Applied Artificial Intelligence
An International Journal

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uaai20

Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Machine Learning

Tianhua Chen, Grigoris Antoniou, Marios Adamou, Ilias Tachmazidis & Pan Su

To cite this article: Tianhua Chen, Grigoris Antoniou, Marios Adamou, Ilias Tachmazidis & Pan Su (2021) Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Machine Learning, Applied Artificial Intelligence, 35:9, 657-669, DOI: 10.1080/08839514.2021.1933761

To link to this article: https://doi.org/10.1080/08839514.2021.1933761

Published online: 02 Jun 2021.

APPLIED ARTIFICIAL INTELLIGENCE
2021, VOL. 35, NO. 9, 657–669
https://doi.org/10.1080/08839514.2021.1933761

Taylor & Francis
Taylor & Francis Group

Check for updates

Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Machine Learning

Tianhua Chen[a], Grigoris Antoniou[a], Marios Adamou[b], Ilias Tachmazidis[a], and Pan Su[c]

[a]School of Computing and Engineering, University of Huddersfield, Huddersfield, UK; [b]School of Human and Health Sciences, University of Huddersfield, Huddersfield, UK; [c]School of Control and Computer Engineering, North China Electric Power University, Baoding, China

ABSTRACT

Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder that includes symptoms such as inattentiveness, hyperactivity and impulsiveness. It is considered as an important public health issue, and prevalence of diagnosis has increased as awareness of the disease grew over the past years. Supply of specialist medical experts has not kept pace with the increasing demand for assessment, both due to financial pressures on health systems and the difficulty to train new experts, resulting in growing waiting lists. Patients are not being treated quickly enough causing problems in other areas of health systems (e.g. increased GP visits, increased risk of self-harm and accidents) and more broadly (e.g. time off work, relationship problems). Advances in machine learning make it possible to attempt to diagnose ADHD based on the analysis of relevant data, and this could inform clinical practice. This paper reports on findings related to the mental health services of a specialist Trust within the UK's National Health Service (NHS). The analysis studied data of adult patients who underwent diagnosis over the past few years, and developed a diagnostic model for ADHD in adults. The results demonstrate that it is indeed possible to correctly diagnose ADHD patients with promising statistical accuracy.

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common neuropsychiatric conditions with a pooled worldwide prevalence estimated at approximately 5% in school-aged children with persistence of impairing symptoms in adulthood in up to 65% of cases. The pooled estimated prevalence of ADHD in adults is approximately 2.5% (Thapar and Cooper 2016). ADHD is characterized by a persistent and impairing pattern of inattention and/or hyperactivity/impulsivity that causes significant impairment across domains (American Psychiatric Association 2013). Along with these three main symptomatic clusters, people with ADHD also present with deficits in

CONTACT Tianhua Chen T.Chen@hud.ac.uk School of Computing and Engineering, University of Huddersfield, Huddersfield UK, HD1 3DR

executive functions, behavior and emotion regulation, and motivation (Asherson et al. 2016).

For the people who are diagnosed, the modes of interventions for primary ADHD symptoms with robust evidence base are pharmacological and psychological (National Institute for Health and Care Excellence (NICE) 2018). The first line treatment for adult ADHD is psychostimulants (Fields, Johnson, and Hassig 2017). Medication is safe and effective, with 70% of patients reported improvement compared to 7% of controls (Fields, Johnson, and Hassig 2017; Spencer et al. 2001).

The adverse effects of untreated ADHD are well documented with negative effects on academic outcomes (Arnold et al. 2020; Langberg et al. 2011; Loe and Feldman 2007), social functioning (Cook et al. 2014), employment (Adamou et al., 2013) but also life itself leading to increased mortality (Dalsgaard et al. 2015). The total yearly costs to the individual and state combined were recently estimated to be €17,769 per person, per year (Simone 2018) so there is strong impetus for interventions.

For the UK, the National Institute for Health and Clinical Excellence (NICE) suggested in 2008 that the standard benchmark rate for referral to a Service in adults is 25 per 100,000 per year (CG72 Attention deficit hyperactivity disorder (ADHD): Audit support (adults) 2008). The largest challenge at the moment for the adult population, bearing in mind the relative recency of acceptance amongst the professional community that ADHD can persist into adulthood (Asherson et al. 2010), is the dearth of clinicians appropriately trained and confident to place the diagnosis. Such bottleneck prevents patients receiving appropriate treatments and hence contributes to the morbidity of the adult ADHD.

