QUESTIONNAIRE SIMPLIFICATION FOR FAST RISK ANALYSIS OF CHILDREN'S MENTAL HEALTH

Kimberly Carpenter,[1] *Pablo Sprechmann,*[2] *Marcelo Fiori,*[3],
Robert Calderbank,[2] *Helen Egger,*[1] *and Guillermo Sapiro*[2]

[1] Psychiatry and Behavioral Science Department; Duke University
[2]Electrical and Computer Engineering Department; Duke University
[3] Instituto de Ingeniería Eléctrica, Universidad de la República, Uruguay

ABSTRACT

Early detection and treatment of psychiatric disorders on children has shown significant impact in their subsequent development and quality of life. The assessment of psychopathology in childhood is commonly carried out by performing long comprehensive interviews such as the widely used *Preschool Age Psychiatric Assessment* (PAPA). Unfortunately, the time required to complete a full interview is too long to apply it at the scale of the actual population at risk, and most of the population goes undiagnosed or is diagnosed significantly later than desired. In this work, we aim to learn from unique and very rich previously collected PAPA examples the intercorrelations between different questions in order to provide a reliable risk analysis in the form of a much shorter interview. This helps to put such important risk analysis at the hands of regular practitioners, including teachers and family doctors. We use for this purpose the alternating decision trees algorithm, which combines decision trees with boosting to produce small and interpretable decision rules. Rather than a binary prediction, the algorithm provides a measure of confidence in the classification outcome. This is highly desirable from a clinical perspective, where it is preferable to abstain a decision on the low-confidence cases and recommend further screening. In order to prevent over-fitting, we propose to use network inference analysis to predefine a set of candidate question with consistent high correlation with the diagnosis. We report encouraging results with high levels of prediction using two independently collected datasets. The length and accuracy of the developed method suggests that it could be a valuable tool for preliminary evaluation in everyday care.

Index Terms— childhood development, boosting, mental health, questionnaire, network analysis

1. INTRODUCTION

There is increasing evidence suggesting that anxiety disorders are present in very young children [1, 2], yet anxiety in the preschool period often goes undetected and untreated [3]. In the current sample, less than 5% of children suffering from an impairing anxiety disorder had been referred for further evaluation or treatment [2]. The ability to quickly and reliably detect and intervene with anxiety disorders in early childhood, during a period when the childs brain is still developing, could directly alter the childs developmental trajectory [4, 5], putting the child at decreased risk for psychiatric illnesses later in life.

The PAPA is a comprehensive interview that assesses symptoms for a range of psychiatric diagnoses in children ages 2-5 [6]. The PAPA uses a highly structured protocol, with required questions and probes. When symptoms are reported, their frequency, duration and dates of onset are also collected, to determine whether they meet the symptom overlap and duration criteria for the various Diagnostic and Statistical Manual (DSM) diagnoses. While shorter measures (e.g. CBCL½-5 [7]) can identify individuals with similar disorders in less time, they do not provide the same level of detail about the symptoms that comprise the disorders. By getting this level of detail, one is able to not only determine the presence or absence of a disorder, but to also understand the normal distribution of severity, frequency, and duration of symptoms across the population, which is imperative for determining where the cut-points between normal variation and clinical symptomatology lie for these disorders. The time required to complete an interview that collects this level of data, however, is extremely time prohibitive and would be impossible to incorporate into everyday clinical care. Thus, there is a need for developing tools that can be used in daily care settings for assessing risk, and then referring the child for the full evaluation when necessary. This way, we expect not to miss the 95% of the population currently missed.

