), DBN, and CNN) but could not achieve more than 91% of mean accuracy as shown in Table 12. Furthermore, we tried LSTM with the traditional as well as proposed features to decode and model the time-sequential information to determine the emotional states. Table 13 and chart in Fig. 25 show the performance of three different approaches to the first dataset where the proposed approach shows the superiority by achieving 98% of mean accuracy over two other approaches.

Besides, another straight-forward approach was applied where the direct presence of the symptoms from “Appendix” was checked to take the binary decision of depression or non-depression. This approach was applied on the whole dataset rather than splitting into training and testing since it was a simple rule-based classification. The direct presence of one or more symptoms-based approach achieved the accuracy of 84.20% where 1684 depression texts were correctly classified among a total of 1807 depression texts and 1730 non-depression texts correctly classified among 10,000 non-depression text. Since there are different ways to express self-depression in texts of

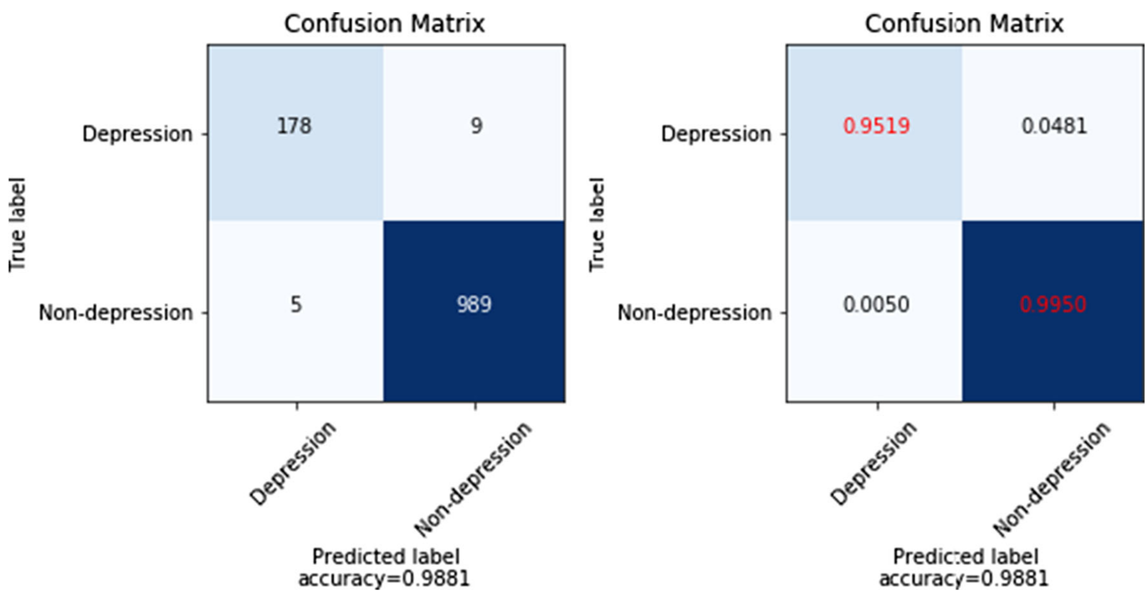


Fig. 15 Confusion matrix of fold-3 in the first dataset using proposed approach

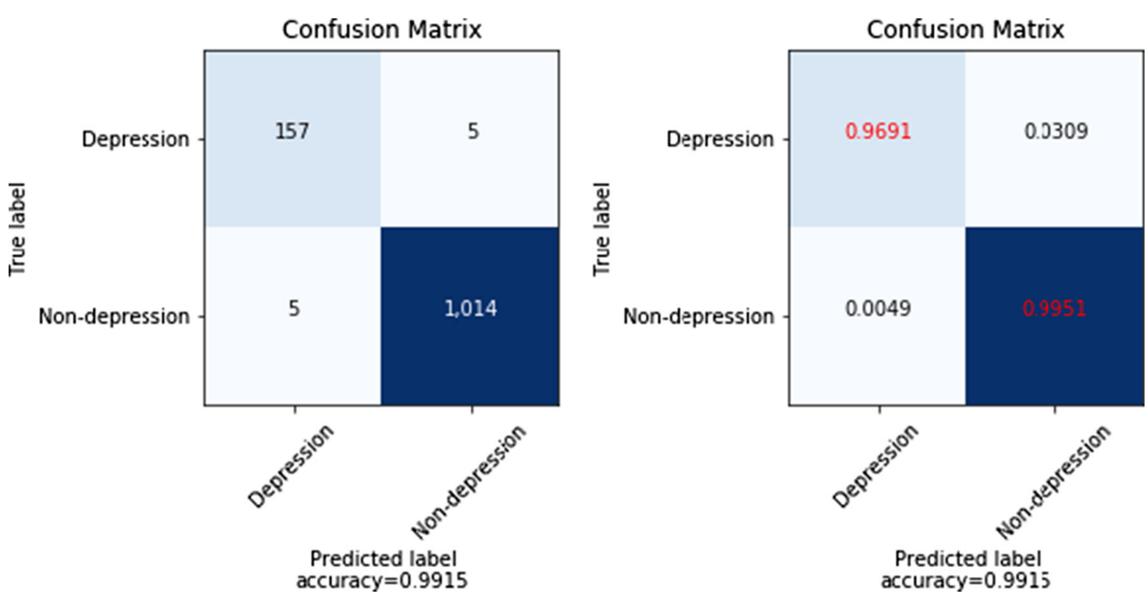


Fig. 16 Confusion matrix of fold-4 in the first dataset using proposed approach

different length, it is hard to apply just a binary rule to determine the depression in the text. Hence, it is better to combine the base words from all the symptoms to define collection of features for depression to apply some complicated algorithms such as sequence-based machine learning algorithm using LSTM-based RNN that has been applied in this work.

5.3 Second dataset and experiments

For the second dataset, a total of 21,470 text samples were obtained consisting of depression—and non-depressions texts. From which, 1470 were depression texts and rest of the 20,000 were non-depression texts. We applied fivefold cross validation for the second phase experiments with the proposed approach, i.e. using RNN on the robust features. Only the results using the proposed approach is reported here since it showed the best results than the other approaches as shown in the experiments of the first phase, i.e. first dataset. Figures 26, 27, 28, 29, 30 represent the confusion matrices of fivefold used in the second experiments where each fold consist of 80% data as training and rest 20% as testing. The experimental results show a remarkable performance of the proposed features followed