Recent advance in machine learning has enjoyed a number of successes in medical applications (Chen et al. 2021a; Chen et al. 2021b). To address this challenge, we wanted to investigate if there was a way by which using clinical information collected from a Service which delivers clinical pathway which is compliant with NICE recommendations (i.e. gold standard), can create a decision tool which can automate the process of making a diagnosis. The clinical data collected in this paper is from a NHS specialist mental health provider in the form of screening questionnaires and clinical interviews, which are routinely collected when a new patient is referred. We are not aware of this being achieved anywhere else in mental health populations whereby an AI algorithm will make diagnostic decisions based on the form of data we use.

The experimental evaluation demonstrates that by applying machine learning, we can achieve a diagnostic accuracy of 85%. This is a very promising result. We are currently working on the computation of confidence scores, with the view that confident recommendations can be adopted, while more borderline cases will be treated as currently by clinicians, all within an NHS clinical pathway. This way, clinician productivity will be increased, while avoiding any incorrect diagnoses. We are in the process of validating the clinical and financial viability within the Trust through a pilot deployment.

The remainder of this paper is organized as follows. Section 2 describes the available data. Section 3 provides details of our data analysis and its outcomes, while section 4 discusses current and future work.

Data Collection

A National Health Service specialist mental health provider (South West Yorkshire Partnership NHS Foundation Trust-SWYPFT) made available for analysis all anonymized data for assessments made of ADHD patients in the period between 2014 and 2017. Overall, there were 69 such patients. For all these patients, the data contained information which included demographics and a number of validated self-reported screening questionnaires and clinical interviews.

Each patient contains a client ID, which is used to join with other entries related to the patient – see below – and demographic information about age, gender and post code. All this information is in (semi-)structured form. Table 1 summarizes the descriptive statistics of the demographics information. No difference in age was assessed between the two gender through t-test (p-value = 0.06).

The screening questionnaires included the Conner's ADHD Rating Scales (Conners, Erhardt, and Sparrow 1999), the Drug Abuse Screening Test (DAST-10) (Skinner 1982), the Iowa Personality Disorder Screen (IOWA) (Langbehn et al. 1999), the Alcohol Use Disorders Identification Test (AUDIT) (Saunders et al. 1993), the Mood Disorder Questionnaire (MDQ) (Hirschfeld 2002), the GAD-7 measuring Generalized Anxiety (Spitzer et al. 2006), and the Patient Health Questionnaire (PHQ-9), which measures the severity of depression. The clinical interviews were both structured and unstructured. The structured interviews were made using the Diagnostic Interview for ADHD in adults (DIVA) (Ramos-Quiroga et al. 2019) and unstructured where captured in the text of the final medical report which

Table 1. Demographics information.

Age	Average (std)
Whole population	33.01 (9.931)
Men	31.36 (10.85)
Women	36.13 (7.12)
Gender	*Number of subjects (%)*
Male	45 (65.2%)
Female	24 (34.8%)
Post Code	
Wakefield	42 (60.9%)
Barnsley	27 (39.1%)

was also provided. In addition, we had data from the scores of the Sainsburys Risk Assessment Tool (Morgan 2000) and results from the objective measurement of ADHD symptoms obtained using the QBTest (Reh et al. 2015).

Experimental Study

This section reports the experiments conducted using machine learning algorithms on predicting ADHD diagnosis.

Predictive Analysis Set-up

As explained in Section 2, the original data comprises of a succession of assessments, each of which describes the potential patient from one particular perspective, as well as medical note that records personal and family history. Having discussed with clinical experts, an assessment-centered analysis is devised, with the goal of predicting whether a referred patient is with ADHD given the available information. The objective is to implement a predictive model that assesses referrals according to the patient's risk for being with ADHD and that can be made operational in a clinical environment.

In order to generate an assessment-centered data set for constructing a predictive model, patient demographics, self-reported assessment, Conner's Adult ADHD Rating Scale (short version) on the basis of both – self-report and observe mode, the Qb Test, and the diagnostic interview for ADHD in Adults are jointed to form an overall assessment report. Once the main assessment data is formed, three groups of predictive analysis is designed as follows, depending on how the risk assessment data and medical note is used.

