Journal of Biomedical Informatics 123 (2021) 103921

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

Prediction of anxiety disorders using a feature ensemble based bayesian neural network

Hao Xiong [a,*], Shlomo Berkovsky [a], Mia Romano [b], Roneel V. Sharan [a], Sidong Liu [a], Enrico Coiera [a], Lauren F. McLellan [b]

[a] Centre for Health Informatics, Australian Institute of Health Innovation, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, Australia
[b] Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, Australia

ARTICLE INFO

Keywords:
Bayesian neural network
Feature uncertainty
Feature ensemble
Anxiety disorder predictions

ABSTRACT

Anxiety disorders are common among youth, posing risks to physical and mental health development. Early screening can help identify such disorders and pave the way for preventative treatment. To this end, the Youth Online Diagnostic Assessment (YODA) tool was developed and deployed to predict youth disorders using online screening questionnaires filled by parents. YODA facilitated collection of several novel unique datasets of self-reported anxiety disorder symptoms. Since the data is self-reported and often noisy, feature selection needs to be performed on the raw data to improve accuracy. However, a single set of selected features may not be informative enough. Consequently, in this work we propose and evaluate a novel feature ensemble based Bayesian Neural Network (FE-BNN) that exploits an ensemble of features for improving the accuracy of disorder predictions. We evaluate the performance of FE-BNN on three disorder-specific datasets collected by YODA. Our method achieved the AUC of 0.8683, 0.8769, 0.9091 for the predictions of Separation Anxiety Disorder, Generalized Anxiety Disorder and Social Anxiety Disorder, respectively. These results provide initial evidence that our method outperforms the original diagnostic scoring function of YODA and several other baseline methods for three anxiety disorders, which can practically help prioritizing diagnostic interviews. Our promising results call for investigation of interpretable methods maintaining high predictive accuracy.

1. Introduction

Anxiety disorders is a common mental problem among youth and adolescents [42]. Inappropriate or late discovery and treatment of such disorders can severely affect the individual's wellbeing. Presently, cognitive behavioural therapy (CBT) is an established treatment for anxiety disorders owing to its first line intervention and efficacy [7,18,21]. Although CBT treatment are efficacious, many anxious youths may not receive proper treatment [35], partially attributable to the lack of accessible and efficient diagnostic instruments. For instance, individuals may be uncertain whether they have an anxiety disorder and need treatment [29].

In this work, we focus on predictions of three disorders: separation anxiety disorder (fear of separating from attachment figures), generalized anxiety disorder (excessive anxiety and worry across a range of domains), and social anxiety disorder (fear of negative evaluation by others during social situations). For these, we propose a novel Bayesian Neural Network (BNN) approach to address the anxiety disorder prediction problem. The proposed model is trained and evaluated using datasets collected by YODA [34]. Instead of using the raw data, we exploit the features selected from the data for training and inference. The features can be considered as indicative representatives of the raw data and are selected by feature selection algorithms.

However, a single set of features may not be informative and diversified enough. Hence, we develop and study a Feature Ensemble based Bayesian Neural Network (FE-BNN) that leverages the advancements in BNNs to integrate an ensemble of feature sets. We exploit the Markov Chain Monte Carlo approximation to sample a large number of features from the latent space. As a result, the inference of the model can be interpreted as averaging the predictions made by each feature set in the sampled ensemble. The ensemble feature set achieves high predictive accuracy for all three disorders.

* Corresponding author.
E-mail address: hao.xiong@mq.edu.au (H. Xiong).

https://doi.org/10.1016/j.jbi.2021.103921
Received 19 April 2021; Received in revised form 16 September 2021; Accepted 29 September 2021
Available online 8 October 2021
1532-0464/

1.1. Related work

The instruments currently used to assess youth disorders include the Anxiety Disorders Interview Schedule (ADIS) [45], Development and Wellbeing Assessment (DAWBA) [16], Diagnostic Interview Schedule for Children (DISC) [10,22], DISC screeners [50], and the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) [24]. These instruments facilitate informal interviews, devised to identify the disorder symptoms. Normally, these instruments are administered via personal interaction between a clinician and a child/parent. Instead of face-to-face interaction, conducting phone assessment is a viable alternative [33].

However, traditional diagnostic procedures often require hours of a qualified clinical psychologist's time, to accurately assess and diagnose a patient. As a result, the concept of online administration of diagnostic tools has been proposed. Online tools can be partially automated, are easy to access, diminish interviewer- and clinician-associated errors, and alleviate limitations related to the clinician availability [6,25,28,49,51]. To increase the diagnosis validity, online instruments often factor in additional clinical reviews, which still reduce the load on the clinicians [15,47].

As the accuracy of the online instruments is often inferior to that of the traditional assessments [34], previous works focused on machine learning methods for improving their accuracy. For instance, to predict generalized anxiety disorder, Support Vector Machine was applied to multimodal bio-behavioural data [19], and Logistic Regression and Bayesian Network were applied to heart rate variability data [5]. Besides, several works applied common machine learning classifiers to predict post-traumatic stress disorder and social anxiety disorder [9,23,37,39,44,53]. Other methods also employed feature selection to extract features out of the raw data and then feed the selected features into the classifiers, to further improve the accuracy [23,37,39,44,56,57]. In addition, [41] presented an extensive review of machine learning methods for anxiety disorder predictions. In this work, we exploit the recent advancement in deep learning and propose a Bayesian ensemble approach to enhance the accuracy of the predictions.

1.2. Significance and objectives

Anxiety disorders are frequently comorbid with depressive disorders or chronic physical health problems. If untreated, anxiety disorders are likely to have adverse effects on a patient's (and their family's) functioning and quality of life, and induce increasingly high health and social costs. Hence, it is of a paramount importance to diagnose and treat anxiety disorders early. In this work, we investigate the ability of ensemble-based deep learning methods to enhance the accuracy of anxiety disorder predictions in youths. The proposed Bayesian deep learning approach exploits an ensemble of more informative and diverse data features. These have the potential to produce more accurate predictions of anxiety disorders than several baseline methods, such as the currently deployed scoring function, as well as machine learning and deep learning baselines.

In practice, our approach would allow clinicians to predict patients' anxiety scores using the data prospectively collected by the online tool. The predicted scores can be leveraged to (i) prioritize diagnostic clinical interviews, such that patients with higher predicted scores are interviewed first and (ii) focus the interviews on the types of anxiety most likely to be a problem for the patient. While other methods can also support the prioritization of focus of the interviews, the proposed method is shown to be more accurate than the baselines. This means that the diagnostic clinical interviews would include fewer false positive patients, who do not have anxiety disorders but are misclassified as anxious. Consequently, more accurate predictions and prioritization have the potential to efficiently use the time of clinical psychologists, improve the quality of health care, and reduce the associated costs.

Table 1
Summary of the datasets for the three disorders.

Datasets	Dim assessments × questions	Disorder	Control
Separation Anxiety Disorder	69 × 19	39	30
Generalized Anxiety Disorder	171 × 32	76	95
Social Anxiety Disorder	132 × 28	58	74

2. Method and materials

In this section, we describe the YODA datasets, formalize the problem, introduce the Feature Ensemble based Bayesian Neural Network, and present our experimental design.

2.1. Data description

McLellan et al. [34] developed an online diagnostic tool, Youth Online Diagnostic Assessment (YODA), to assess and diagnose a range of anxiety disorders in children and adolescents. During the assessment, parents complete multiple screening questionnaires, containing questions pertaining to anxiety disorders their children may have. Screening questions lead each diagnostic category, with negative responses triggering a skip of the remaining questions for that disorder. YODA responses are scored based on an automated scoring function aligned with the DSM (Diagnostic and Statistical Manual) criteria [34].