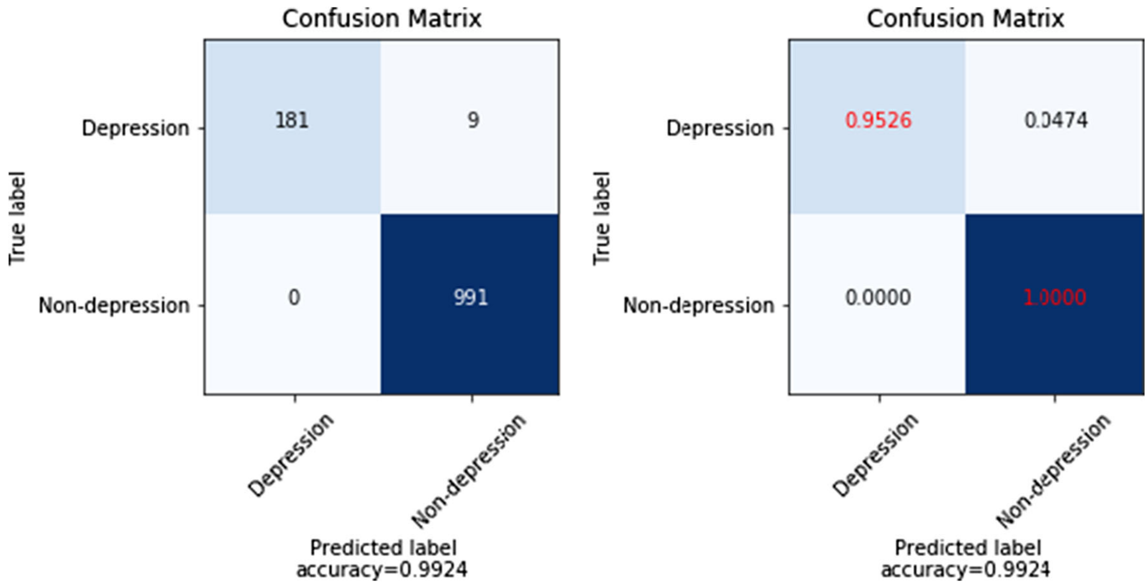


Fig. 17 Confusion matrix of fold-5 in the first dataset using proposed approach

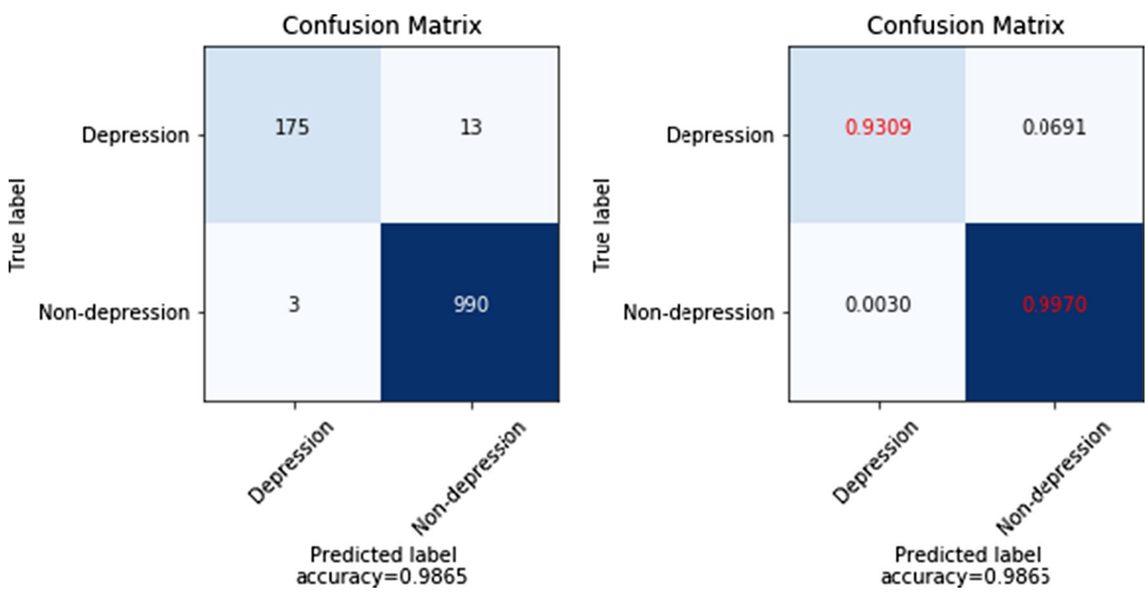


Fig. 18 Confusion matrix of fold-6 in the first dataset using proposed approach

by one-hot and LSTM where the mean recall rate of depression and non-depression is 0.98 and 0.99, respectively. The mean accuracy is 0.99 that shows the robustness of the proposed approach.

In summary, the above experimental results show the overall efficiency of the proposed depression prediction system using depression symptom-based features and time-sequential LSTM-based machine learning model. The proposed system shows better results than existing latest approaches for depression prediction. For instance, in [12], the work is basically based on a measuring scale considering depression, anxiety and stress, which is a point-based measuring scale obtained by writing four different kind of letters by the candidates. The candidates collected by formal advertisements were asked to write these letters whereas in our database, the participants wrote the text spontaneously expressing their necessity to seek assistance over a national portal. The model used [12] is logistic regression, a simple and basic machine learning model which is usually simple linear model and hence, should not generally fit well where the sample data is distributed nonlinearly. On the contrary, our work adopted time-sequential LSTM-based machine learning model that can separate both linearly and nonlinearly distributed samples from

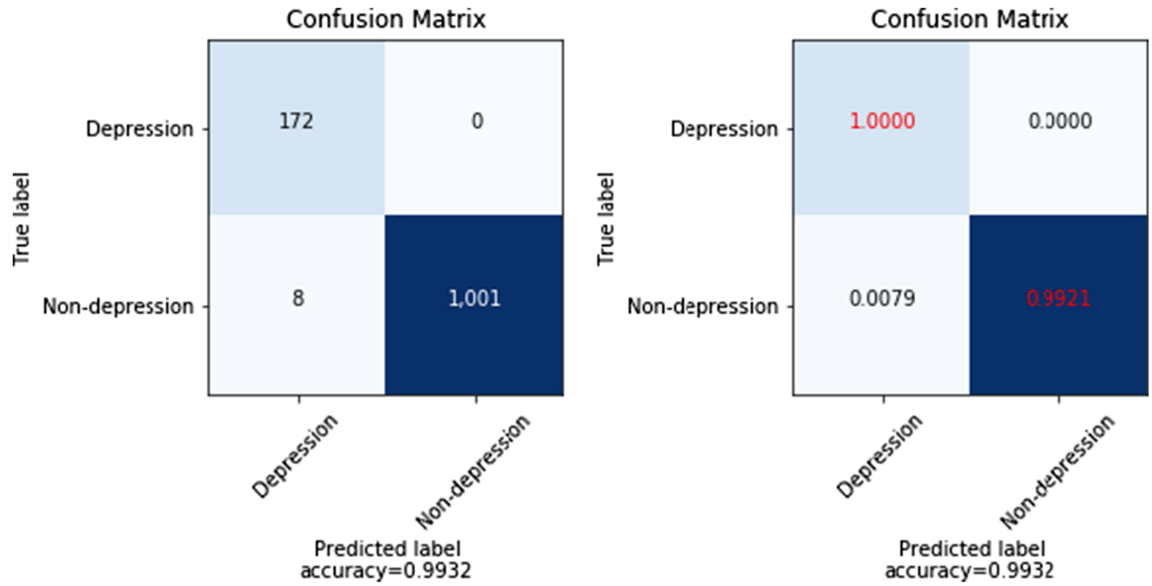


Fig. 19 Confusion matrix of fold-7 in the first dataset using proposed approach

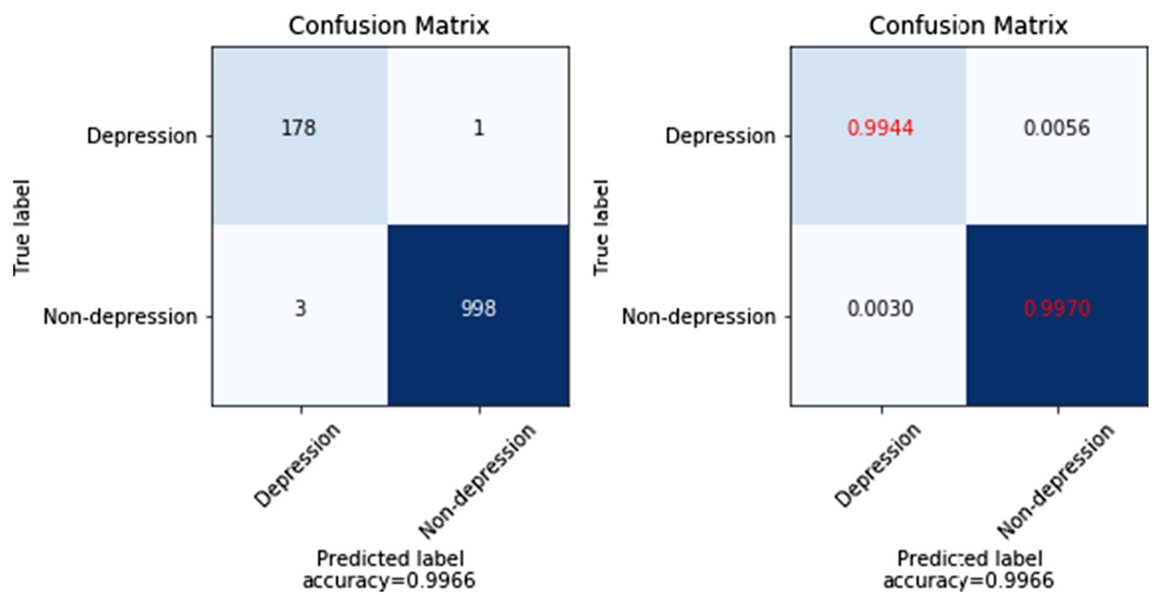


Fig. 20 Confusion matrix of fold-8 in the first dataset using proposed approach

different classes. The proposed approach also overpowers other popular deep learning models such as DBN and CNN which are usually used for non-sequential event modelling.