Construct the predictive model by purely using the main assessment data, which consisted of 28 variables. Note that missing values occasionally occur for some of variables. One conventional way to handle this scenario is to fill with the average value for a continuous variable and mode value for a discrete variable where missing values apply.

Build the predictive model on the basis of joining the main assessment and the risk assessment data, which results in an additional 66 variables and 94 variables in total. Note that each variable from the risk assessment data is binary that only take 'yes' or 'no' for a given assessment question.

The medical note takes down a comprehensive personal record, which is utilized to recognize significant life events such as aggression, prison, bully problem, which may contribute to the ADHD diagnosis.

Predictive Analysis Pipeline

In order to test how accurate a particular predictive model performs, when it is asked to make new predictions for data it has not seen yet in practice, k-fold cross validation is used whereby a model is given a dataset of known data on which training is run and an independent dataset of unknown data against which the model is tested. In order to implement this, a given data set is divided into k subsets. Each time, one of the k subsets is used as the test set and the other k-1 subsets are put together to form a training set. Then the average error across all k trials is computed. Considering the number of available dataset is relatively small, which only comprises 69 patients in total, the leave-one-out cross validation is utilized, which is k-fold cross validation taken to its logical extreme, with k equal to the number of data points in the set. That is to say, the learning method is trained on all the data except for one patient and a prediction is made for this particular patient. The average error is computed on the basis where each single patient has used to test the performance of a given model and used to evaluate the learning method.

The next issue comes to which particular machine learning method to use for the acquisition of a final model to use in practice. Given that the cost of model misclassification is potentially high, explanations with respect to how a machine made conclusion is derived play a significant role informing clinicians making unbiased decisions in combination with medical domain knowledge. From interpretability viewpoint, a transparent model is preferred so that the generated system is able to reason about how it reaches a conclusion and provide explanation of its reasoning to end users. Performance-wise, the model should be accurate enough to make correct decisions. As a result, a number of popular machine learning techniques were tested:

Support Vector Machine (Shawe-Taylor and Sun 2011) is a sequential optimization algorithm, which aims to construct a multidimensional hyper-plane that optimally discriminates between the two classes by maximizing the margin of the two data clusters.

Logistic Regression (Hosmer, Lemeshow, and Sturdivant 2013) is a statistical model for building linear logistic regression models, which models the probability of output in terms of input. When used as a classifier, a cutoff value is chosen and classifiers inputs with probability greater than the cutoff as one class, below as the other class.

Naive Bayes (Robert 2014) is a simple probabilistic learning classifier that assigns class labels to problem instances, represented as vectors of feature values, based on direct application of the Bayesian theorem with strong independence assumptions.

Random Forest (Liaw and Wiener 2002) is a powerful ensemble learning method by constructing a multiple of decision tree classifiers at training time and outputting the class that is most representative among all ensemble members.

Decision Tree (Bhargava et al. 2013) is an algorithm to generate a tree that begins with the original training set at the root node. On each iteration of the algorithm, it iterates though every unused attribute and selects one with the largest information gain to produce subsets of the data. It continues to recuse on each subset until all attributes have been used or no more additional gains obtained.

K-nearest Neighbor (Triguero et al. 2012) is the classical machine learning approach, where an instance is classified by a majority vote of its neighbors. It works by assigning an instance to the class most common among its k ($k = 3$ in this experiment) nearest neighbors.

In a nutshell, Figure 1 illustrates the general experimental pipeline of the underlying study. With the full collection of clinical data, which comprise the main assessment data, risk assessment data and the medical notes, three different configures are setup to properly utilize all available data. For each configuration, the leave-one-out cross validation is adopted to train and test a given machine learning model, i.e., to train a model using information from