The YODA tool originally helped to diagnose seven anxiety disorders and was deployed by the Centre for Emotional Health at Macquarie University. Some disorders did not have a sufficient number of diagnosed participants or completed assessment questionnaires to warrant the development and training of machine learning methods. Hence, our work focuses on three specific anxiety disorders, which have sufficient training data available. These are Separation Anxiety Disorder, Generalized Anxiety Disorder, and Social Anxiety Disorder.

In our evaluation, each of the above three anxiety disorders was considered as a standalone dataset. This, we evaluate the performance of the developed methods using three datasets that are summarized in Table 1. The 'Dim' column presents the data dimensionality for each disorder: number of completed assessments (broken down into the Disorder and Control cases shown in the next columns) and the number of questions for the disorder (varies across the disorders). The Disorder/Control status is determined by the clinician's decision following the face-to-face diagnostic interview. The three datasets include a total of 297 participants, all children or adolescents aged 6 to 16. The mean age was 9.34 years and 45.8% were male.

The parents of the children completed the diagnostic assessments of YODA. The assessments contained questions addressing various symptoms, e.g., "*does the child get distressed when he/she needs to separate from particular family members or home*?". The answers used either binary responses or severity/frequency scales. Note that the parents might have provided inaccurate responses due to misinterpreting the questions or just being unaware of the symptoms. Likewise, when assessing the severity of a symptom, the parents inherently introduced noise into the collected data. The proposed FE-BNN method addresses the uncertainty and noises associated with self-reported data.

2.2. Problem formulation

We focus on diagnosis of three disorders: separation anxiety disorder, generalized anxiety disorder, and social anxiety disorder. The problem can be cast as a binary classification task, which receives an input describing the symptoms of a patient and predicts whether an individual has the disorder. An overview of the proposed classification method is illustrated in Fig. 1. We initially introduce three key concepts: input, feature, and prediction.

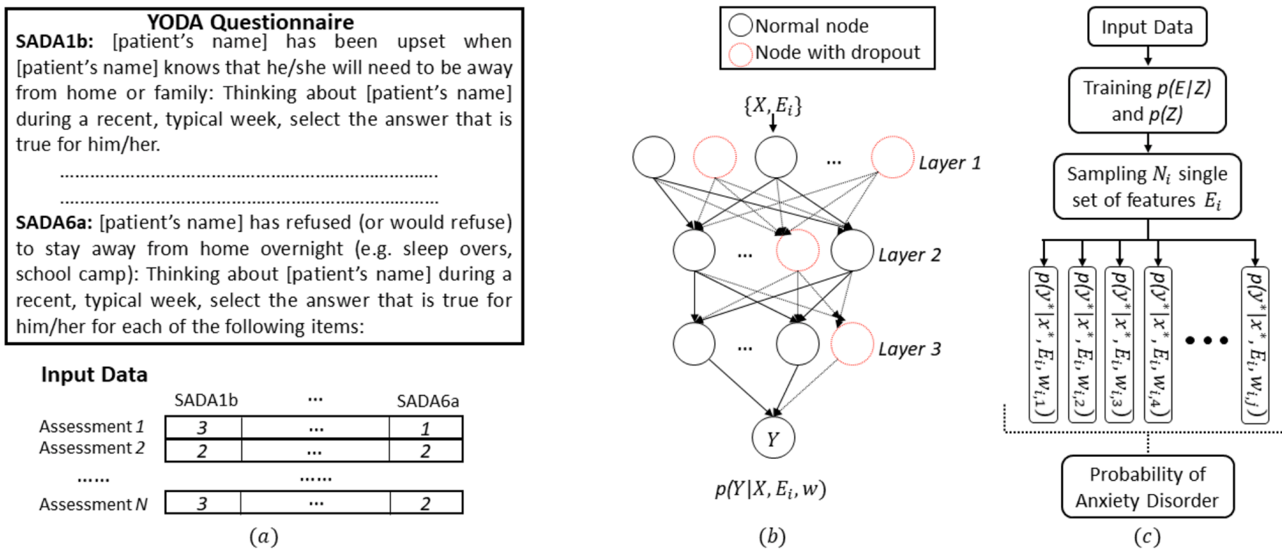

Fig. 1. Pipeline of the proposed anxiety disorder predictions. (a) Sample questions in the YODA questionnaire and pre-processed input data. (b) Network architecture (different anxiety disorders have a different number of layers). (c) Ensemble inference process deployed for anxiety disorder predictions.

(1) **Input:** For anxiety disorder assessments, consider a questionnaire consisting of D questions answered by the patient (or their parent). The questions refer to various symptoms of the disorder and their severity. The answers to the questions can be encoded as either binary 'yes/no' or symbolic values on a predefined scale, e.g., 5 values ranging from 'very little' to 'very much'. Upon a patient answering the D questions, their input is represented as $x^{1\times D}$. Suppose there are N patients answering the questionnaire; then, the entire input is represented as $X^{N\times D}$.

(2) **Feature:** It is common for some information in x to be noisy or redundant. Consequently, feature selection is performed, in order to extract a subset of most informative features out of the raw input data. We use a binary vector $c^{1\times D}$ to denote the elements of x that are selected, such that $|c| = |x|$. The values that are set to 1 in c indicate that their corresponding elements in x are selected. The selected features will be considered as predictors for our model.

(3) **Prediction:** Given the input x and selected features c, the goal of the model is to estimate the predicted output variable $y \in \{0,1\}$. In the context of anxiety disorder predictions, y is a binary variable indicating whether the patient is predicted to be diagnosed with a disorder by the clinician. For the cohort of N patients, $Y^{N\times 1}$ represents the diagnostic outcome for each anxiety disorder.

2.3. Bayesian neural network

We utilize Bayesian Neural Network (BNN) as our backbone. BNN incorporates the uncertainty of the weights into the deep neural network and derives the Bayesian equivalent of the network. As such, it places the prior probability over the model weights ω, such that $\omega \sim p(\omega)$, where $p(\omega)$ is the standard Gaussian prior.

Suppose f^{ω} is the output of a network. Then, the anxiety disorder predictions can be cast as a classification task with the following likelihood model:

$$p\left(Y|X,\omega\right) = \prod_{i=1}^{N} p\left(y_i|x_i,\omega\right) = \prod_{i=1}^{N} softmax\left(\mathbf{f}^{\omega}\left(x_i\right)\right), \tag{1}$$

where $y_i \in \{0,1\}$ (0 = normal, 1 = disorder). Predictive distribution $p(y^*|x^*, X, Y)$ infers the output y^* given testing input x^* and training data X, Y. As per [13], it is defined as:

$$p\left(y^*|x^*,X,Y\right) = \int p\left(y^*|x^*,\omega\right)p\left(\omega|X,Y\right)d\omega. \tag{2}$$

However, posterior distribution $p(\omega|X,Y)$ renders the integral in Eq. 2 intractable. Thus, variational distribution $q(\omega)$ is introduced to approximate $p(\omega|X,Y)$ and make Eq. 2 tractable. Then, the task at hand becomes minimization of the Kullback–Leibler (KL) divergence between $q(\omega)$ and $p(\omega|X,Y)$:

$$KL\left(q\left(\omega\right)||p\left(\omega|X,Y\right)\right) = -\int q\left(\omega\right)\log\frac{p(\omega|X,Y)}{q(\omega)} \tag{3}$$

According to the Bayes theorem, $p\left(\omega\middle|X,Y\right) = \frac{p(Y|X,\omega)p(\omega)}{p(Y|X)}$, such that Eq. 3 can be re-written as:

$$KL(q(\omega)||p(\omega|X,Y)) = \\ -\int q\left(\omega\right)\log p\left(Y\middle|X,\omega\right) + KL\left(q\left(\omega\right)\middle|\middle|p\left(\omega\right)\right) + const \tag{4}$$

Then, Eq. 4 can be approximated as a dropout network with a penalty term.

$$KL\left(q\left(\omega\right)||p\left(\omega|X,Y\right)\right) = \underbrace{-\sum_{i=1}^{N}\log p\left(y_i\middle|x_i,\widehat{\omega}_i\right)}_{\text{network with dropout}} + \underbrace{||\omega||^2}_{\text{penalty}} + const. \tag{5}$$

As mentioned, ω refers to the parameters of the whole network. By applying the dropout, some parameters will be dropped. Therefore, $\widehat{\omega}_i$ denotes the network parameters remaining after the dropout.