5.4 XAI to explain the ML decisions

Humans are basically restrained to accept approaches that are not interpretable or trustworthy, pushes the demand for transparent AI to increase. Hence, focusing only on performance of the AI models, gradually makes the systems towards unacceptance. Though there is a trade-off between the performance and transparency of machine learning models, improvements in the understanding of the models via explainability can however lead to the correction of the model's deficiencies as well. Therefore, with the target of overcoming the limitations of accepting the current generation AI models, XAI should focus on machine learning techniques to produce more and more explainable models while upholding a high level of accuracy. Besides, they can also make it happen for humans to appropriately understand, trust, and manage the emerging AI phenomena as much as possible. Explainability is a main factor to gain confidence of whether a model would act as intended for a given problem. Most certainly, it is a property of any

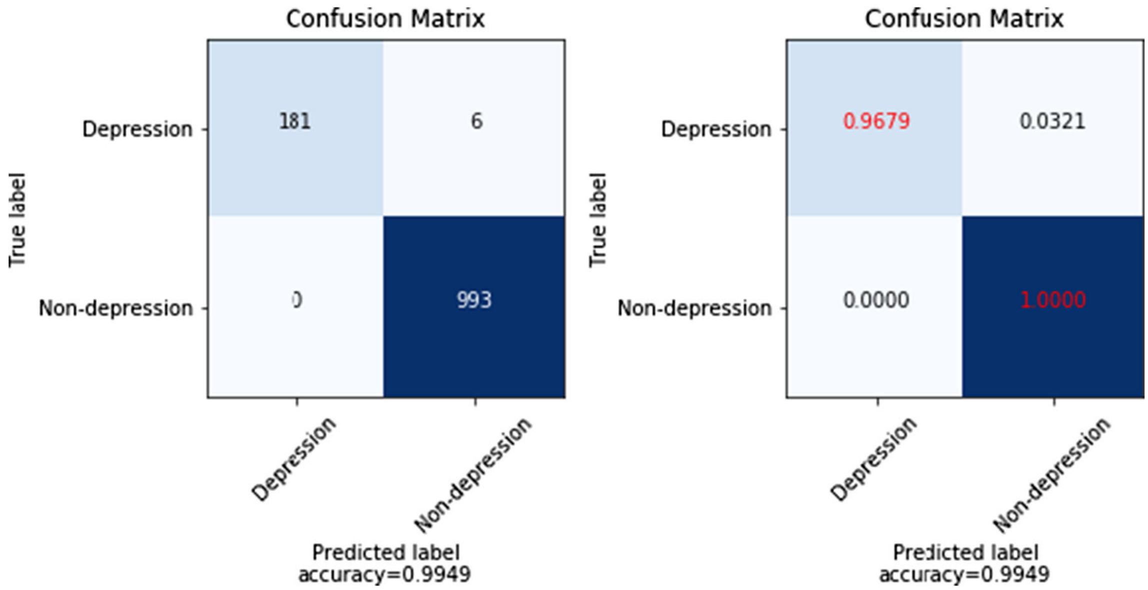


Fig. 21 Confusion matrix of fold-9in the first dataset using proposed approach

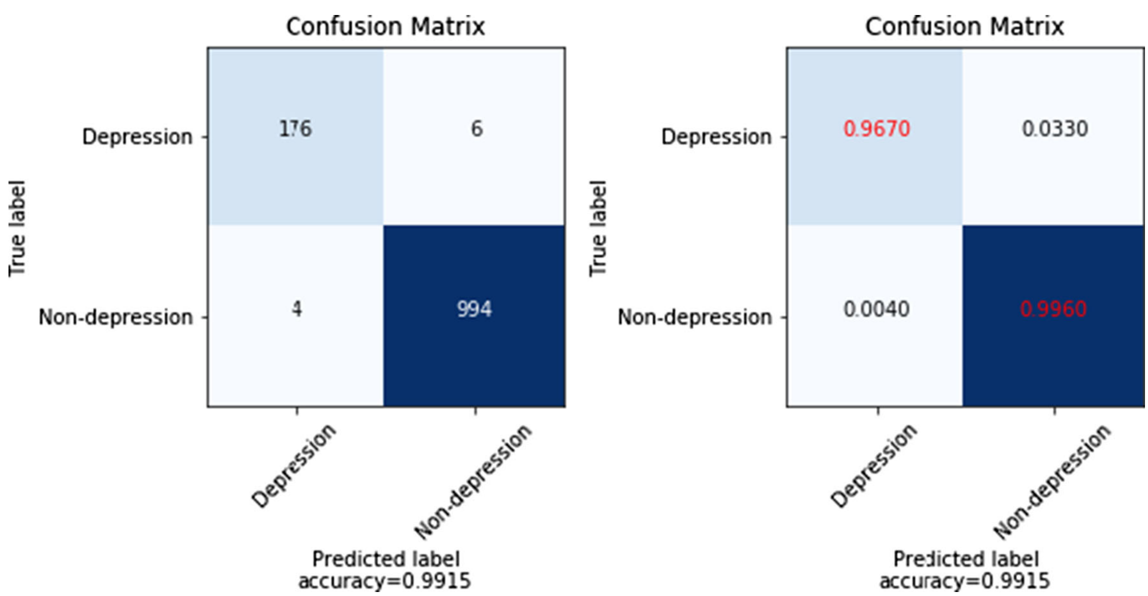


Fig. 22 Confusion matrix of fold-10 in the first dataset using proposed approach

explainable model. Local explanations in AI models handle explainability by dividing the model's complex solutions space into several less complex solution subspaces which are relevant for the whole model. These explanations can utilize some approaches with the differentiating property to explain the model to some basic extent.

Most of the techniques of model simplification are based on rule extraction techniques. The most popular contributions for local post-hoc explanation is based on the approach called Local Interpretable Model-Agnostic Explanations (LIME) [35]. LIME basically generates locally linear models for the predictions of a machine learning model to explain it. It falls under category of the rule-based local explanations by simplification. Explanations by simplification builds a whole new system based on the trained model to be explained. Then, the new simplified

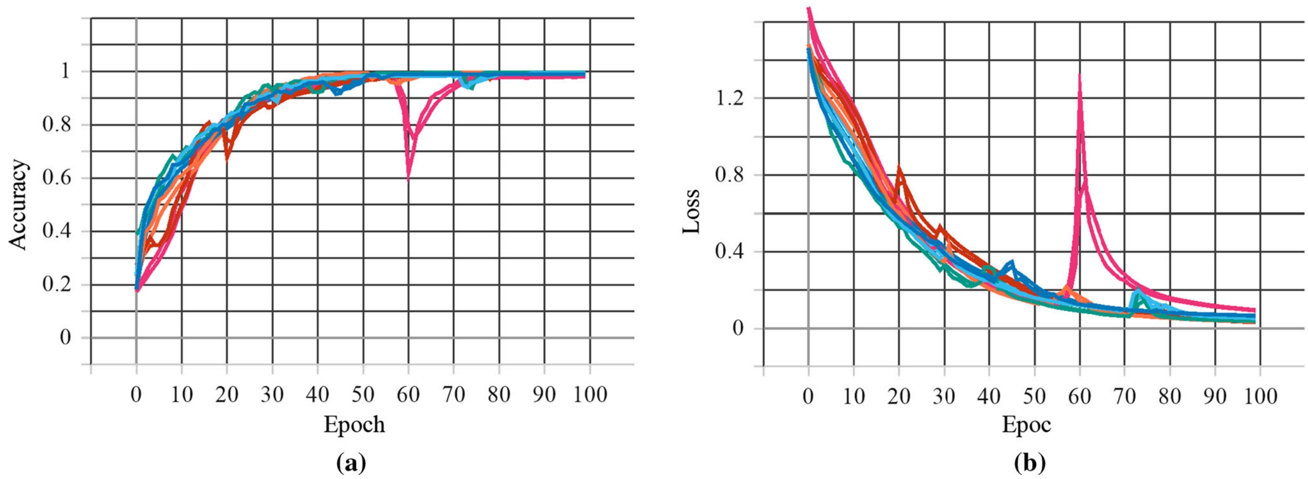
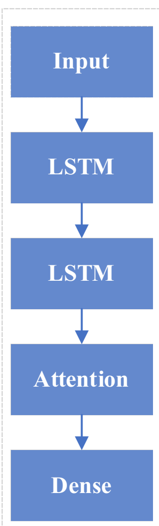