A wealth of data has been collected using the full PAPA interview on children ages 2-5 in pediatric primary care. These interviews have all been conducted in a lab setting where sufficient time is set aside for the completion of the entire interview. This high-value and unique data contains very important information from which meaningful statistical relations between symptomatology and the different disorders can be obtained. In this paper, we propose to use this data in order to move away from the full interview and determine if individual interview items or clusters of items can predict, with high reliability and low false positive rates, that the child is likely to have a specific diagnosis. That way, if there is a subset of the full interview items that, if endorsed, reliably identifies a significant proportion of children who are diagnosed with the disorder based on the full interview, while maintaining the validity of the full interview, then those "high value" items could be prioritized during a shorter clinical interview. This would not only decrease the time it would take to make a risk assessment, but also decrease the amount of training required to complete the interview. This would permit the use of the PAPA, or a subset of it, in a larger range of settings, such as busy pediatric primary care clinics and school counseling departments. Furthermore, by understanding the confidence with which subsets of items predict a specific disorder, we will be able to provide both the parent and clinician with information about how confident they can be that a child does or does not have a disorder. This will also allow the identification of children for whom we have insufficient confidence about their status, who can then be referred to a specialty clinic for the full, comprehensive interview. This will enable us to quantify best clinical practices in a usable, accessible,

Work partially supported by ONR, NGA, NSF, ARO, and AFOSR.

and reproducible way. The obtained data-driven tool is very flexible as it can be naturally revised and updated as more data gets collected or adapted to a specific cultural environment.

2. PRIOR WORK AND PROPOSED METHOD

Given the set of completed interviews with their corresponding diagnosis, our problem can be thought as a classic classification problem in which the feature vectors are a subset of the questions used in the full interviews and the labels are given by the diagnosis.

The problem of finding compact classifiers has been previously studied in the literature. Many works have independently proposed to use alternating decision trees (ADtrees) [8]. This algorithm was first proposed in the context of automatic call classification in spoken language dialogue [9]. The appealing property of this algorithm is that, by combing boosting and decision trees, the obtained classifiers are more compact and easy to interpret. ADtree produces decision trees with considerably fewer nodes (weak learners) than other boosted decision tree algorithms (such as C4.5 [10]) and with no performance degradation. The compactness of this classifier was previously exploited in several medical applications such as early diagnosis of dengue fever [11], modeling disease trait information [12], heart-disease diagnostics [8], and for shortening observation-based screening and diagnosis of autism [13]. In [13] the authors present an extensive experimental evaluation in which ADtrees obtain the best performances while, at the same time, being the most compact alternative.

All the above-mentioned examples apply off-the-shelf ADtree algorithms to their particular case of study. In [13], the classifier used with ADtrees considers a small number of questions only because it is run over a small number of iterations. This is not optimal in the sense that one could keep adding nodes without increasing the number of questions to improve performance. On the other hand, choosing the number of boosting iterations is a difficult task, normally done via cross validation. As was reported in [8], ADtrees are known to overfit the data, and the situation is more delicate for small datasets. This is particularly important when adding a node implies asking a new question. Hence we would like to have an objective measure of when the questions used by ADtrees are statistically relevant for the diagnosis and not just for improving the classification accuracy on the training set.

In this work, in addition to using for the first time the unique data coming from the PAPA, we propose to mitigate these problems by pre-processing the data using network analysis. The idea is to use network inference to first identify the subset of questions that are statistically correlated with the different diagnosis. We then propose to construct the classifier by using a two-step approach:

- We first select a small subset of relevant questions, explained in Section 3.
- We then apply the ADtrees algorithm, briefly described in Section 4, only to those pre-selected questions.

3. NETWORK-BASED QUESTIONS SELECTION

The understanding of the dependencies between questions, and more important the relation between questions and the presence of a disorder, is crucial in order to reduce the size of the questionnaire, producing a shorter interview.

We now characterize and visualize these dependencies in terms of networks or graphs. We consider each question as a node in the graph, adding the diagnosis as a node as well, and we infer an edge between two nodes whenever the corresponding questions are related, according to some criterion. The weight of each edge is in concordance with the strength of the dependency. A very meaningful way to describe these dependencies is in terms of the conditional correlations. We represent the answers to the questionnaire as a matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$, where each column $\mathbf{X}_i$ contains the answers to the question i for n subjects. Then, the weight of the edge between nodes i and j in the graph represents the correlation between $\mathbf{X}_i$ and $\mathbf{X}_j$ given all the rest of the variables $\mathbf{X}_k$, $k \neq i, j$. The absence of an edge means that $\mathbf{X}_i$ and $\mathbf{X}_j$ are conditionally independent.