2.4. Feature ensemble based Bayesian neural network

As the raw input data is self-reported, thus, potentially noisy or containing redundant information, features selected out of the raw input can alternatively be utilized. This allows various classifiers exploiting the selected features to achieve more accurate classification. We illustrate how to exploit the features using the framework of BNN for anxiety disorder predictions.

Suppose, for input X, we have the binary vector c indicating the selected features. Now, the BNN is $p(\omega|Y,X,c)$. Given input x in X, its corresponding features are selected by c. Hence, the training of BNN relies on the labels Y and a subset of features of X selected by c. Accordingly, the predictive distribution becomes:

$$p\left(y^*|x^*, c^*, X, Y, c\right) = \int p\left(y^*|x^*, c^*, \omega\right) p\left(\omega|X, Y, c\right) d\omega. \tag{6}$$

Here, c^* is the binary vector indicating the position of key features in test input x^*. As can be seen, Eq. 6 does not change the standard BNN, as the features of X are used for training and prediction.

However, combining multiple predictive features may be preferred over using a single feature set. This is due to the observation that one set of features can only describe x from a single, potentially restricted perspective. In contrast, multiple feature sets can surface more informative characteristics of x. Based on this assumption, we propose a *Feature Ensemble based Bayesian Neural Network* (FE-BNN). Thus, we introduce E to denote an ensemble of features and explore the disorder predictions using this ensemble under the framework of BNN. Now, the BNN $p(\omega|X, Y, E)$ and the predictive distribution can be re-written as:

$$p\left(y^*|x^*, X, Y\right) = \int p\left(y^*|x^*, E, \omega\right) p\left(\omega|X, Y, E\right) p\left(E|Z\right) p\left(Z\right) d\omega dE dZ \tag{7}$$

Here, variable Z is a latent variable of E utilized to generate the ensemble of features and prior $p(Z)$ is placed over Z. As the integral in Eq. 7 is intractable, Markov Chain Monte Carlo is used to approximate it [2,31,36]. Here, the Monte Carlo approximation of Eq. 7 is:

$$p\left(y^*|x^*, X, Y\right) \approx \frac{1}{N_i}\sum_{i}^{N_i}\frac{1}{N_j}\sum_{j}^{N_j} p\left(y^*|x^*, E_i, \omega_{i,j}\right). \tag{8}$$

In Eq. 8, N_i samples of latent variables Z_i are firstly drawn from $p(Z)$. For each Z_i, its corresponding E_i is sampled from $p(E|Z)$. In similar to c, E_i is a binary vector used to extract features from input. Lastly, N_j weight matrices $\omega_{i,j}$ are sampled from $p(\omega|X, Y, E_i)$, standard BNN accepting a single set of features extracted via E_i as an input. In Eq. 8, $\frac{1}{N_j}\sum_j^{N_j} p\left(y^*\middle|x^*, E_i, \omega_{i,j}\right)$ represents the prediction using a single set of features. Essentially, there are N_i individual BNN models trained for the corresponding N_i feature sets. Thus, Eq. 8 averages the predictions generated using the ensemble of features, that is FE-BNN. This resembles a voting process, where the final prediction depends on the majority of predictions, each using a single set of features.

We set $N_j = 500$ for all the anxiety disorders. As for N_i, we set it to 69, 94 and 87 for the separation anxiety disorder, generalized anxiety disorder and social anxiety disorder, respectively. In theory, larger N_i and N_j values facilitate more samples and more accurate predictions. However, we found in offline experiments that the predictive accuracy hardly improves when N_i and N_j are greater than the above values. For example, for the separation anxiety disorder, N_i=69 yields the optimal AUC value reported below. For N_i=74 AUC improves by 0.05% only, while the training time increases by 7.96%. Thus, the selected values of N_i and N_j offer the best trade-off considering both the training time and predictive accuracy.

2.4.1. Derivation of $p(E|Z)$ and $p(Z)$

The distribution of $p(E|Z)$ and $p(Z)$ can be obtained by auto-encoding variational Bayes [27], where the marginal distribution of E is:

$$p\left(E\right) = \int p\left(E|Z\right) p\left(Z\right) dZ. \tag{9}$$

Here, prior $p(Z)$ is a centered isotropic multivariate Gaussian $N(Z; 0, I)$ and $p(E|Z)$ is a multivariate Gaussian, the distribution parameters of which are derived from Z using a fully-connected neural network with one hidden layer. As the integral in Eq. 9 is intractable, a variational distribution $q(E|Z)$ is introduced. Thus, the log likelihood of $p(E)$ is:

$$log(p(E)) = KL(q(Z|E)||p(Z|E)) + \mathscr{L}(E). \tag{10}$$

Here, $q(Z|E)$ and $p(E|Z)$ are called encoder and decoder, respectively. Besides, $q(Z|E) \sim N(\mu, \sigma^2 I)$ is a multivariate Gaussian with a diagonal covariance structure. μ and σ are the outputs of a multilayer perceptron, and $\mathscr{L}(E)$ is the variational bound of E represented as:

$$\mathscr{L}\left(E\right) = -KL\left(q\left(Z|E\right)||p\left(Z\right)\right) + \int q\left(Z|E\right) logp\left(E|Z\right) dZ. \tag{11}$$

The variational bound $\mathscr{L}(E)$ is intractable and should be approximated by Markov Chain Monte Carlo. Then, the parameters of $q(Z|E)$ and $p(E|Z)$ are optimized using Stochastic Gradient Descent (SGD) [43]. A detailed implementation of auto-encoding variational Bayes is presented in [27].

Algorithm 1. Feature Ensemble Based Bayesian Neural Network (FE-BNN)

Input: x^*, y^*, X, Y
Output: $p(y^*|x^*, X, Y)$
1. **Initialization:** train variational auto-encoder to initialize $p(E|Z)$ and $p(Z)$.
2. **for** i = 1:N_i **do**
3. sample $Z_i \sim p(Z)$
4. sample $E_i \sim p(E|Z_i)$
5. **for** j = 1:N_j **do**
6. sample $\omega_{i,j}$ with dropout by training BNN over the feature extracted by E_i
7. **end for**
8. **end for**
9. approximate $p(y^*|x^*, X, Y)$ using Eq. 8

In our case, the features selected by a feature selector can be used to train the auto-encoder. We experimented with several feature selection methods and chose Lasso due to its strong performance [48]. By adjusting the threshold in Lasso, different features can be selected from a single input. After training, the obtained $p(E|Z)$ and $p(Z)$ can be used to generate samples used in Eq. 8. The complete FE-BNN method is outlined in Algorithm 1.