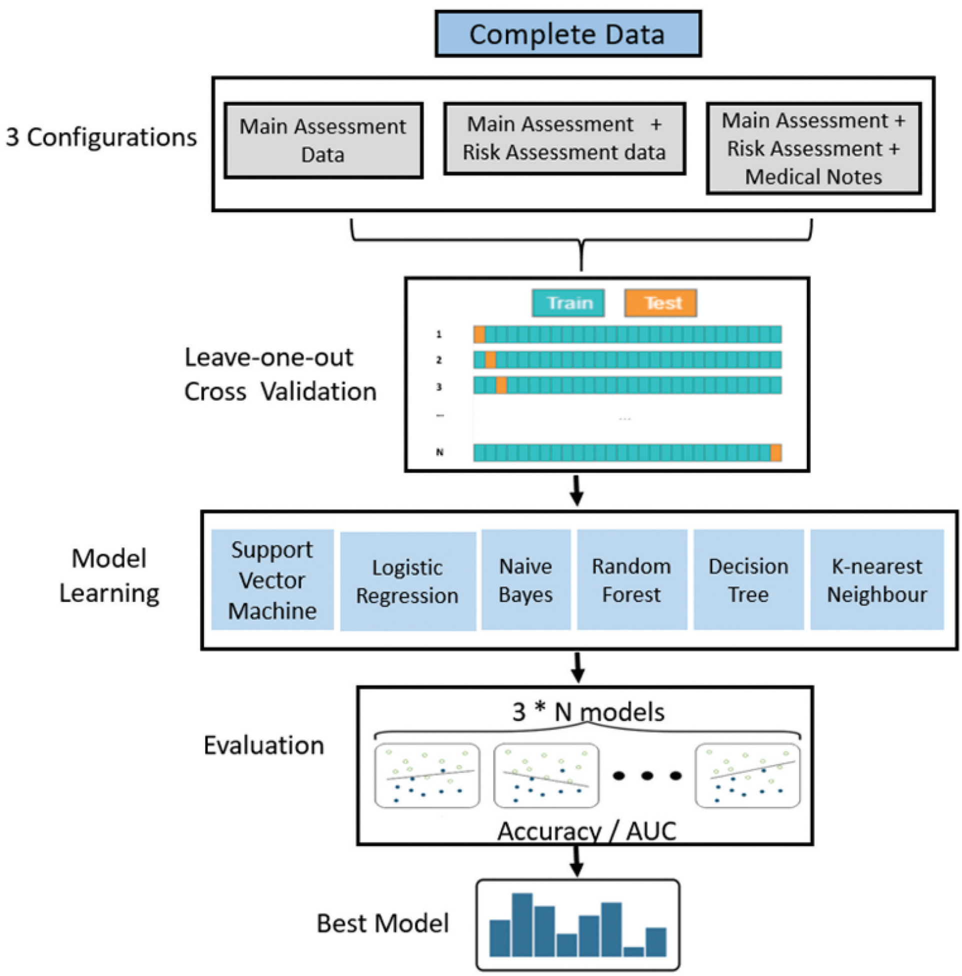

Figure 1. Experimental pipeline.

any N-1 patients out of N total patients, and test the established model using the remaining patient. The performance of a model will be evaluated with accuracy and AUC score. A total number of six popular machine learning algorithms will be employed in an effort to select the best model for clinical use.

Experimental Analysis

To demonstrate the performance of machine learning algorithms for the predictive modeling of ADHD diagnosis, experiments were conducted using the scikit-learn open source machine learning library for the Python programming language, which integrates the implementation of all aforementioned ML approaches with default settings unless otherwise explicitly specified. Performances are reported as accuracy, which is the percentage of correct predictions, i.e., the resultant model predicts positive in case the patient to be diagnosed is with ADHD and negative in case the patient is without ADHD. A perfect classification model would always make correct predictions, resulting in 100% accuracy. In addition, Area Under the Receiver Operating Characteristic (ROC) curve (AUROC or just AUC) is also reported, which illustrates the performance for a binary classification problem, when a threshold is varied on the predictions. AUC is the curve of sensitivity (a.k.a. true positive rate), plotted against 1-specificity (a.k.a. false positive rate), which is independent of the prior class distribution, i.e., percentages of positive and negative samples. A perfect classification would produce AUC = 1, while random guessing would produce a 0.5 AUC.

Table 2 summaries the performance on the main assessment report, which consisted of 27 variables. Most algorithms achieves accuracy in the range of 70–80%, with the decision tree algorithm having accomplished the highest accuracy as highlighted in bold, followed by random forest and Naive Bayes. In terms of AUC, the three algorithms that achieve the top 3 best accuracies are also competent with each other, resulting in very close AUC. It is worth noting that the experiment at this stage aims to identify the optimum machine learning algorithm for this ADHD predictive modeling task, hence necessary to compare their performances.

Table 2. Experimental results on the main assessment data.