It is well known that when the data $\mathbf{X}$ follows a Gaussian distribution, the entries of the inverse covariance matrix $\mathbf{\Sigma}^{-1}$ correspond to rescaled conditional correlations. This implies in particular that the sparsity pattern of $\mathbf{\Sigma}^{-1}$ reveals the structure of the graph of conditional correlations. A great number of applications are based on this property, from social sciences and economics to biology [14]. In most of these applications, the underlying graph structure is known to be sparse, and the so-called Graphical Lasso approach [15] has proven to be successful for these cases. The formulation of the Graphical Lasso is the convex optimization problem given below, and it consists on a Maximum Likelihood estimation of $\mathbf{\Sigma}^{-1}$ with a sparsity promoting penalty term:

$$\min_{\mathbf{\Theta}\succ 0} \ \mathrm{tr}(\mathbf{S\Theta}) - \log\det\mathbf{\Theta} + \lambda||\mathbf{\Theta}||_{\ell_1}, \qquad (1)$$

where S is an estimator of the covariance matrix and the ℓ_1 norm is $||\mathbf{\Theta}||_{\ell_1} = \sum_{i,j}|\mathbf{\Theta}_{ij}|$.

Although the previous properties hold for Gaussian distributions, recent results show that similar statements can be made for more general models. For instance, when the variables $\mathbf{X}_i$ follow an arbitrary discrete distribution, under some extra conditions, the inverse covariance matrix reflects the structure of the conditional correlation graph, and the Graphical Lasso program is a consistent method for recovering this structure [16].

The strategy for selecting the most relevant questions is therefore the following. We first build the conditional dependency graph by running the Graphical Lasso pn the questionnaire, including the diagnosis column, and then we look for the nodes connected to the diagnosis node. The use of these conditional correlations is very important, since it gets rid of indirect connections, keeping only the direct dependencies. For instance, suppose that question i is related to question j, but only question i is relevant to the diagnosis. In that case, considering only the conditional dependencies one would select only the question i, although both questions would have positive correlation with the diagnosis.

The problem of selecting the best question subset is very hard, since it is not tractable by exhaustive search due to the exponential size of the set of parts. We tackle this problem by selecting those questions which are directly correlated with the diagnosis, using the training dataset. More specifically, we estimate the empirical covariance matrix $\mathbf{S}$ from the training data, and run the Graphical Lasso (1) to obtain an estimate of the adjacency matrix of the conditional correlation graph, see Figure 1. This graph contains very meaningful information about the dependence between questions. However, in this part we focus on the connections between the questions and the diagnosis node, information which is contained in the column of the adjacency matrix corresponding to the diagnosis. We sort this vector in descending order, see Figure 1 for an example. Due to the inherent sparsity of the graph, and the capacity of the Graphical Lasso to recover this structure, the number of questions with positive conditional correlation with the diagnosis, q, is significantly smaller than the total number of questions, p. We pre-select only the questions

with non zero correlation to the diagnosis to be used by the ADtree algorithm. If $I = \{i_1, \ldots, i_q\}$ is the set of indexes of the selected questions, we denote by $\mathbf{X}_I \in \mathbb{R}^{n \times q}$ the reduced questionnaire.

4. ALTERNATING DECISION TREES

Alternating decision trees are an extension of both, binary decision trees and voted stumps [8]. The structure of the ADtrees classifier is given by a series of decision nodes and prediction layers organized in the form of a tree. The decision nodes are given by very simple rules based on individual variables. In our problem, they can take the form of quantitative comparison such as "the patient feels in a certain way more than x times per week or not", or a simple binary decision such as "the patient presents a give symptom or not". Following each decision node comes a prediction layer that assigns a real-valued coefficient to the each of the binary outputs of the decision rule. Formally, for the t^{th} node in the tree we can define a function $r_t : \mathbb{R}^q \rightarrow \mathbb{R}$ that is zero for responses that do not meet the conjunction of conditions that lead to this particular node in the tree, takes a value a_t if the node's condition is satisfied and a value b_t otherwise. For a given response $\mathbf{x} \in \mathbb{R}^q$, the classification rule is the sign of the sum of the individual node predictions,