2.4.2. Implementation details

The proposed FE-BNN method was implemented with PyTorch [40]. It consists of one input layer, one output layer, and hidden layers. The size of the input layer depends on the input (i.e., number of questions/features) and the size of the output layer is 2 (binary prediction for each disorder). The number of hidden layers depends on the target disorder (one for the separation anxiety disorder, two for generalized anxiety, and one for social anxiety) and this was determined experimentally (results reported below). The size of each hidden layer was fixed to 200. The dropout rates for the input layer and hidden layer were set to 0.2 and 0.5, respectively. The network was optimized using SGD with learning rate of 0.01 [43]. We trained the model using the CPU only.

Note that the feature selection methods may be trained stochastically, which could result in a different set of features being selected. To mitigate this, we performed an exhaustive search on the feature selection methods. Specifically, the feature selection methods use a threshold hyper-parameter to select different sets of features. In general, low threshold values select more features and vice versa. In our experiments, we used variables a and b to define an interval $[a,b]$, which was divided by 100 to obtain 100 thresholds within $[a,b]$. After that, 100 feature sets were selected and fed into the classifiers to find the optimal feature set. Combining with the classifiers, the optimal feature set should induce the best accuracy of the predictions. In the experiments, we observed that 100 feature sets were repetitive and it was sufficient to select a fixed set of features outperforming other feature sets.

3. Results

We evaluate the performance of the proposed method on data pertaining to three disorders-separation anxiety disorder, generalized anxiety disorder, and social anxiety disorder-all collected by YODA [34]. Using the three datasets, we experimentally compare the performance of

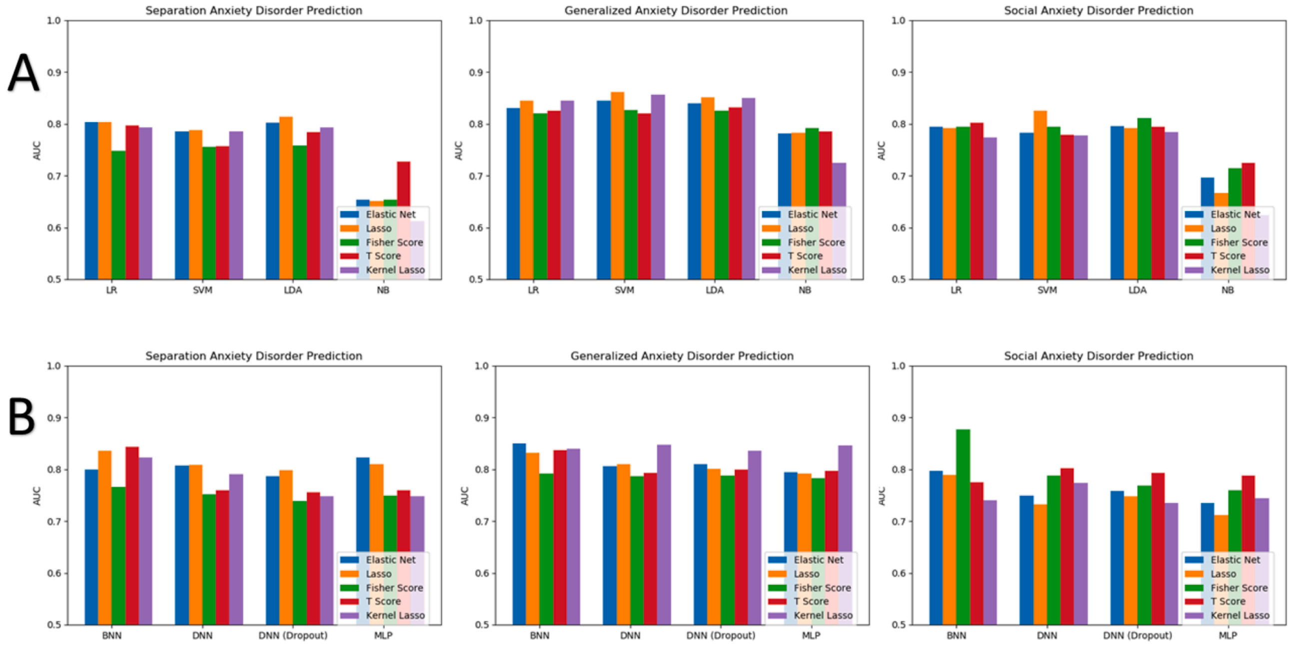

Fig. 2. Combinations of classifiers and feature selectors: A. machine learning based classifiers and feature selectors; B. neural network based classifiers and feature selectors.

our method with a number of baselines, including YODA's scoring function, machine learning based classifiers, deep learning based classifiers, combination of feature selectors and classifiers, and ensemble methods. We also report an in-depth study of the number of hidden layers in the neural network.

3.1. Experimental design

3.1.1. K-Fold cross validation

K-Fold cross validation is commonly used in machine learning, particularly useful for evaluating methods on datasets with a limited sample size. *K*-Fold cross validation uses parameter *K* denoting the number of groups that the data set is split into. The experimental procedure is as follows: (1) the dataset is randomly split into *K* folds; (2) one fold is considered as a test data and the rest as a training data; (3) the model is trained and tested accordingly; (4) the process is repeated for other test folds; and (5) performance is averaged across the *K* folds. Here, we set $K = 5$.

3.1.2. Predictive algorithms

We evaluate the performance of the proposed FE-BNN approach for anxiety disorder predictions against several baseline methods:

- The original YODA scoring function [34].
- *Classifiers* refers to the Logistic Regression (LR) [20], Support Vector Machine (SVM) [3], Linear Discriminant Analysis (LDA) [12,30], Naïve Bayes (NB) [52], Multi-Layer Perceptron (MLP) [1], Deep Neural Network (DNN) and Bayesian Neural Network (BNN) [13] classifiers. LR, SVM, LDA, and NB are traditional classifiers, whilst MLP, DNN and BNN are deep learning based classifiers. Here, DNN is a non-Bayesian variant of BNN and has two versions: one without dropout (denoted as DNN) and another one with dropout (denoted as DNN (Dropout)). Note that the dropout in DNN (Dropout) can avoid the potential overfitting, and does not act as Bayesian approximation to perform Monte Carlo integration. BNN is a Bayesian approach that uses dropout for Bayesian approximation.
- *Classifiers + FS* first deploy feature selectors to select the most informative features from the data and then feed these into the classifiers for training and prediction. The evaluated feature selection methods include ElasticNet [55], Lasso [48], Kernel Lasso [32], T Score [54] and Fisher Score [17]. We exhaustively evaluate all the combinations of these five feature selection methods with the above eight classifiers. Due to space limitations, for every classifier we present only the results achieved by the best-performing feature selector combination.
- *Ensemble Methods* combine the predictions of several estimators in order to improve the generalizability and robustness of one estimator. The evaluated ensemble methods include Random Forest (RF) [4], AdaBoost [8] and Gradient Tree Boosting (GradBoost) [11]. RF builds estimators independently and averages their predictions, while AdaBoost and GradBoost build estimators sequentially to reduce the bias of the combined estimator.

3.1.3. Evaluation metrics

We treat the anxiety disorder predictions as a classification task. Thus, sensitivity (with standard deviation, std), specificity (with std), and Area Under the Curve (AUC, with std) are considered as the evaluation metrics. Sensitivity measures the ratio of participants with a disorder, who are correctly classified and, respectively, specificity measures the ratio of the correctly classified healthy participants. For sensitivity and specificity, we use a threshold of 0.5 to discretize the classifiers' output. AUC refers to the area under the sensitivity–specificity curve that communicates how well the model distinguishes between the classes. Higher AUC values indicate a better accuracy of the model.