Fig. 23 **a** Accuracy and **b** loss of 10-folds during experiments on the first dataset using the proposed approach




```
Model: "model"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
input_1 (InputLayer)            [(None, 1, 189)]     0
__________________________________________________________________________________________________
lstm (LSTM)                     (None, 1, 50)        48000       input_1[0][0]
__________________________________________________________________________________________________
lstm_1 (LSTM)                   (None, 1, 10)        2440        lstm[0][0]
__________________________________________________________________________________________________
last_hidden_state (Lambda)      (None, 10)           0           lstm_1[0][0]
__________________________________________________________________________________________________
attention_score_vec (Dense)     (None, 1, 10)        100         lstm_1[0][0]
__________________________________________________________________________________________________
attention_score (Dot)           (None, 1)            0           last_hidden_state[0][0]
                                                                 attention_score_vec[0][0]
__________________________________________________________________________________________________
attention_weight (Activation)   (None, 1)            0           attention_score[0][0]
__________________________________________________________________________________________________
context_vector (Dot)            (None, 10)           0           lstm_1[0][0]
                                                                 attention_weight[0][0]
__________________________________________________________________________________________________
attention_output (Concatenate)  (None, 20)           0           context_vector[0][0]
                                                                 last_hidden_state[0][0]
__________________________________________________________________________________________________
attention_vector (Dense)        (None, 128)          2560        attention_output[0][0]
__________________________________________________________________________________________________
dense (Dense)                   (None, 2)            258         attention_vector[0][0]
==================================================================================================
Total params: 53,358
Trainable params: 53,358
Non-trainable params: 0
__________________________________________________________________________________________________
```

Fig. 24 The emotional state model structure and parameters based on attention over LSTM units

model usually tries to optimize its resemblance to its predecessor model functions while reducing the complexity and at the same time, keeping a similar performance. Therefore, once the machine learning decision is obtained, XAI algorithm LIME is applied to see the importance of the features and probabilities towards the decision. Hence, we can understand the presence of the feature importance in the input for the decision, that helps understanding the

Table 12 Prediction accuracy using different approaches to all subjects (%)

Approaches	Mean accuracy (%)
One-hot + Logistic Regression	84
TF-IDF + Logistic Regression	85
One-hot + Decision Trees	82
TF-IDF + Decision Trees	81
One-hot + SVM	83
TF-IDF + SVM	85
One-hot + ANN	88
TF-IDF + ANN	87
One-hot + DBN	89
TF-IDF + DBN	89
One-hot + CNN	91
TF-IDF + CNN	91

outcomes of the system. Figure 31 shows the total class probabilities, top 10 features, their probabilities, and automatically highlighted features in a sample input text using LIME. As can be seen in right side of the figure, features towards depression get higher weights altogether than non-depression class, indicates the person to be in depression mode. The input text, features, and highlights were originally in Norwegian language since the database is from a Norwegian national portal to interact with youth, but the figure shows the corresponding translated text in English for better readability and understanding of the approach. According to the decision from machine learning model and explanations from LIME, the sample text consists of depression. To be noted, the ground truth for the sample text in the figure was the same as the model's prediction (i.e., depression), indicating the robustness of the model's decision and explanation.

Furthermore, Fig. 32 shows summarized probabilities of top 10 features for a paraphrased non-depression example text using LIME. In the figure, left side represents the original part after applying the algorithm and right side the corresponding representation in English for better readability as well as understandability. The overall probabilities of the non-depression text from the machine learning model for depression and non-depression classes were 0.001 and 0.999, respectively.

6 Conclusion

To automatically detect depression symptoms in text for decision support in health care is important. In this work, a multimodal human depression prediction approach has been investigated based on one-hot approach on robust features based on describing depression symptoms and deep learning method, RNN. First, the young users' text data has been obtained from *ung.no,* a public information channel targeting young people in Norway. Then, one-hot method is applied after sequentially extracting the words

Table 13 Prediction accuracy using different approaches to all subjects (%)

Emotional State	Typical One-hot + LSTM	TF-IDF + LSTM	Proposed Features + One-hot + LSTM
Depression	0.86	0.64	0.97
Non-depression	0.96	0.93	0.99
Mean	0.91	0.79	0.98