Machine Learning Method	#patients	#variables	Accuracy (%)	AUC
Support Vector Machine	69	27	72.464	0.784
Logistic Regression	69	27	72.464	0.795
Decision Tree	69	27	82.609	0.866
K Nearest Neighbor	69	27	59.420	0.558
Random Forest	69	27	81.159	0.866
Naive Bayes	69	27	75.362	0.870
Averaged	69	27	73.91 ± 8.30	0.79 ± 0.12

In addition to the main assessment data, the experiment carries on utilizing the risk assessment data as well, which evaluates a potential patient historical behavior that might be related to the occurrence of ADHD. On the basis of the additional 66 variables resulted from the risk assessment, the aforementioned machine learning algorithms construct predictive models joining the main assessment and the risk assessment data, resulting in 94 variables in total.

According to Table 3, in spite of a significant larger number of variables, performances of the resulting algorithms generally match that in Table 2. That is, most algorithms still perform in the level of 70–80%. However, performances of most algorithms have generally improved (except k-nearest neighbor and Random Forest). This is also expected from a clinical viewpoint, considering a lot of relevant and targeted information is now embedded in the training by utilizing the risk assessment which is specifically carried out as a clinical activity. Overall, the decision tree is a clear winner being the only algorithm with accuracy above 80%, as well as top AUC value.

The experiment is carried out further, with a view to utilizing the information embedded in the text medical notes, which record the details pertaining to the development of ADHD symptoms over the course of growing up. Once the medical scripts are collected, a number of pre-processing steps are necessary for the generation of clean documents for further processing. These include the following:

Tokenize the reviews such that each review is represented as a collection of words for text analysis;

Convert all text data to lowercase, so that the words of different cases could be treated the same to remove redundancy;

Erase punctuation and symbols, which can safely be ignored without sacrificing the meaning of the sentence;

Remove a list of stop words such as ‘and’ and ‘the’ that does not add much meaning to a sentence;

Lemmatize the words to reduce words to their dictionary forms such that for example, ‘am’, ‘are’ and ‘is’ can all be converted to ‘be’.

The extraction of a collection of raw medical scripts can be summarized as shown in Figure 2 by a word cloud that creates a visual representation of the text data, where the prominence of individual terms is reflected by size and font. Following the bag-of-words approach where each medical note is

Table 3. Experimental results on the main assessment and risk assessment data.

Machine Learning Method	#patients	#variables	Accuracy (%)	AUC
Support Vector Machine	69	93	76.812	0.806
Logistic Regression	69	93	75.362	0.815
Decision Tree	69	93	85.507	0.871
K Nearest Neighbor	69	93	59.420	0.559
Random Forest	69	93	75.362	0.804
Naive Bayes	69	93	72.464	0.740
Averaged	69	93	74.15 ± 8.47	0.77 ± 0.11

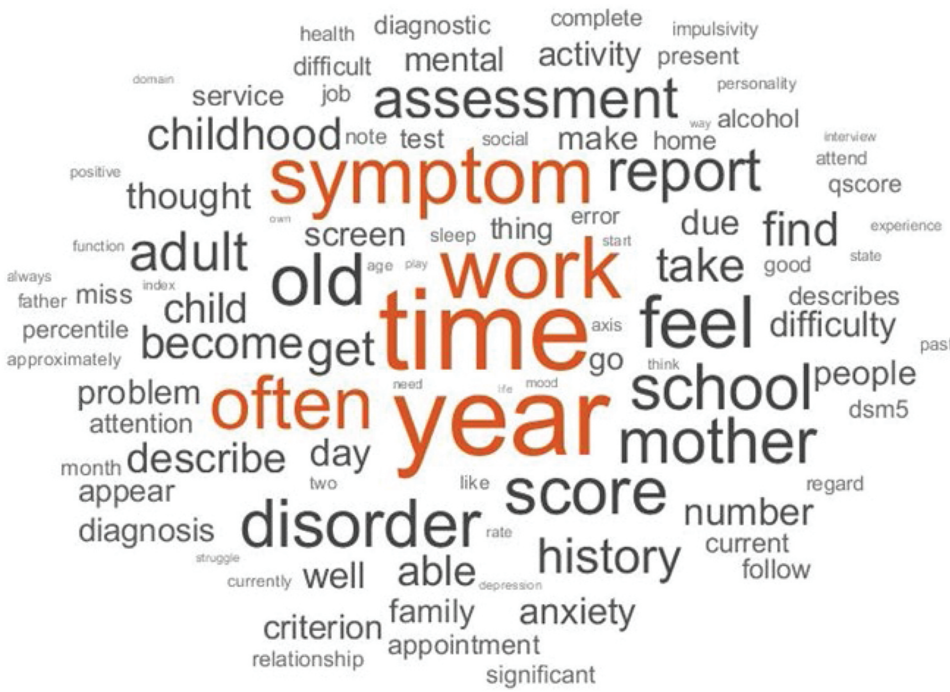

Figure 2. Word cloud of medical notes.