3.2. Classifier and feature selection combinations

We experimented with 40 combinations of eight classifiers (LR, SVM, LDA, NB, BNN, DNN, DNN (Dropout), MLP) and five feature selection methods (Elastic Net, Lasso, Kernel Lasso, T Score, Fisher Score). In Fig. 2, we present the AUC achieved by these combinations for each of the three disorders. The best combination for each classifier will be benchmarked against the proposed FE-BNN method.

In general, BNN slightly outperformed the other classifiers and NB performed worst across all the disorders. For most classifiers, there was no substantial variability in AUC across the feature selection methods. That said, Lasso generally outperformed other methods and Fisher Score

Table 2
Separation Anxiety Disorder predictions.

Methods		Sensitivity (*std*)	Specificity (*std*)	AUC (*std*)
YODA		0.8205 (0.1760)	0.4333 (0.2816)	0.6269 (0.1812)
Classifiers	LR	0.6923 (0.2066)	0.5000 (0.2863)	0.6427 (0.2651)
	SVM	0.7436 (0.2519)	0.3667 (0.2702)	0.5915 (0.2881)
	LDA	0.6410 (0.2439)	0.6333 (0.3258)	0.6419 (0.2644)
	NB	0.7692 (0.1817)	0.4333 (0.2341)	0.6538 (0.2055)
Neural Networks	BNN	0.9231 (0.0571)	0.5000 (0.2989)	0.6794 (0.1682)
	DNN	0.5641 (0.1903)	0.5333 (0.2737)	0.5068 (0.2351)
	DNN (Dropout)	0.5897 (0.1384)	0.5000 (0.2702)	0.5017 (0.2181)
	MLP	0.4872 (0.1588)	0.5667 (0.2887)	0.5350 (0.2470)
Classifiers + FS	LR + Lasso	0.8462 (0.0918)	0.5333 (0.2409)	0.8030 (0.1353)
	SVM + Lasso	0.6923 (0.1828)	0.7333 (0.1807)	0.7876 (0.0827)
	LDA + Lasso	0.8462 (0.0918)	0.7000 (0.1021)	0.8137 (0.1024)
	NB + T Score	0.8974 (0.1817)	0.3000 (0.2341)	0.7274 (0.2055)
Neural Networks + FS	BNN + T Score	0.7949 (0.1510)	**0.8667 (0.1393)**	0.8436 (0.1192)
	DNN + Lasso	0.7436 (0.1828)	0.6333 (0.1021)	0.8090 (0.0914)
	DNN (Dropout)+ Lasso	0.7692 (0.1828)	0.5667 (0.0908)	0.7982 (0.1290)
	MLP + Elastic Net	0.8718 (0.1828)	0.6000 (0.1636)	0.8226 (0.0719)
Ensemble Methods	RF	0.6410 (0.1967)	0.6333 (0.1693)	0.7368 (0.1809)
	AdaBoost	0.6410 (0.2179)	0.5667 (0.1714)	0.6077 (0.1473)
	GradBoost	0.7692 (0.0768)	0.5667 (0.3021)	0.7167 (0.1026)
FE-BNN		**0.9744 (0.0571)**	0.6667 (0.3429)	**0.8683 (0.1248)**

Table 3
Generalized Anxiety Disorder predictions .

Methods		Sensitivity (*std*)	Specificity (*std*)	AUC (*std*)
YODA		0.6974 (0.1306)	**0.8737 (0.0874)**	0.7855 (0.0731)
Classifiers	LR	0.7105 (0.0782)	0.7579 (0.0867)	0.7967 (0.0260)
	SVM	0.6711 (0.1145)	0.8105 (0.1076)	0.8133 (0.0490)
	LDA	0.7632 (0.0881)	0.7579 (0.1177)	0.8018 (0.0377)
	NB	0.7105 (0.1376)	0.7789 (0.1084)	0.7849 (0.0289)
Neural Networks	BNN	0.8026 (0.1219)	0.8421 (0.1116)	0.8360 (0.0271)
	DNN	0.6579 (0.1641)	0.7579 (0.0958)	0.7744 (0.0440)
	DNN (Dropout)	0.6315 (0.1201)	0.7473 (0.1436)	0.7688 (0.0511)
	MLP	0.6053 (0.1088)	0.8421 (0.1095)	0.7830 (0.0378)
Classifiers + FS	LR + Lasso	0.7763 (0.1756)	0.7579 (0.1070)	0.8450 (0.0191)
	SVM + Lasso	0.8289 (0.1010)	0.7684 (0.1133)	0.8613 (0.0178)
	LDA + Lasso	0.7895 (0.1576)	0.8000 (0.0912)	0.8510 (0.0194)
	NB + Fisher Score	0.7237 (0.1376)	0.7895 (0.1084)	0.7917 (0.0289)
Neural Networks + FS	BNN + Elastic Net	0.8289 (0.0643)	0.8105 (0.1089)	0.8493 (0.0417)
	DNN + Kernel Lasso	0.8026 (0.0997)	0.7579 (0.0665)	0.8476 (0.0522)
	DNN (Dropout)+ Kernel Lasso	0.6578 (0.1167)	0.8315 (0.0829)	0.8354 (0.0651)
	MLP + Kernel Lasso	0.6184 (0.1067)	0.8526 (0.0863)	0.8457 (0.0434)
Ensemble Methods	RF	0.7105 (0.1263)	0.8000 (0.1189)	0.8470 (0.0501)
	AdaBoost	0.6711 (0.1078)	0.6842 (0.1786)	0.7537 (0.0969)
	GradBoost	0.6184 (0.0954)	0.7579 (0.1030)	0.7785 (0.0492)
FE-BNN		**0.8289 (0.0881)**	0.8211 (0.1089)	**0.8769 (0.0451)**

was inferior in many cases. Considering the combinations of classifiers and feature selectors, we note that BNN + T Score and BNN + Lasso achieved the best AUC for the separation anxiety disorder, a range of combinations achieved comparably high AUC for the generalized anxiety disorder, and BNN + Fisher Score outperformed other combinations for the social anxiety disorder predictions.

3.3. Performance of FE-BNN and baselines

Tables 2–4 show the sensitivity, specificity, and AUC scores obtained by the evaluated methods for the predictions of separation anxiety disorder, generalized anxiety disorder, and social anxiety disorder, respectively. We compare the proposed FE-BNN with four groups of classifiers: original YODA scoring function, classifiers using all the data (machine learning based and neural networks), combinations of classifiers and feature selectors (only the best performing one for each classifier), and ensemble methods. The highest score for each metric is highlighted in bold.

First, we observe that AUC of the YODA scoring function is inferior to most machine learning methods, justifying the application of machine learning for anxiety disorder predictions. Second, we observe that the addition of feature selection boosts across the board the performance of the classifiers, supporting our initial assumption that the raw data may be noisy and redundant for generating accurate predictions. Averaging the AUC of all the seven classifiers + FS methods and comparing to the AUC of classifiers only, we observe an improvement of 31.89% for the separation anxiety disorder predictions, 5.39%-for the generalized anxiety disorder, and 8.08%- for the social anxiety disorder. In terms of AUC, sensitivity and specificity, the studied ensemble methods are generally comparable to classifiers, but weaker than classifiers + FS.

Most importantly, AUC of the proposed FE-BNN consistently outperforms all the other methods across all three disorders under investigation. Unlike feature-based classifiers, integrating an ensemble of features into a BNN allows to consider a broader and more informative feature set and, as a result, the generated disorder predictions are more accurate. Although FE-BNN outperforms all other baselines for AUC, it should be noted that BNN in combination with T-score and Fisher Score achieves a higher specificity for the separation anxiety disorder and a higher sensitivity for the social anxiety disorder. Also, the YODA scoring function achieves a higher specificity for the generalized anxiety disorder, implying that YODA can accurately rule out disorders for healthy participants. Nevertheless, considering AUC as a single metric reflecting the overarching accuracy, FE-BNN steadily shows its superiority over all the evaluated baselines.