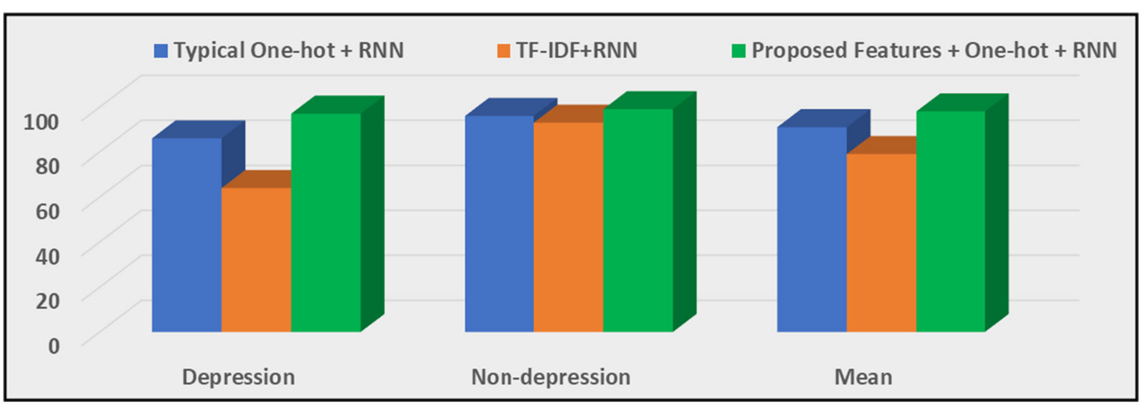


Fig. 25 Performance of three different approaches to the first dataset

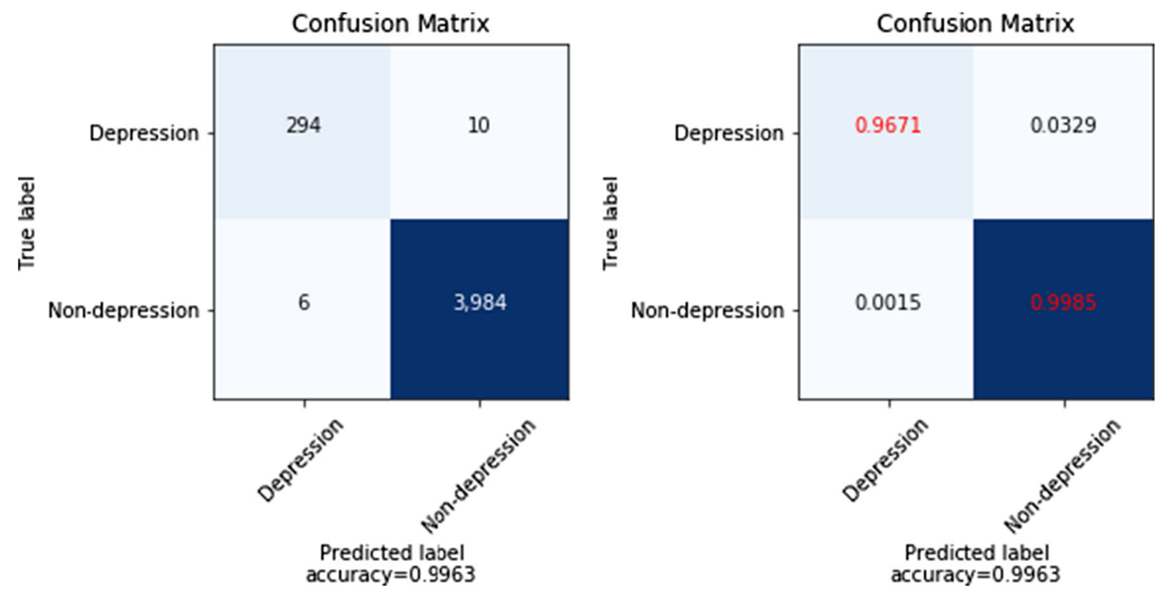


Fig. 26 Confusion matrix of fold-1 in the second dataset using proposed approach

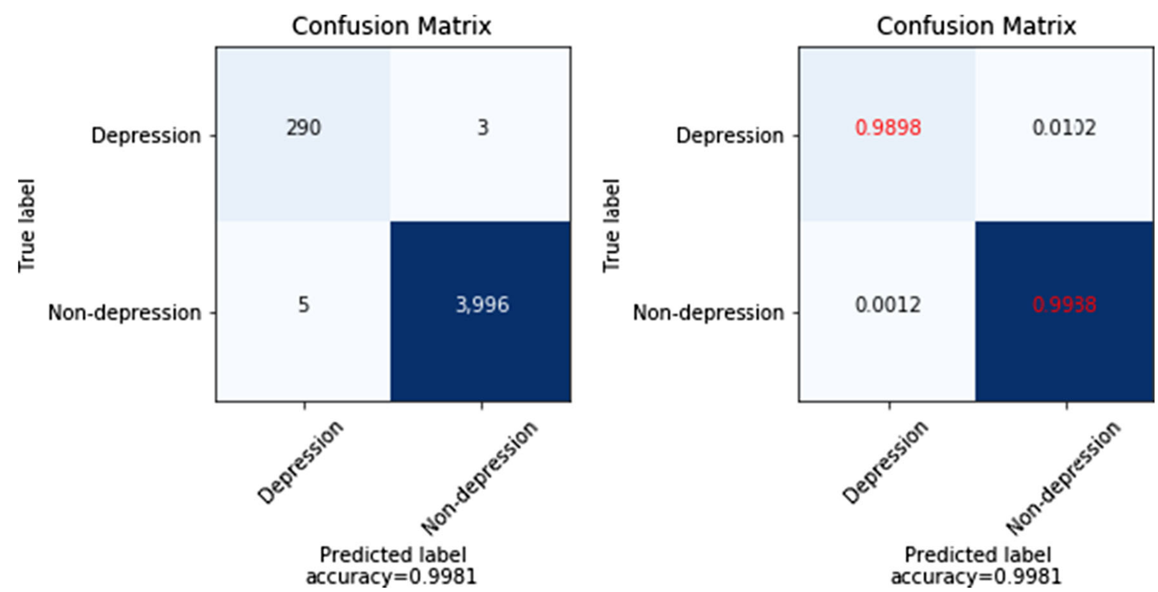


Fig. 27 Confusion matrix of fold-2 in the second dataset using proposed approach

from different sentences and words representing the symptoms of depression. Furthermore, the one-hot features have been applied to train a deep RNN based on LSTM method to model two different emotional states: depression and non-depression. Finally, the trained RNN has been used for predicting the underlying emotional state in unknown sensor text data. Using the proposed approach, 98% and 99% mean prediction performance has been achieved on first and second dataset consists of around 11,807 and 21,807 texts, respectively. Whereas, the traditional approaches could achieve maximum of 91% mean recognition performance, indicating the robustness of the proposed approach. The proposed approach outperforms the other traditional approaches such as using the proposed

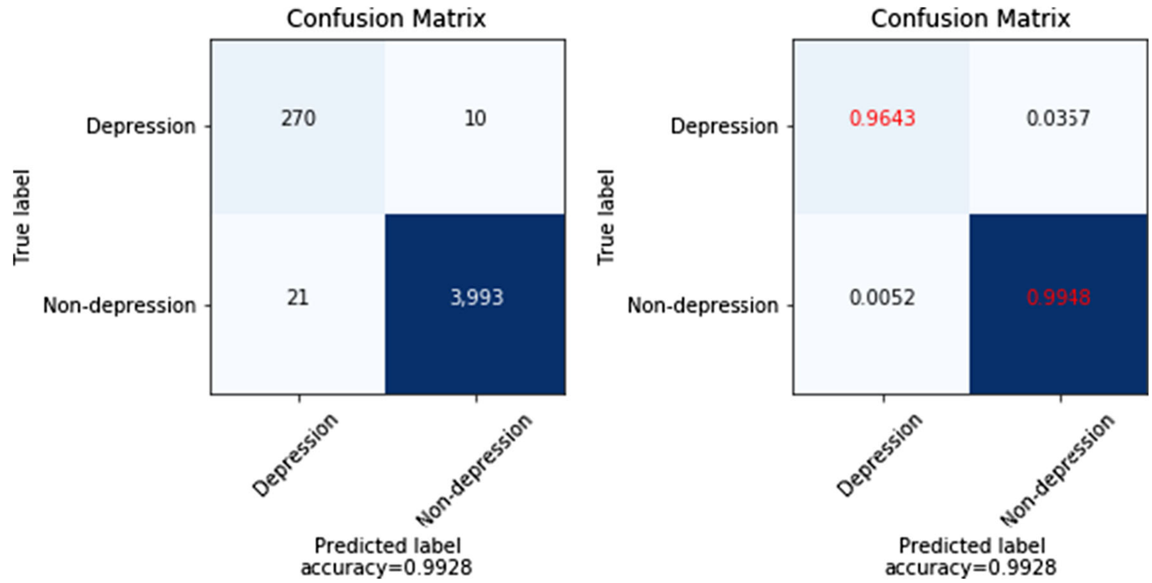


Fig. 28 Confusion matrix of fold-3 in the second dataset using proposed approach

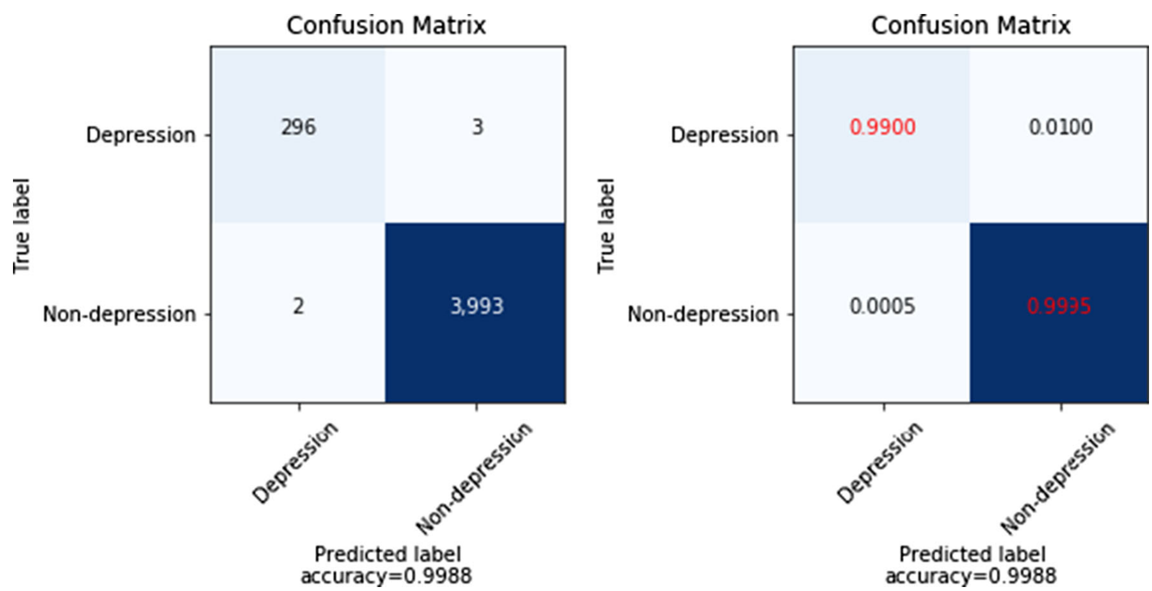


Fig. 29 Confusion matrix of fold-4 in the second dataset using proposed approach

features with logistic regression, DBN, and CNN models as well as using typical one-hot and TF-IDF features with RNN. Besides, an XAI algorithm, LIME has been utilized to see whether the proposed system generates meaningful explanations to support its decision. Thus, the features used in this work can be used to support the machine learning decisions and to contribute to design effective user interface for better affective care. The deep learning-based efficient system can be explored in greater levels with comprehensive dataset. Detection of depression symptoms in texts can be applied in mental health care services for real-time analysing and predicting normal as well as severe states of mood disorders in smart environments combined with latest technologies. For instance, smart chatbot systems providing informational support about depression can