represented as a matrix of the length that is equal to the number of unique terms in the returned corpus, a total number of 7609 terms are extracted. However, most of extracted text terms may not be informative to the diagnosis of ADHD, which may even overfit the classification algorithms considering the small number of patients. Having recognized the significance of life events such as aggression, prison, bully problem, which may contribute to the ADHD diagnosis, a total number of 13 pre-defined keywords are provided by clinical experts, which do not overlap with those covered by the risk assessment.

Table 4 shows the performances of the resulting classifiers on the basis of simultaneously using the main assessment, risk assessment as well as the variables extracted from medical scripts. Although decision tree still remains the one with top accuracy (jointly with Naive Bayes), its performance has decreased to 79.710% when utilizing the additional variables extracted from

Table 4. Experimental results using all available data.

Machine Learning Method	#patients	#variables	Accuracy (%)	AUC
Support Vector Machine	69	106	76.812	0.819
Logistic Regression	69	106	75.362	0.810
Decision Tree	69	106	79.710	0.865
K Nearest Neighbor	69	106	59.420	0.559
Random Forest	69	106	71.014	0.777
Naive Bayes	69	106	79.710	0.739
Averaged	69	106	73.67 ± 7.69	0.76 ± 0.11

medical notes, in comparison to 85.507% when text scripts are not used. For the other classifiers, the additional information from medical notes does not significantly impact their performances, resulting in only slight decreased averaged accuracy and AUC overall. These indicate that the use of medical note may not necessarily help improve the predictive capability, and could even damage the model by overfitting a given classifier, which may of course be remedied if more patients are included in the data analysis.

As a result, the findings of the above experiments may be summarized as follows:

Decision tree has been the best overall classifier in comparison to five popular alternatives, achieving 3 highest accuracies and 2 top AUC values.

Decision tree algorithm generates a set of IF-THEN rules, each of which provides a diagnosis specified by the condition. The rule base is interpretable, offering a means to explain how a conclusion is derived, which is necessary for a data-driven model to be employed in practice. In case of rules against medical knowledge, clinicians can easily make changes or simply delete abnormal rules.

Risk assessment data has helped improve performance of the decision tree from 82.609% to 85.507% if put into use in conjunction with the main assessment data. As such risk assessment data should be utilized to generate the final model.

Although medical notes comprise a comprehensive set of personal information, the inclusion of a set of key words extracted does not necessarily facilitate performance increment. Instead it actually overfits the existing model, resulting in a worse performance, which is likely attributed to two reasons: (1) the information extracted from medical notes are redundant or even conflicting with those from main assessment and risk assessment data; (2) more patients are necessary to fully make use of medical notes to output a robust model. Nevertheless, medical notes may be utilized in the future when more patients are available, but are excluded into use to avoid overfitting at the moment.

In a nutshell, from the perspective of both performance and interpretability, decision tree is selected as the machine learning algorithm to output final predictive model; while the main assessment and risk assessment data are used to train the model.

Conclusion and Future Work

In order to automate the diagnosis of ADHD, a number of popular machine learning techniques have been used to analyze data related to patients of an NHS who have been suspects of potential patients. The data sources included structured patient information as well as unstructured textual medical notes. Six popular machine learning algorithms have been applied to this cohort of patients, with decision tree learning algorithm identified as the optimum choice to construct the final predictive model, owing to its superior performance and interpretability. Another outcome of the underlying study shows that the joint use of main assessment and the risk assessment data would generate the best performance for decision tree.

Whilst being promising, it would be beneficial to collect more patients' data in order to build a more robust model. It would also be interesting to investigate the use of alternative approaches to constructing the predictive model, e.g., using the recently proposed fuzzy rule-based models like the ones proposed in (Chen et al. 2016; Chen et al. 2018), which may work better while dealing with the uncertainty and linguistic imprecision embedded in the cognitive tests.

Other lines of current and future work are to:

Develop a user-friendly tool incorporating the predictive model.

Validate the predictive model in a trial deployment within the NHS Trust.