Table 4
Social Anxiety Disorder predictions.

Methods		Sensitivity (*std*)	Specificity (*std*)	AUC (*std*)
YODA		0.6552 (0.0775)	0.7297 (0.1555)	0.6925 (0.0773)
Classifiers	LR	0.6724 (0.0916)	0.7568 (0.1544)	0.7754 (0.1097)
	SVM	0.5172 (0.1263)	0.8243 (0.1369)	0.7174 (0.1174)
	LDA	0.7241 (0.0797)	0.7973 (0.1468)	0.7640 (0.1015)
	NB	0.5345 (0.0478)	0.6892 (0.1254)	0.6966 (0.1058)
Neural Networks	BNN	0.7759 (0.1103)	0.7432 (0.1599)	0.7775 (0.1231)
	DNN	0.6552 (0.1041)	0.7568 (0.1607)	0.7463 (0.0910)
	DNN (Dropout)	0.6034 (0.1352)	0.7432 (0.1483)	0.7491 (0.0801)
	MLP	0.6034 (0.0512)	0.7432 (0.1517)	0.7330 (0.1132)
Classifiers + FS	LR + T Score	0.6897 (0.0916)	0.7838 (0.1544)	0.8027 (0.1097)
	SVM + Lasso	0.7069 (0.0920)	0.7973 (0.1715)	0.8250 (0.0953)
	LDA + Fisher Score	0.7241 (0.0797)	0.8108 (0.1468)	0.8115 (0.1015)
	NB + T Score	0.5172 (0.0478)	0.7973 (0.1254)	0.7248 (0.1058)
Neural Networks + FS	BNN + Fisher Score	**0.8103 (0.0884)**	0.7973 (0.1394)	0.8772 (0.1371)
	DNN + T Score	0.6034 (0.1099)	0.8108 (0.1565)	0.8017 (0.0981)
	DNN (Dropout)+T Score	0.6724 (0.1306)	0.8108 (0.1543)	0.7928 (0.0840)
	MLP + T Score	0.6207 (0.0783)	0.8243 (0.1358)	0.7882 (0.0972)
Ensemble Methods	RF	0.6379 (0.1218)	0.7973 (0.1634)	0.8096 (0.1136)
	AdaBoost	0.5517 (0.1259)	0.7703 (0.0804)	0.7013 (0.0872)
	GradBoost	0.6034 (0.1192)	0.7297 (0.2142)	0.7295 (0.1254)
FE-BNN		0.7586 (0.0799)	**0.8511 (0.1489)**	**0.9091 (0.0876)**

Note that the performance of both BNN and FE-BNN, which relies on the backbone of BNN, is relatively stable. The AUC scores of these methods are mostly close to 0.8 or even higher, alleviating the adverse effects of the limited and noisy self-reported training data. That said, the exact accuracy of these methods still depends on the devised neural network architecture, which is one of the key determinants of performance.

3.4. BNN sensitivity analysis

The number of hidden layers in Neural Network based methods is an important parameter to study. For four classifiers (BNN, DNN, DNN (Dropout), MLP), four best performing classifiers + FS (BNN + FS, DNN + FS, DNN (Dropout)+FS, MLP + FS), and FE-BNN, we modify the number of hidden network layers to assess how this affects the performance of the methods. Fig. 3 shows the AUC scores, with the number of hidden layers gradually increasing from 1 to 5.

Overall, the proposed FE-BNN outperforms the other eight methods for all three disorders. However, the optimal number of hidden layers fluctuates across the disorders. The general trend observed is that the AUC of FE-BNN is inversely correlated with the number of hidden layers. This can be explained by the fact that a larger number of hidden layers increases the number of network parameters, which boosts the likelihood of overfitting thus, degrading the predictive accuracy of FE-BNN.

4. Discussion

As shown in Tables 2–4, the performance of neural networks is similar to that of other machine learning based classifiers. At the first sight, it seems that deep learning methods do not improve the predictions of anxiety disorders. However, anxiety disorders predictions benefit from the Bayesian variant of neural network, which considers the uncertainty of the network weights. For the separation anxiety and social anxiety disorders, BNN with feature selection outperforms both machine learning based and deep learning based baselines. It has been found that predictions of Bayesian networks considering such an uncertainty are more accurate than those determinant predictions of non-Bayesian, ordinary networks [13,14,26,38].

Other machine learning based probabilistic models, such as Gaussian processes, may also be applied to the disorder classification problem. However, since we consider feature ensembles in the probabilistic model, the inference process involves several integration operations. Integrating the variables in Eq. 4 is not straightforward, as these variables appear non-linear. To the best of our knowledge, the dropout as approximation [13] is the most practical way to address the integration issue in BNNs, especially compared to other probabilistic models. Another benefit of adding dropout to the neural network is that our model becomes more robust to overfitting [46], which is likely to occur in small datasets like YODA. Hence, we choose BNN as the backbone, on which we build our feature ensemble based method.

We observe in Tables 2–4 that the feature selection based methods in the Classifiers + FS and Neural Network + FS groups generally outperform other baseline methods. These methods exploit key features selected for the predictions, which eliminates redundant and noisy features and hence improves the predictive accuracy. The proposed ensemble FE-BNN is in essence similar to those feature selection based methods. In Eq. 8, we sample N_i times to obtain the ensemble of features. Suppose that each sample contains m features and the total number of sampled features is mN_i. A feature may be selected $k \geqslant 0$ times within the

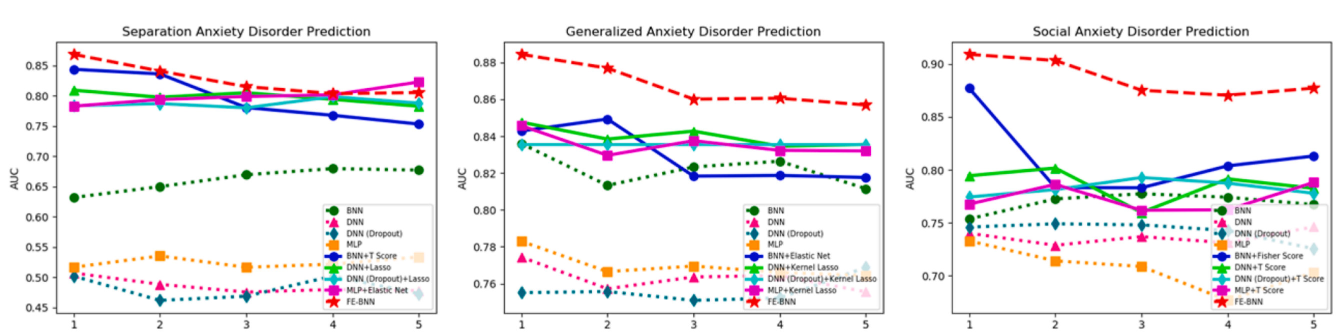

Fig. 3. Impact of the number of hidden layers in deep learning based methods on AUC.

mN_i features and we quantify its weight by $k/(mN_i)$, which communicates the frequency of a selected feature in the ensemble. As a result, features with higher frequency (and weight) will be considered as more important by and contribute more to the predictions. This differentiates FE-BNN from those feature selection methods, which treat the selected features equally important. We posit this is the main reason for the observed strong performance of FE-BNN. We demonstrate the ensemble of features used by FE-BNN and their respective weights in the supplementary material.