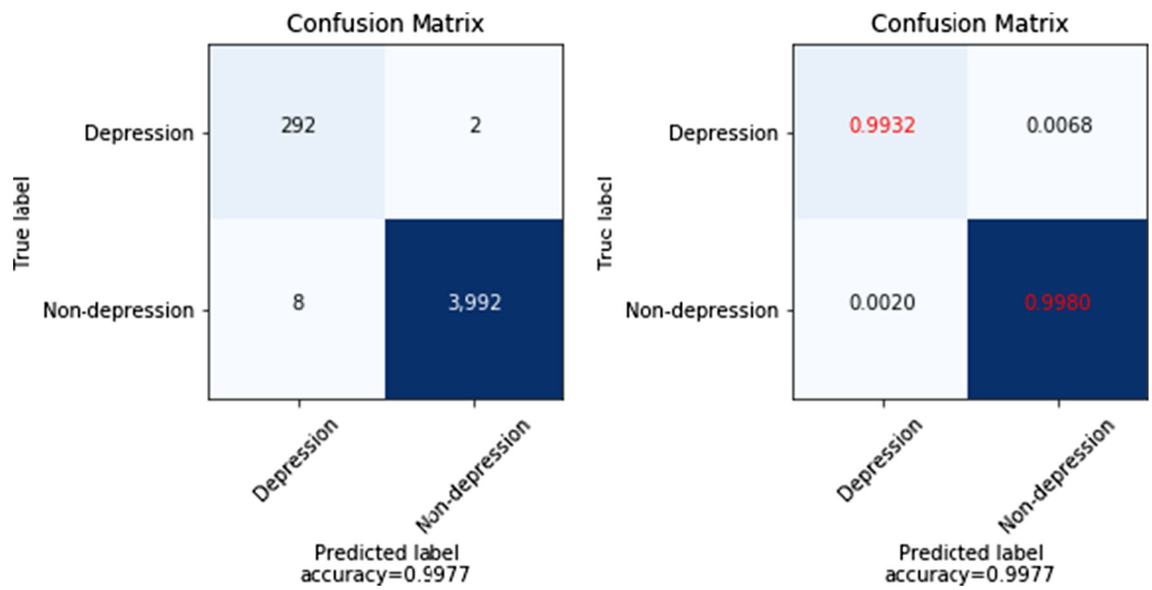


Fig. 30 Confusion matrix of fold-5 in the second dataset using proposed approach

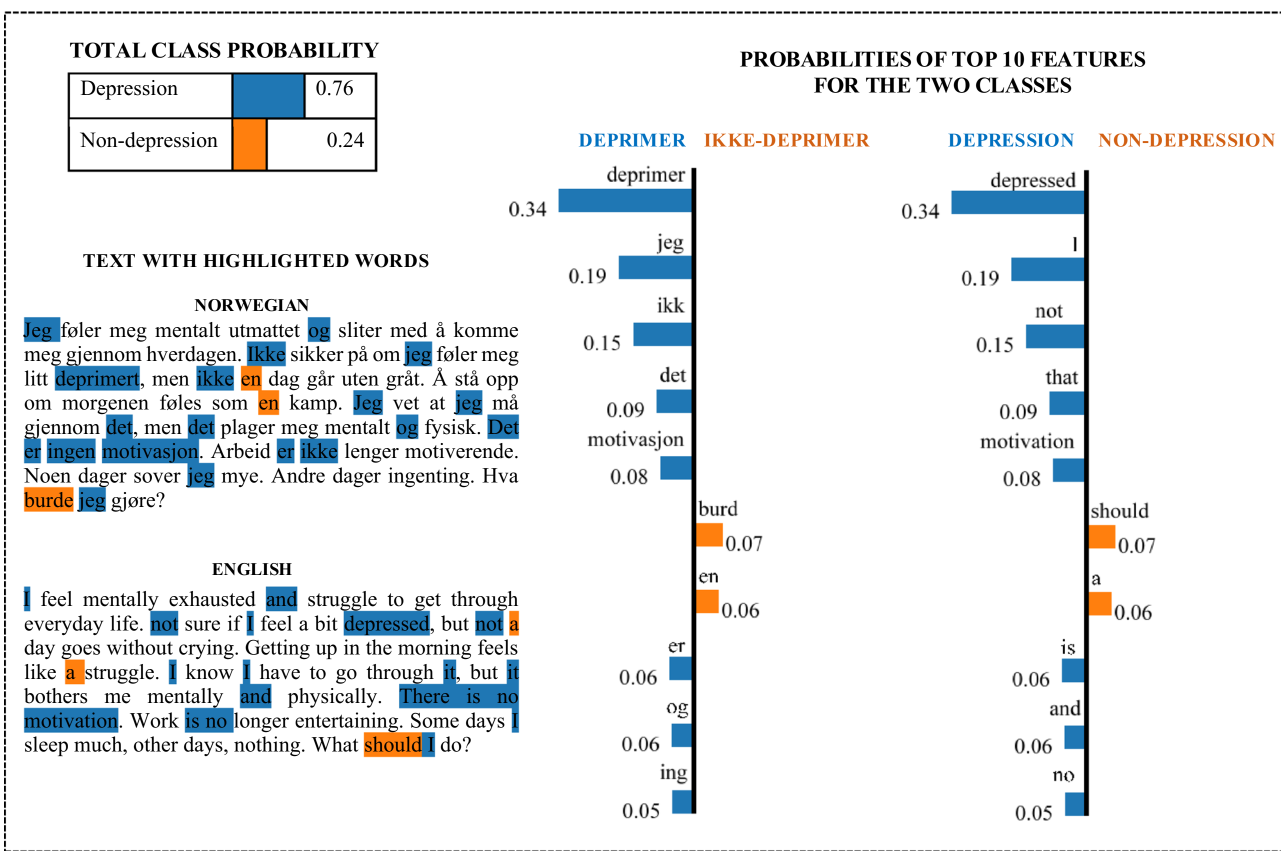


Fig. 31 Total class probabilities, top 10 features, their probabilities, and automatically highlighted features in a sample input text using LIME

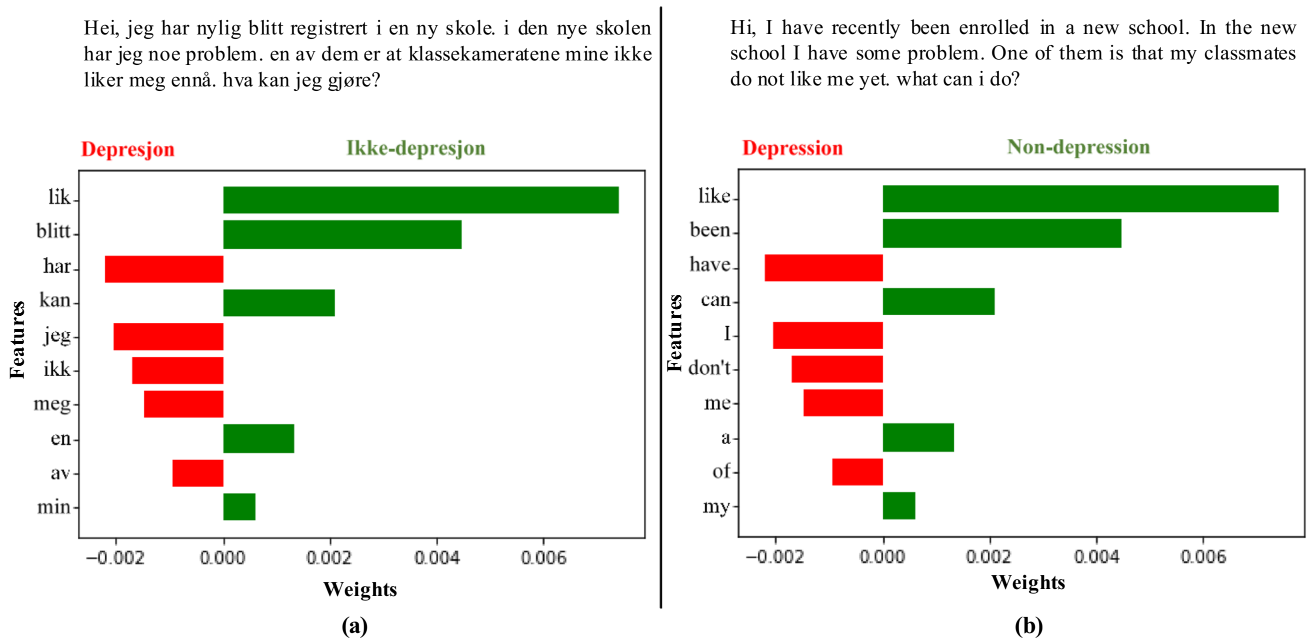


Fig. 32 **a** Probabilities of top 10 features for a normal non-depression text (on the top) using LIME and **b** corresponding English on the right

be a feasible solution for both health professionals working with youth and youths struggling with mental health issues.