Compute a confidence score and associated confidence thresholds; confident recommendations can be adopted, while more borderline cases will be treated as currently by clinicians, all within an NHS clinical pathway.

Deal with missing values more effectively by exploiting advanced knowledge interpolation techniques (Chen et al. 2019).

Perform feature selection (Su et al. 2017; Chen et al. 2019) to remove attributes that may be irrelevant, redundant or even misleading.

References

Adamou, M., M. Arif, P. Asherson, T.-C. Aw, B. Bolea, D. Coghill, G. Guðjónsson, A. Halmøy, P. Hodgkins, U. Müller et al. 2013. Occupational issues of adults with ADHD. *BMC Psychiatry*. 13(1):59. doi:10.1186/1471-244X-13-59.

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders*. 5 ed. Arlington, VA: American Psychiatric Association.

Arnold, L. Eugene, et al. "Long-term outcomes of ADHD: academic achievement and performance.„ Journal of attention disorders 24.1 (2020): 73–85

Asherson, P., J. Buitelaar, S. V. Faraone, L. A. Rohde. 2016. Adult attention-deficit hyperactivity disorder: Key conceptual issues. *The Lancet Psychiatry*. 3(6):568–78. doi:10.1016/S2215-0366(16)30032-3.

Asherson, P., M. Adamou, B. Bolea, U. Muller, S. D. Morua, M. Pitts, J. Thome, S. Young. 2010. Is ADHD a valid diagnosis in adults? Yes. *BMJ*. 340(mar26 1):c549. doi:10.1136/bmj.c549.

Bhargava, Neeraj, Girja Sharma, Ritu Bhargava, and Manish Mathuria. "Decision tree analysis on j48 algorithm for data mining.„ Proceedings of international journal of advanced research in computer science and software engineering 3, no. 6 (2013).

CG72 Attention deficit hyperactivity disorder (ADHD): Audit support (adults). 2008, National Institute for Health and Clinical Excellence.

Chen, T., C. Shang, P. Su, E. Keravnou-Papailiou, Y. Zhao, G. Antoniou, Q. Shen. 2021a. Medical analytics for healthcare intelligence – Recent advances and future directions. *Artificial Intelligence in Medicine* 111,0933–3657.

Chen, T., C. Shang, P. Su, E. Keravnou-Papailiou, Y. Zhao, G. Antoniou, Q. Shen. 2021b. A decision tree-initialised neuro-fuzzy approach for clinical decision support. *Artificial Intelligence in Medicine* 111:101986.

Chen, T., C. Shang, P. Su, Q. Shen. 2018. Induction of accurate and interpretable fuzzy rules from preliminary crisp representation. *Knowledge-Based Systems* 146:152–66. doi:10.1016/j.knosys.2018.02.003.

Chen, T., Q. Shen, P. Su, C. Shang. 2016. Fuzzy rule weight modification with particle swarm optimisation. *Soft Computing*. 20(8):2923–37. doi:10.1007/s00500-015-1922-z.

Chen, T., Shang, C., Yang, J., Li, F., & Shen, Q. (2019). A new approach for transformation-based fuzzy rule interpolation. IEEE Transactions on Fuzzy Systems, 28(12), 3330–3344.

Chen, Tianhua, Pan Su, Changjing Shang, Richard Hill, Hengshan Zhang, and Qiang Shen. "Sentiment Classification of Drug Reviews Using Fuzzy-rough Feature Selection.„ In 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6. IEEE, 2019

Conners, C. K., D. Erhardt, and E. Sparrow. 1999. *Conners' adult ADHD rating scales*. New York: MHS.

Cook, J., E. Knight, I. Hume, A. Qureshi. 2014. The self-esteem of adults diagnosed with attention-deficit/hyperactivity disorder (ADHD): A systematic review of the literature. *ADHD Attention Deficit and Hyperactivity Disorders*. 6(4):249–68. doi:10.1007/s12402-014-0133-2.

Dalsgaard, S., S. D. Østergaard, J. F. Leckman, P. B. Mortensen, M. G. Pedersen. 2015. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: A nationwide cohort study. *The Lancet*. 385(9983):2190–96. doi:10.1016/S0140-6736(14)61684-6.

Fields, S. A., W. M. Johnson, and M. B. Hassig. 2017. Adult ADHD: Addressing a unique set of challenges. *The Journal of Family Practice* 66 (2):68–74.