Despite achieving promising results, our work is not without limitations. *First*, we exploited three datasets collected by YODA, focusing on Australian population, and addressing three anxiety disorders. Further experiments are needed to validate replicability and generalizability of our method on other populations and disorders. Such a work is, however, hampered by a lack of detailed and validated public datasets in the mental health space. Collating and preparing a dataset like YODA is a substantial task, especially considering the individual diagnostic interviews that need to be conducted by a clinical psychologist for every patient. The datasets exploited in this work took several years to collect and other public datasets of this kind are unfortunately not available. However, the data was collected in a common urban setting, with a population of an average socio-economic level, and the data collection relied of questionnaires, which is a common practice for online diagnostic procedures. We posit that the developed methods are likely to be applicable to similar populations and accurately predict other anxiety disorders and mental conditions.

Second, the proposed method is currently a black box that takes the YODA questionnaire responses as an input and directly outputs the disorder predictions. Other than the predictions themselves, our model cannot point out specific parts of the questionnaire that are the key determinants of the predicted diagnosis. This may be a limitation for translation into clinical practice, as psychologists may be willing to inspect and focus the interviews on specific patient answers when validating the diagnosis. In the supplementary material, we demonstrate the ensembles of features utilized by our method and compare them with the features selected by the feature selection algorithms. While features that occur frequently in the ensemble are potentially more important, exploring the explainability of the predictive model itself is beyond the scope of our work.

Predictions generated in this study are sought to be exploited for prioritizing the diagnostic interviews rather than making clinical decisions. Hence, the interviews of the predicted true positives are supposed to be prioritized over those of the true negatives. However, false positives mean that patients without disorders will be prioritized and interviewed sooner, while false negatives mean that the interviews of patients with disorders will be deferred. While sensitivity is deemed more important generally, deciding on the appropriate trade-off and prioritizing either sensitivity or specificity depends on the volume of patients in a clinic and availability of psychologists. Despite this, such prioritization will not affect the final diagnosis, since the clinicians will still interview the patients and make the decisions.

5. Conclusions

In this work, we focused on the predictions of psychological anxiety disorders in youth. Although automated online systems have been developed, their accuracy may not be sufficiently high to warrant clinical translation. To improve the predictive accuracy, we proposed and evaluated FE-BNN, a Feature Ensemble based Bayesian Neural Network. We evaluated FE-BNN on three disorder-specific datasets collected by the online YODA tool and benchmarked it against a collection of machine learning methods and the original scoring function of YODA. The obtained results clearly demonstrate the superiority of FE-BNN over the evaluated methods, not requiring the involvement of clinical psychologists.

In the future, we primarily intend to investigate how the proposed ensemble approach can be made more interpretable and explainable, so that it can highlight the key features and questions affecting the attained outcome. Clinicians will directly benefit from such an explainability, as this can practically save time and increase the transparency of machine learning methods deployed for clinical decision-support in mental health applications.

CRediT authorship contribution statement

Hao Xiong: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Shlomo Berkovsky:** Supervision, Formal analysis, Writing - review & editing. **Mia Romano:** Data curation, Resources. **Roneel V. Sharan:** Formal analysis. **Sidong Liu:** Formal Analysis. **Enrico Coiera:** Writing – review & editing. **Lauren F. McLellan:** Data curation, Resources, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknolwedgement

This work was supported by an Australian Rotary Health Research Grant, the NHMRC Centre of Research Excellence in Digital Health and the NHMRC Partnership Centre for Health System Sustainability.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jbi.2021.103921.

References

[1] L.B. Almeida, Multilayer perceptrons. Handbook of Neural Computation, IOP Publishing Ltd and Oxford University Press, 1997.
[2] A.K. Balan, V. Rathod, K.P. Murphy, M. Welling, Bayesian dark knowledge, in: Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015, pp. 3438–3446.
[3] A. Bogomolov, B. Lepri, M. Ferron, F. Pianesi, A.S. Pentland, Daily stress recognition from mobile phone data, weather conditions and individual traits, in: 22nd ACM International Conference on Multimedia, 2014, pp. 477–486.
[4] L. Breiman, Random forests, Machine Learn. 45 (2001) 5–32.
[5] M. Chatterjee, G. Stratou, S. Scherer, L.P. Morency, Context-based signal descriptors of heart-rate variability for anxiety assessment, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2014, pp. 3631–3635.
[6] R. Elford, H. White, R. Bowering, A. Ghandi, B. Maddiggan, K.S. John, M. House, J. Harnett, R. West, A. Battcock, A randomized, controlled trial of child psychiatric assessments conducted using video conferencing, J. Telemed Telecare 6 (2000) 73–82.
[7] D.L. Ewing, J.J. Monsen, E.J. Thompson, S. Cartwright-Hatton, A. Field, A meta-analysis of transdiagnostic cognitive behavioural therapy in the treatment of child and young person anxiety disorders, J. Clin. Child Adolescent Psychol. 43 (2015) 562–577.
[8] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1997) 119–139.
[9] A. Frick, M. Gingnell, A.F. Marquand, K. Howner, H. Fischer, M. Kristiansson, S.C. R. Williams, M. Fredrikson, T. Furmark, Classifying social anxiety disorder using multivoxel pattern analyses of brain function and structure, Behav. Brain Res. 259 (2014) 330–335.
[10] P.J. Frick, S.O. Lilienfeld, M. Ellis, B. Loney, P. Silverthorn, The association between anxiety and psychopathy dimensions in children, J. Abnorm. Child Psychol. 27 (1999) 383–392.
[11] J. Friedman, Greedy function approximation: A gradient boosting machine, Ann. Stat. 29 (2001) 1189–1232.
[12] J.H. Friedman, R. Tibshirani, T. Hastie, Section 4.3. The Elements of Statistical, Learning (2008) 106–119.
[13] Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: representing model uncertainty in deep learning, in: 33rd International Conference on International Conference on Machine Learning, 2016, pp. 1050–1059.
[14] Y. Gal, R. Islam, Z. Ghahramani, Deep bayesian active learning with image data, in: 34th International Conference on Machine Learning, 2017, pp. 1183–1192.
[15] R. Goodman, Online child and adolescent mental health surveys can be good enough, Soc. Psychiatry Psychiatr. Epidemiol. 48 (2013) 1317–1325.

[16] R. Goodman, T.J. Ford, H. Richards, The development and well-being assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology, J. Child Psychol. Psychiatry Allied Discip. 41 (2000) 645–655.

[17] Q.Q. Gu, Z.H. Li, J.W. Han, Generalized fisher score for feature selection, in: Twenty-Seventh Conference on Uncertainty in Artificial Intelligence, 2011, pp. 266–273.

[18] C.K. Higa-McMillan, S.E. Francis, L. Rith-Najarian, B.F. Chorpita, Evidence base update: 50 years of research on treatment for child and adolescent anxiety, J. Clin. Child Adolescent Psychol. 45 (2016) 91–113.

[19] K. Hilbert, U. Lueken, M. Muehlhan, K. Beesdo-Baum, Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural mri data: A multimodal machine learning study, Brain Behav. 7 (2009) e00633.

[20] D.W. Hosmer, S. Lemeshow, Applied logistic regression, John Wiley and Sons, 2000.

[21] A.C. James, G. James, F.A. Cowdrey, A. Soler, A. Choke, Cognitive behavioural therapy for anxiety disorders in children and adolescents, Cochrane Database Systematic Rev. (2013).

[22] P.S. Jensen, H.K. Watanabe, J.E. Richters, M. Roper, E.D. Hibbs, A.D. Salzberg, S. Liu, Scales, diagnoses, and child psychopathology: Ii. comparing the cbcl and the disc against external validators, J. Abnorm Child Psychol. 24 (1996) 151–168.