Appendix

Symptoms (Norwegian)	Symptoms (Translated)
alle hater meg	everyone hates me
alt er jævlig	everything is damn
alt er så jævlig	everything is so damn
alt var jævlig	everything was damn
alt var så jævlig	everything was so damn
ikke apetitt	no appetite
lite apetitt	little appetite
ingen apetitt	no appetite
avslutte livet	end life
bli borte fra alt	get away from it all
bryr meg ikke om noe	I do not care about anything
ikke bryr meg om noe	do not care about anything
brydde meg ikke om noe	did not care about anything
ikke brydde meg om noe	did not care about anything
jeg burde bli glad	I should be happy
depremert	depressed
deprimert	depressed
depresjon	depression
deprisjon	depression
Deppa	depressed

Symptoms (Norwegian)	Symptoms (Translated)
distensert meg fra	distanced me from
distansert meg fra	distanced me from
Dø	die
Død	death
Dødd	died
Dør	dying
ikke mer energi	no more energy
tom for energi	empty of energy
lite energi	little energy
ingen energi	no energy
ikke har noen energi	have no energy
Energiløs	energyless
lavt energinivå	low energy level
tappet for energi	drained of energy
aldri nok energi	never enough energy
som et forferdelig menneske	as a terrible human being
vondt inni meg	hurt inside me
meg god nok	me good enough
meg god nokk	me good enough
Gråt	crying
Grine	grine
går ikke ut mer	does not go out anymore
ikke går ut mer	does not go out anymore
ikke gå ut mer	do not go out anymore
det grusomt	it cruel
det ille	the bad
har det vondt	is in pain

Symptoms (Norwegian)	Symptoms (Translated)
hatt det vondt	had it hurt
har det så vondt	is in so much pain
hatt det så vondt	had it so painful
har det så sinnyskt vondt	it hurts so insanely
hatt det så sinnyskt vondt	had it so insanely painful
hater å leve	hate living
hater livet mitt	hate my life
helt på bunnen	at the very bottom
alt er håpløst	everything is hopeless
alt føles helt håpløst	everything feels completely hopeless
har mistet håp	have lost hope
jeg mister håp	I lose hope
Håpløshet	hopelessness
jeg ikke får til noe	I do not get anything
jeg ikke får til noen	I do not get to anyone
jeg ikke kan gjøre noe riktig	I can not do anything right
jeg ikke gjør noe riktig	I'm not doing anything right
ikke klarer å tenke	unable to think
ikke klarer og tenke	unable to think
ikke leve lenger	no longer live
ikke lyst til å gjøre noe	not wanting to do anything
Meningsløs	meaningless
ingenting har mening	nothing makes sense
ingenting har noen mening	nothing has any meaning
ser ikke noe mening	sees no meaning
ikke morsomt lenger	no fun anymore
ikke overskudd til noe	no profit to anything
ikke tro på meg selv	do not believe in myself
ikke være sosial lenger	not be social anymore
blitt usosial	become antisocial
indre uro	inner turmoil
ingen bryr seg om meg	nobody cares about me
ingen glede	no joy
ingen liker meg	nobody likes me
ingen som liker meg	no one like me
ingen lykke	no happiness
ingen liker meg	nobody likes me
ingen som liker meg	no one like me
ingen savner meg	no one misses me
ingen vil savne meg	no one will miss me
ingenting føles	nothing feels
ingenting interesserer meg	nothing interests me
mistet interesse	lost interest
ingenting å leve for	nothing to live for
jeg er en vanskelig person	I am a difficult person
klarer ikke leve	unable to live
Ukonsentrert	unconcentrated
ikke konsentrere meg	do not concentrate
ikke å konsentrere meg	not to concentrate
ikke og konsentrere meg	not and concentrate

Symptoms (Norwegian)	Symptoms (Translated)
til å konsentrere meg	to concentrate
med å konsentrere meg	with concentrating
ikke lenger konsentrasjon	no longer concentration
mistet konsentrasjon	lost concentration
mista konsentrasjon	lose concentration
meg langt nede	me far down
meg så langt nede	me so far down
lei av livet	tired of life
lei meg	sad
leve med meg selv	live with myself
meg likegyldig	me indifferent
følelse av likegyldighet	feeling of indifference
likegyldigheten	indifference
jeg er likegyldig	I'm indifferent
lite initiative	blu initiative
låser meg inne	locks me inside
mistet matlyst	lost appetite
ikke matlyst	not appetite
ingen matlyst	no appetite
liten matlyst	small appetite
har ikke matlyst	have no appetite
meg ubetydelig	me insignificant
mistet motivasjon	lost motivation
ikke motivasjon	not motivation
lite motivasjon	little motivation
demotivert	demotivated
motivasjonen er borte	the motivation is gone
motivasjonen er vekk	the motivation is gone
mørkeste tanker	darkest thoughts
de mørke skyene	the dark clouds
mørkt hull	dark hole
mørkt sted	dark place
mørke tanker	dark thoughts
alt er mørkt	everything is dark
nedfor	down
nedenfor	below
nedstemt	voted down
tenke negativt	think negatively
negative tanker	negative thoughts
negativt inni meg	negative inside me
nervøs følelse	nervous feeling
nervøs hele tiden	nervous all the time
nytteløst	useless
oppgitt	tired
selvmord	suicide
selvmordstanker	suicidal thoughts
skyver vennene mine vekk	pushes my friends away
skyver venner vekk	pushes friends away
sliten	tired
jeg sliter	I'm struggling

Symptoms (Norwegian)	Symptoms (Translated)
sliter med meg	struggling with me
sluttet jeg å være med på	I stopped participating
maten smaker ingenting	the food tastes nothing
meg som en taper	me as a loser
sove	sleep
søvn	sleep
sovne	to fall asleep
sover bort	sleeping away
stenger meg inne	shuts me in
stengte meg inne	locked me inside
suicid	suicide
suisid	suicide
ende livet mitt	end my life
ta livet mitt	take my life
ende mitt eget liv	end my own life
ta livet av meg	take my life
ta mitt eget liv	take my own life
tar livet mitt	takes my life
tar mitt eget liv	takes my own life
tenke på døden	think of death
jeg få ting til å gå fortere	I make things go faster
helt tom	completely empty
tomhet	emptiness
er jeg tom	am I empty
jeg er tom	I'm empty
trist	sad
konstant trøtt	constantly tired
konstant trett	constantly tired
alltid trøtt	always tired
alltid trett	always tired
tare	tear
meg ubetydelig	me insignificant
meg ubrukelig	me useless
umotivert	unmotivated
utbrent	burnt out
jeg er utslitt	I'm exhausted
jeg er så utslitt	I'm so exhausted
føler meg så utslitt	feel so exhausted
er bare helt utslitt	is just completely exhausted
psykisk utslitt	mentally exhausted
jeg er veldig utslitt	I'm very exhausted
uutholdelig	unbearable
vekk fra denne verdenen	away from this world
vekker ikke følelser lenger	does not evoke emotions anymore
ingenting vekker følelser	nothing evokes emotions
meg verdiløs	me worthless
verdiløs jeg er	worthless I am
jeg er verdiløs	I'm worthless
ønsker å være død	wants to be dead
ønsket å være død	wanted to be dead

Symptoms (Norwegian)	Symptoms (Translated)
vil ikke leve	will not live
ikke vil leve	will not live
ikke ville leve	would not live
ville ikke leve	would not live
noe mer å leve for	something more to live for

Acknowledgements This work was supported by the Research Council of Norway under Grant Number 262848.