$$\text{class}(\mathbf{x}) = \text{sign}\left(\sum_{t=1}^{T} r_t(\mathbf{x})\right), \tag{2}$$

where T is the total number of nodes in the tree. Hence, ADtrees can be thought as a majority vote over a particular type simple prediction rules. ADtrees uses AdaBoost [17] to learn the parameters on each individual classifier. The algorithm iteratively grows the tree one node at a time, see [8, 17] for details. The selection of T to avoid overfitting is a delicate task (particularly for small datasets) [8] and in general is done via cross-validation. Note that the number of questions actually used does not depend on T. It could happen that the algorithm always uses questions from a subset to define the nodes. In our application, asking the questions is costly while adding nodes to the classifier is not, thus, we can restrict the algorithm to use few relevant questions but can increase the number of decision nodes so as to obtain more complex decision rules. The absolute value of the additive expansion in (2) gives a measure of confidence in the classification results. This was observed experimentally [8] and theoretically [18], providing a way to recover the class probability.

5. EXPERIMENTAL EVALUATION

Study Design and Participants: Data was acquired from independent samples of participants recruited as part of two separate studies: *(i)* The Preschool Age Psychiatric Assessment (PAPA) Test-Retest study ($n = 307$) [6] and *(ii)* The Duke Preschool Anxiety Study (PAS) ($n = 917$) [2]. Both studies recruited children aged 2 to 5 years old attending Duke University Pediatric Primary Care Clinics for both well-child and sick-child visits. These clinics care for a diverse population of families, drawn from the city of Durham and the surrounding rural areas of Durham County. Inclusion criteria were *(i)* the child was between 24 and 71 months old and *(ii)* the child attended the pediatric clinic during a screening period. Exclusion criteria were *(i)* the child was not accompanied by a parent/legal guardian who could provide consent, *(ii)* the parent/legal guardian lacked adequate fluency in English to complete the screen, *(iii)* the index child was known to have $IQ < 70$, autism, or other pervasive developmental disorders, *(iv)* the child's sibling was participating in the study, or *(v)* the provider decided that the child was too medically ill at the visit for the parent to be approached about the study.

T-R dataset: The PAPA Test-Retest study included 1,073 parents with children ages 24 to 71 months screened with the CBCL½-5 questionnaire. Children who obtained a T score greater or equal than 55 on the total symptom score of the CBCL½-5, which identifies the top 30% of the general population according to the CBCL½-5 norms [7], were considered screen high. The number of screen highs in our sample was 307, 246 (stratified by age, gender, and race) of which were selected for recruitment in the study. Of this 246, 193 completed the interview phase. A random sample of 20% of parents whose children had a T score smaller or equal than 55 (screen lows) were invited to take part in the test-retest phase of the study. Interviewers were blind to the parent's screen status and to the results from the first interview at the time of the second interview. 307 patients completed both PAPA interviews.

PAS dataset: The Preschool Anxiety Study included 3,433 parents screened using a 10-item anxiety screening measure developed using data from the earlier Test-Retest study from our group [6]. Children were identified as screen positive for anxiety if the parents endorsed 4 or more of the 10 items on this screener. All children who screened high ($n = 944$) and a random sample of 189 who did not screen high were selected to participate in an in-home assessment. Of the 1,132 children selected to participate in the in-home assessment, 917 eligible parents completed the PAPA and checklist assessments about the child's psychiatric symptoms and temperament, the parent's own personality traits, symptoms of anxiety, and depression, and distress in the parent-child relationship. The Duke University Medical Center Institutional Review Board approved both of these studies.