[23] K.I. Karstoft, I. Galatzer-Levy, A. Statnikov, Z.G. Li, A.Y. Shalev, Bridging a translational gap: using machine learning to improve the prediction of ptsd, BMC Psychiatry 15 (2015) 1–7.

[24] J. Kaufman, B. Birmaher, D. Brent, U. Rao, C. Flynn, P. Moreci, D. Williamson, N. Ryan, The schedule for affective disorders and schizophrenia for school-age children, University of Pittsburgh Medical Center, Pittsburgh, 1996.

[25] J. Kaufman, B. Birmaher, D. Brent, U. Rao, C. Flynn, P. Moreci, D. Williamson, N. Ryan, Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (k-sads-pl): initial reliability and validity data, J. Am. Acad. Child Adolesc. Psychiatry 36 (1997) 980–988.

[26] A. Kendall, V. Badrinarayanan, R. Cipolla, Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding, in: Proceedings of the British Machine Vision Conference (BMVC), 2017, pp. 1–12.

[27] D.P. Kingma, M. Welling, Auto-encoding variational bayes, in: International Conference on Learning Representations, 2014.

[28] S.J. Ko, G.A. Wasserman, L.S. McReynolds, L.M. Katz, Contribution of parent report to voice disc-iv diagnosis among incarcerated youths, J. Am. Acad. Child Adolesc Psychiatry 43 (2004) 868–877.

[29] D. Lawrence, S. Johnson, J. Hafekost, K.B. de Haan, M. Sawyer, J. Ainley, S. R. Zubrick, The mental health of children and adolescents, in: Report on the second Australian Childand Adolescent Survey of Mental Health and Wellbeing, 2015.

[30] O. Ledoit, M. Wolf, I. Honey, Shrunk the sample covariance matrix, J. Portfolio Manage. 30 (2004) 110–119.

[31] C.Y. Li, C.Y. Chen, D. Carlson, L. Carin, Preconditioned stochastic gradient langevin dynamics for deep neural networks, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016, pp. 1788–1794.

[32] F.H. Lochovsky, G. Wang, G.Y. Yeung, The kernel path in kernelized lasso, in: Eleventh International Conference on Artificial Intelligence and Statistics, 2007, pp. 580–587.

[33] H.J. Lyneham, R.M. Rapee, Agreement between telephone and in-person delivery of a structured interview for anxiety disorders in children, J. Am. Acad. Child Adolesc Psychiatry 44 (2005) 274–282.

[34] L.F. McLellan, M. Kangas, R.M. Rapee, L. Iverach, V.M. Wuthrich, J.L. Hudson, H. J. Lyneham, The youth online diagnostic assessment (yoda): Validity of a new tool to assess anxiety disorders in youth, Child Psychiatry Hum. Dev. 52 (2021) 270–280.

[35] K.R. Merikangas, J.P. He, M.E. Burstein, J. Swendsen, S. Avenevoli, B. Case, K. Georgiades, L. Heaton, S. Swanson, M. Olfson, Service utilization for lifetime mental disorders in u.s. adolescents: Results of the national comorbidity survey adolescent supplement (ncs-a), J. Am. Acad. Child Adolesc Psychiatry 50 (2011) 32–45.

[36] R.M. Neal, Bayesian learning via stochastic dynamics, in: Proceedings of Advances in Neural Information Processing Systems, 1993, pp. 475–482.

[37] S. Omurca, E. Ekinci, An alternative evaluation of post traumatic stress disorder with machine learning methods, in: IEEE International Symposium on Innovations in Intelligent Systems and Applications, 2005, pp. 1–7.

[38] J.I. Orlando, P. Seebock, H. Bogunovic, S. Klimscha, C. Grechenig, S. Waldstein, B. S. Gerendas, U. Schmidt-Erfurth, U2net: A bayesian u-net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological oct scans, in: IEEE 16th International Symposium on Biomedical Imaging, 2019, pp. 1141–1445.

[39] S.P. Pantazatos, A. Talati, F.R. Schneier, J. Hirsch, Reduced anterior temporal and hippocampal functional connectivity during face processing discriminates individuals with social anxiety disorder from healthy controls and panic disorder, and increases following treatment, Neuropsychopharmacology 39 (2014) 425–434.

[40] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems, 2019, pp. 8024–8035.

[41] E.G. Pintelas, T. Kotsilieris, I.E. Livieris, P. Pintelas, A review of machine learning prediction methods for anxiety disorders, in: International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, 2018, pp. 8–15.

[42] R.M. Rapee, C.A. Schniering, J.L. Hudson, Anxiety disorders during childhood and adolescence: Origins and treatment, Annual Rev. Clin. Psychol. 5 (2009) 311–341.

[43] H. Robbins, S. Monro, A stochastic approximation method, Annals Mathe. Stat. (1951) 400–407.

[44] G.N. Saxe, S.S. Ma, J.W. Ren, C. Aliferis, Machine learning methods to predict child posttraumatic stress: A proof of concept study, BMC Psychiatry 17 (2017) 223.

[45] W.K. Silverman, L.M. Saavedra, A.A. Pina, The anxiety disorders interview schedule for children for dsm-iv: Child and parent versions. Psychological Corporation, 1996.

[46] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from over?tting, J. Machine Learn. Res. 15 (2014) 1929–1958.

[47] M.P. Steenhuis, M. Serra, R.B. Minderaa, C.A. Hartman, An internet version of the diagnostic interview schedule for children (disc-iv): Correspondence of the adhd section with the paper-and-pencil version, Psychol. Assess. 21 (2009) 231–234.

[48] R. Tibshirani, Shrinkage and selection via the lasso, J. Roy. Stat. Soc. 58 (1996) 267–288.

[49] G.A. Wasserman, L.S. McReynolds, C.P. Lucas, P. Fisher, L. Santos, The voice disc-iv with incarcerated male youths: Prevalence of disorder, J. Am. Acad. Child Adolesc Psychiatry 41 (2002) 314–321.

[50] C.F. Weems, D.J. Feaster, V.E. Horigian, M.S. Robbins, Parent and child agreement on anxiety disorder symptoms using the disc predictive scales, J. Abnorm. Child Psychol. 18 (2010) 213–216.

[51] P. West, H. Sweeting, G. Der, J. Barton, C. Lucas, Voice-disc identified dsm-iv disorders among15-year-olds in the west of scotland, J. Am. Acad. Child Adolesc Psychiatry 42 (2003) 941–949.

[52] H. Zhang, The optimality of naïve bayes, in: FLAIRS, 2004, pp. 477–486.

[53] W.J. Zhang, X. Yang, S. Lui, Y.J. Meng, L. Yao, Y. Xiao, W. Deng, W. Zhang, Q. Y. Gong, Diagnostic prediction for social anxiety disorder via multivariate pattern analysis of the regional homogeneity, BioMed Res. Int. (2015).

[54] N. Zhou, L.P. Wang, A modified t-test feature selection method and its application on the hap map genotype data, Genomics Proteom. Bioinformat. 5 (2008) 242–249.

[55] H. Zou, T. Hastie, Regularization and variable selection via the elasticnet, J. Roy. Stat. Soc. (2005) 301–320.

[56] M.D. Nemesure, M.V. Heinz, H. Raphael, N.C. Jacobson, Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence, Sci. Rep. 11 (2021) 1980.

[57] M. Arif, A. Basri, G. Melibari, T. Sindi, N. Alghamdi, N. Altalhi, M. Arif, Classification of Anxiety Disorders using Machine Learning Methods: A Literature Review, Insights Biomed Res 4 (1) (2020) 95–110.