Funding Open access funding provided by SINTEF AS.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. WHO (2021) https://www-who-int.ezproxy.uio.no/news-room/fact-sheets/detail/suicide.
2. Häfner H, Maurer K, Trendler G, an der Heiden W, Schmidt M (2005) The early course of schizophrenia and depression. Eur Arch Psychiatry Clin Neurosci 255(3):167–173
3. Fusar-Poli P, Nelson B, Valmaggia L, Yung AR, McGuire PK (2014) Comorbid depressive and anxiety disorders in 509 individuals with an at-risk mental state: impact on psychopathology and transition to psychosis. Schizophr Bull 40(1):120–131
4. World Health Organization (2017) Depression and other common mental disorders: global health estimates (No. WHO/MSD/MER/2017.2)
5. Kessler RC, Bromet EJ (2013) The epidemiology of depression across cultures. Annu Rev Public Health 34:119–138. https://doi.org/10.1146/annurev-publhealth-031912-114409
6. Depression WHO (2017) Other common mental disorders: global health estimates. World Health Organization, Geneva, pp 1–24
7. Weinberger AH, Gbedemah M, Martinez AM, Nash D, Galea S, Goodwin RD (2018) Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. Psychol Med 48(8):1308–1315
8. Cambria E, White B (2014) Jumping NLP curves: a review of natural language processing research. IEEE Comput Intell Mag 9(2):48–57
9. Picard RW (2003) Affective computing: challenges. Int J Human Comput Stud 59(1–2):55–64

10. Cambria E (2016) Affective computing and sentiment analysis. IEEE Intell Syst 31(2):102–107
11. Shangfei W, Guozhu P, Zhuangqiang Z, Zhiwei Xu (2019) Capturing emotion distribution for multimedia emotion tagging. IEEE Trans Affect Comput. https://doi.org/10.1109/taffc.2019.2900240
12. Havigerová JM, Haviger J, Kučera D, Hoffmannová P (2019) Text-based detection of the risk of depression. Front Psychol 10:513
13. Tao J, Tan T (2005) Affective computing: a review. In: Proceedings of the international conference on affective computing and intelligent interaction, pp 981–995
14. Yang H, Willis A, De Roeck A, Nuseibeh B (2012) A hybrid model for automatic emotion recognition in suicide notes. Biomed Inform Insights 5(1):17–30
15. Desmet B, Hoste VR (2013) Emotion detection in suicide notes. Expert Syst Appl 40(16):6351–6358
16. Allouch M, Azaria A, Azoulay R, Ben-Izchak E, Zwilling M, Zachor DA (2018) Automatic detection of insulting sentences in conversation. Paper presented at: Proceedings of the 2018 IEEE international conference on the science of electrical engineering in Israel (ICSEE); pp 1–4; IEEE.
17. John S, Ederyn W, Bruce C (1976) The social psychology of telecommunication. Wiley, London
18. Strapparava C, Mihalcea R (2008) Learning to identify emotions in text. Paper presented at: Proceedings of the 2008 ACM symposium on applied computing, pp 1556–1560; ACM
19. Ding Y, Chen X, Fu Q, Zhong S (2020) A depression recognition method for college students using deep integrated support vector algorithm. IEEE Access 8:75616–75629. https://doi.org/10.1109/ACCESS.2020.2987523
20. Yildirim E, Avci E (2018) Classificationof groutability of sands by linear discriminant analysis. ZBORNIK RADOVA GEO-EXPO 2018, Oct. 2018.
21. Vannatta RA, LaVenia KN (2020) Linear discriminant analysis. In Atkinson P, Delamont S, Cernat A, Sakshaug JW, Williams RA (eds) SAGE research methods foundations. doi: https://doi.org/10.4135/9781526421036889610
22. Gardner-Lubbe S (2020) Linear discriminant analysis for multiple functional data analysis. J Appl Stat pp 1–17
23. Uddin MZ, Torresen J (2019) Activity recognition using smartphone sensors, robust features, and recurrent neural network. In: 2019 13th international symposium on medical information and communication technology (ISMICT), May 2019.
24. Kiranyaz S, Ince T, Gabbouj M (2016) Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans Biomed Eng 63(3):664–675
25. Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554
26. Deboeverie F, Roegiers S, Allebosch G, Veelaert P, Philips W (2016) Human gesture classification by brute-force machine learning for exergaming in physiotherapy. In: Proceedings of IEEE conference on computational intelligence and games (CIG), Santorini, pp 1–7
27. Graves A, Mohamed A, Hinton G (2013) Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 6645–6649
28. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
29. Zaremba W, Sutskever I, Vinyals O (2014) Recurrent neural network regularization. arXiv preprint arXiv: 1409.2329
30. Gers FA, Schraudolph NN, Schmidhuber J (2003) Learning precise timing with LSTM recurrent networks. J Mach Learn Res 3:115–143
31. Sak H, Senior AW, Beaufays F (2014) Long short-term memory recurrent neural network architectures for large scale acoustic modeling. INTERSPEECH, pp 338–342
32. Kumari T, Mishra A (2020) Deep networks and deep learning algorithms. In: Artificial intelligence trends for data analytics using machine learning and deep learning approaches, pp 199–214
33. Liu P, Wang J, Guo Z (2020) Multiple and complete stability of recurrent neural networks with sinusoidal activation function. IEEE Trans Neural Netw Learn Syst, pp 1–12
34. Uddin MZ, Hassan M, Alsanad A, Savaglio C (2020) A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare. Inform Fusion. https://doi.org/10.1016/j.inffus.2019.08.004
35. Govindasamy KA, Palanichamy N (2021) Depression detection using machine learning techniques on twitter data. In: 2021 5th international conference on intelligent computing and control systems (ICICCS), May 2021.
36. Dong Y, Yang X (2021) A hierarchical depression detection model based on vocal and emotional cues. Neurocomputing 441:279–290
37. Fortino G, Parisi D, Pirrone V, Di Fatta G (2014) BodyCloud: a SaaS approach for community Body Sensor Networks. Futur Gener Comput Syst 35:62–79
38. Fortino G, Galzarano S, Gravina R, Li W (2015) A framework for collaborative computing and multi-sensor data fusion in body sensor networks. Inform Fusion 22:50–70
39. Zhang Y, Gravina R, Lu H, Villari M, Fortino G (2018) PEA: Parallel electrocardiogram-based authentication for smart healthcare systems. J Netw Comput Appl 117:10–16
40. Wang R, Hao Y, Yu Q, Chen M, Humar I, Fortino G (2021) Depression Analysis and Recognition based on Functional Near-infrared Spectroscopy. IEEE J Biomed Health Inform. doi: https://doi.org/10.1109/JBHI.2021.3076762. Epub ahead of print. PMID: 33929968.
41. Hassan MM, Alam MGR, Uddin MZ, Huda S, Almogren A, Fortino G (2019) Human emotion recognition using deep belief network architecture. Inform Fusion 51:10–18
42. Tjoa E, Guan C (2021) A survey on explainable artificial intelligence (XAI): toward medical XAI. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2020.3027314
43. Wang N, Kosinski M, Stillwell DJ, Rust J (2014) Can well-being be measured using Facebook status updates? Validation of facebook's gross national happiness index. Soc Indic Res 115(1):483–491
44. Baines T, Baines T, Wittkowski A, Wittkowski A (2013) A systematic review of the literature exploring illness perceptions in mental health utilising the self-regulation model. J Clin Psychol Med Settin 20(3):263–274
45. Cohen JR, So FK, Young JF, Hankin BL, Lee BA (2019) Youth depression screening with parent and self-reports: assessing current and prospective depression risk. Child Psychiatry Hum Dev 50(4):647–660
46. Coyne JC, Thompson R, Racioppo MW (2001) Validity and efficiency of screening for history of depression by self-report. Psychol Assess 13(2):163–170
47. Sim AYL, Borthwick A (2018) Record2Vec: unsupervised representation learning for structured records. IEEE Int Confer Data Min 2018:1236–1241. https://doi.org/10.1109/ICDM.2018.00165
48. Yuan X, Li L, Shardt YAW, Wang Y, Yang C (2021) Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. IEEE Trans Industr Electron 68(5):4404–4414. https://doi.org/10.1109/TIE.2020.2984443