GAD and SAD: We evaluated the proposed approach in two common disorders observed in young children: Generalized Anxiety Disorder (GAD) and Separation Anxiety Disorder (SAD). Anxiety was measured using the anxiety disorders module of the Preschool Age Psychiatric Assessment (PAPA) [6]. Diagnoses and symptom scales are generated by combining the answers of questions in the PAPA interview: $i = 27$ questions for GAD and $i = 51$ questions for SAD. The PAPA includes assessment of most DSM diagnostic criteria in sofar as they are relevant to younger children, plus all items in the Diagnostic Classification: 0-3R (Zero to three diagnostic classification system, Washington, DC, 2005). Impairment due to anxiety symptoms was also assessed using the World Health Organization's International Classification of Functioning, Disability and Health (2001). Detecting the cases in which both GAD and SAD cause impairment is of high relevance.

Question selection and classification: We trained several classification trees for both GAD and SAD in different settings. *ADnet* refers to result obtained with the method proposed in this paper, which restricts the ADtrees algorithm to choose questions from the ones pre-selected via network analysis. We refer simply as *ADtree* when the algorithm can choose question with no restrictions. As explained in Section 4 , we selected via cross-validation the number of nodes in the tree which is independent from the number of questions pre-selected by the networks. We consider pairs (n_Q, n_N), where n_Q is the number of questions used by the algorithm and n_N is the number of nodes composing the decision tree. We studied the case of binary classification and a second situation in which we omitted a decision for those tested samples with low confidence. We defined a result as low confidence if the probability of belonging to the chosen class was smaller than 0.8. As a benchmark we compared the obtained classifiers against the C4.5 algorithm [10]. If we do not restrict the number of questions, C4.5 would in general use

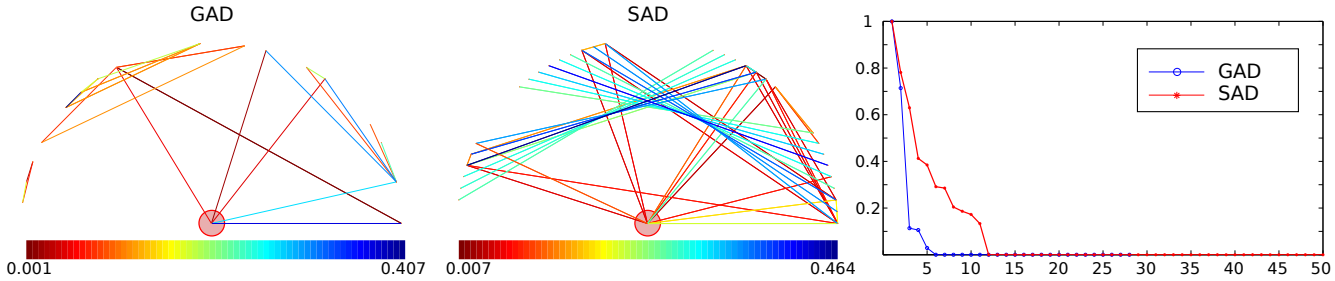

Fig. 1: Network inference. Estimated conditional correlation graphs for GAD (left) and SAD (center) datasets. The center node corresponds to the diagnosis, and questions are distributed in the half-circle. Right: decay of the conditional correlation coefficients between all questions with the diagnosis.

Table 1: Performance comparison for different algorithms for predicting GAD, measured in terms of False Positive Rate (FPR), True Positive Rate (TPR), Accuracy (Acc), and size of the low confidence set (LC). All values are percentages.

Method	GAD binary			GAD conf.			
	FPR	TPR	Acc.	FPR	TPR	Acc.	LC
No imp.							
C4.5 (4)	2.6	100	97.7	2.6	100	97.7	0
ADnet (2,4)	2.6	100	97.7	0	100	100	13.4
ADnet (4,8)	2.6	100	97.7	1.1	100	99.0	5.2
With imp.							
C4.5 (4)	6.2	93.9	93.8	4.1	89.5	95.5	6.5
ADnet (2,4)	4.1	73.2	93.8	0	–	100	19.1
ADnet (4,8)	5.1	84.8	93.8	0.4	100	99.6	15.0

Table 2: Performance comparison for different algorithms for predicting SAD. See Table 1 for notation.. All values are percentages.