Hirschfeld, R. M. 2002. The mood disorder questionnaire: a simple, patient-rated screening instrument for bipolar disorder. *The Primary Care Companion to the Journal of Clinical Psychiatry* 4 (1):9–11. doi:10.4088/PCC.v04n0104.

Hosmer, D. W., Jr, S. Lemeshow, and R. X. Sturdivant. 2013. *Applied logistic regression*, vol. 398. John Wiley & Sons.

Langbehn, D. R., B. M. Pfohl, S. Reynolds, L. A. Clark, M. Battaglia, L. Bellodi, R. Cadoret, W. Grove, P. Pilkonis, P. Links, et al. 1999. The iowa personality disorder screen: Development and preliminary validation of a brief screening interview. *Journal of Personality Disorders*. 13(1):75–89. doi:10.1521/pedi.1999.13.1.75.

Langberg, J. M., B. S. G. Molina, L. E. Arnold, J. N. Epstein, M. Altaye, S. P. Hinshaw, J. M. Swanson, T. Wigal, L. Hechtman. 2011. Patterns and predictors of adolescent academic achievement and performance in a sample of children with attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*. 40(4):519–31. doi:10.1080/15374416.2011.581620.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. *R News* 2 (3):18–22.

Loe, I. M., and H. M. Feldman. 2007. Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology* 32 (6):643–54. doi:10.1093/jpepsy/jsl054.

Morgan, S. 2000. *Clinical risk management: A clinical tool and practitioner manual*. London: The Sainsbury Centre for Mental Health.

National Institute for Health and Care Excellence (NICE). 2018. Attention deficit hyperactivity disorder: Diagnosis and management. NICE guideline [NG87].

Ramos-Quiroga, J. A., V. Nasillo, V. Richarte, M. Corrales, F. Palma, P. Ibáñez, M. Michelsen, G. Van De Glind, M. Casas, J. J. S. Kooij, et al. 2019. Criteria and concurrent validity of

DIVA 2.0: A semi-structured diagnostic interview for adult ADHD. *Journal of Attention Disorders*. 23(10):1126–35. doi:10.1177/1087054716646451.

Reh, V., M. Schmidt, L. Lam, B. G. Schimmelmann, J. Hebebrand, W. Rief, H. Christiansen. 2015. Behavioral assessment of core ADHD symptoms using the QbTest. *Journal of Attention Disorders*. 19(12):1034–45. doi:10.1177/1087054712472981.

Robert, C. 2014. *Machine learning, a probabilistic perspective*. Taylor & Francis.

Saunders, J. B., O. G. AASLAND, T. F. BABOR, J. R. DE LA FUENTE, M. Grant. 1993. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction*. 88(6):791–804. doi:10.1111/j.1360-0443.1993.tb02093.x.

Shawe-Taylor, J., and S. Sun. 2011. A review of optimization methodologies in support vector machines. *Neurocomputing* 74 (17):3609–18. doi:10.1016/j.neucom.2011.06.026.

Simone, V. 2018. *Your attention please: The social and economical impact of ADHD*. London: DEMOS.

Skinner, H. A. 1982. The drug abuse screening test. *Addictive Behaviors* 7 (4):363–71. doi:10.1016/0306-4603(82)90005-3.

Spencer, T., J. Biederman, T. Wilens, S. Faraone, J. Prince, K. Gerard, R. Doyle, A. Parekh, J. Kagan, S. K. Bearman, et al. 2001. Efficacy of a mixed amphetamine salts compound in adults with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*. 58 (8):775–82. doi:10.1001/archpsyc.58.8.775.

Spitzer, R. L., K. Kroenke, J. B. W. Williams, B. Löwe. 2006. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*. 166(10):1092–97. doi:10.1001/archinte.166.10.1092.

Su, P., Shang, C., Zhao, Y., Chen, T. and Shen, Q., 2017, July. Fuzzy rough feature selection based on owa aggregation of fuzzy relations. In 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1–6). IEEE.

Thapar, A., and M. Cooper. 2016. Attention deficit hyperactivity disorder. *Lancet* 387 (10024):1240–50. doi:10.1016/S0140-6736(15)00238-X.

Triguero, I., J. Derrac, S. Garcia, F. Herrera. 2012. A taxonomy and experimental study on prototype generation for nearest neighbor classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 42(1):86–100. doi:10.1109/TSMCC.2010.2103939.