Method	SAD binary			SAD conf.			
	FPR	TPR	Acc.	FPR	TPR	Acc.	LC
No imp.							
C4.5 (5)	0.9	56.8	93.5	0.6	61.6	94.8	3.6
ADnet (5,5)	0.9	56.8	93.5	0	72.0	97.4	10.7
ADnet (5,12)	3.0	60.5	92.2	0	80.0	98.3	14.7
ADtree (10,12)	3.2	59.3	91.9	0.4	38.5	94.0	7.2
With imp.							
C4.5 (5)	3.9	57.1	93.0	3.9	57.1	93.0	0
ADnet (5,5)	3.7	55.1	93.0	0.6	41.2	97.5	14.0
ADnet (8,12)	3.5	59.2	93.5	0.8	53.3	97.9	15.1
ADtree (10,12)	0	4.1	92.3	0	–	92.7	1.5

all (or most) of them. Thus we restricted C4.5 to use the questions pre-selected by the network analysis. In all the cases we trained the classifiers using the PAS dataset (917 samples) and then tested on the T-R dataset (614 samples). We define a patient as positive if they were diagnosed with GAD or SAD and negative if they were not. Given it's relevance, we also trained classifiers to predict when the anxiety disorders, both GAD and SAD, occurred with impairment. In the PAPA, imairment is assessed with different sections of the interview which require another set of 40 questions. We did not include any of those questions in the study, we tried to see if the information given in the anxiety sections was enough to predict impairment. In this case positives where patients that were diagnosed with GAD or SAD and also presented any type of social impairment.

We measured performance in terms of False Positive Rate (FPR), True Positive Rate (TPR), and Accuracy (Acc). The accuracy is defined as, $\mathrm{Acc} = (\mathrm{TP} + \mathrm{TN})/(\mathrm{FP} + \mathrm{FN} + \mathrm{TP} + \mathrm{TN})$, where TP and FP are the true and false positives, and TN and FN are the true and false negatives. We denoted as LC the percentage of the tested cases for which the prediction has low confidence.

We begin by running the network analysis presented in Section 3 for both the GAD and SAD diagnosis (Figure 1). In both cases the number of questions with non-zero correlation to the diagnosis is much smaller than the total number of questions. Moreover, in the case of the GAD dataset, there are two questions clearly much more related to the presence of GAD than the others, and using only these two questions the ADtree classifier already yields excellent results. We selected all the questions positively related to the diagnosis: 5 questions out of 27 were chosen for the GAD dataset, and 11 out of 51 for the SAD case. This reduction is very significant, and the resulting questions are the input to the following classifier stage.

Tables 1 and 2 summarize the results obtained for GAD and SAD respectively, while reducing the number of questions by at least an order of magnitude. We can see that in both cases the algorithm provides a very reliable risk estimator with accuracy values in the upper nineties. We observe that the case of GAD is considerably easier and can be asses with much fewer questions than the SAD, obtaining perfect classification results. We further restricted the per-selected questions to only 2, the ones which had the highest correlation coefficients, see Figure 1. In general, we observe that including more questions makes possible to have much smaller low confidence sets, which implies fewer children sent out for further evaluation. Having different models trained with different number of questions means that the test can be carried out in a serial manner, we accumulate evidence in favor or against the presence of the disorder as we go. If at a certain point we have a very high confidence, then we do not need to continue. On the other hand if after making a short test the confidence is still low, we can ask an additional set of questions to try to improve the confidence in the results. We observe very good performance for predicting when the disorders occur with an impairment, which is quite remarkable since non of the specific questions for assessing impairment where included

Using directly the ADtrees algorithm leads to weaker results than pre-selecting the questions as in ADnet. This can be seen clearly in the case of SAD. The ADtrees achieves better performance on the training set but presents worse generalization, requiring more questions to reach a similar performance as ADnet. The confidence score provided by the ADtree algorithm does not carry as much information as in the case of ADnet.

6. DISCUSSION

We presented a method for reliably measuring the risk of psychopathology in childhood. The method requires significantly fewer questions than the standard interviews. We presented experimental validation using two distinct datasets obtaining encouraging results that suggest that it could be a valuable tool for preliminary evaluation in everyday care. This paper is a first step in the direction of using statistical tools in child development applications. We believe that significant contribution can be made with high impact for children's health.

7. REFERENCES

[1] C. Clark, B. Rodgers, T. Caldwell, Ch. Power, and S. Stansfeld, "Childhood and adulthood psychological ill health as predictors of midlife affective and anxiety disorders: the 1958 british birth cohort," *Archives of General Psychiatry*, vol. 64, no. 6, pp. 668, 2007.

[2] L. Franz, A. Angold, W. Copeland, E. J. Costello, N. Towe-Goodman, and H. Egger, "Preschool anxiety disorders in pediatric primary care: Prevalence and comorbidity," *Journal of the American Academy of Child Adolescent Psychiatry*, 2013.

[3] D. A. Chavira, M. B. Stein, K. Bailey, and M. T. Stein, "Child anxiety in primary care: prevalent but untreated," *Depression and Anxiety*, vol. 20, no. 4, pp. 155–164, 2004.

[4] R. M. Rapee, "The development and modification of temperamental risk for anxiety disorders: Prevention of a lifetime of anxiety?," *Biological Psychiatry*, vol. 52, no. 10, pp. 947–957, 2002.

[5] D. R. Hirshfeld-Becker and J. Biederman, "Rationale and principles for early intervention with young children at risk for anxiety disorders," *Clinical Child and Family Psychology Review*, vol. 5, no. 3, pp. 161–172, 2002.

[6] H. L. Egger, A. Erkanli, G. Keeler, E. Potts, B. K. Walter, and A. Angold, "Test-retest reliability of the preschool age psychiatric assessment (papa)," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 45, no. 5, pp. 538–549, 2006.

[7] T. M. Achenbach and T. M. Ruffle, "The child behavior checklist and related forms for assessing behavioral/emotional problems and competencies," *Pediatrics in Review*, vol. 21, no. 8, pp. 265–271, 2000.

[8] Y. Freund and Ll. Mason, "The alternating decision tree learning algorithm," in *ICML*, 1999, vol. 99, pp. 124–133.

[9] M. Rochery, R. Schapire, M Rahim, N Gupta, G. Riccardi, S. Bangalore, H. Alshawi, and S. Douglas, "Combining prior knowledge and boosting for call classification in spoken language dialogue," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*. IEEE, 2002, vol. 1, pp. I–29.

[10] J. R. Quinlan, *C4. 5: programs for machine learning*, vol. 1, Morgan kaufmann, 1993.

[11] M. N. Kumar, "Alternating decision trees for early diagnosis of dengue fever," *arXiv preprint arXiv:1305.7331*, 2013.

[12] K.-Y. Liu, J. Lin, X. Zhou, and S. TC Wong, "Boosting alternating decision trees modeling of disease trait information," *BMC genetics*, vol. 6, no. Suppl 1, pp. S132, 2005.

[13] D. P. Wall, J. Kosmicki, T. F. Deluca, E. Harstad, and V. A. Fusaro, "Use of machine learning to shorten observation-based screening and diagnosis of autism," *Translational psychiatry*, vol. 2, no. 4, pp. e100, 2012.

[14] M. Fiori, P. Musé, A. Hariri, and G. Sapiro, "Multimodal graphical models via group lasso," *Signal Processing with Adaptive Sparse Structured Representations*, 2013.

[15] O. Banerjee, L. El Ghaoui, A. D'Aspremont, and G. Natsoulis, "Convex optimization techniques for fitting sparse Gaussian graphical models," *Proceedings of the 23rd international conference on Machine learning - ICML '06*, pp. 89–96, 2006.

[16] P. Loh and M. Wainwright, "Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses," in *Advances in Neural Information Processing Systems 25*, pp. 2096–2104. 2012.

[17] Y. Freund and R. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational learning theory*. Springer, 1995, pp. 23–37.

[18] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, vol. 1, Springer New York, 2001.