Fuzzy Data Mining for Autism Classification of Children

Mofleh Al-diabat

Department of Computer Science,
Al Albayt University,
Jordan

Abstract—**Autism is a development condition linked with healthcare costs, therefore, early screening of autism symptoms can cut down on these costs. The autism screening process involves presenting a series of questions for parents, caregivers, and family members to answer on behalf of the child to determine the potential of autistic traits. Often existing autism screening tools, such as the Autism Quotient (AQ), involve many questions, in addition to careful design of the questions, which makes the autism screening process lengthy. One potential solution to improve the efficiency and accuracy of screening is the adaptation of fuzzy rule in data mining. Fuzzy rules can be extracted automatically from past controls and cases to form a screening classification system. This system can then be utilized to forecast whether individuals have any autistic traits instead of relying on the conventional domain expert rules. This paper evaluates fuzzy rule-based data mining for forecasting autistic symptoms of children to address the aforementioned problem. Empirical results demonstrate high performance of the fuzzy data mining model in regard to predictive accuracy and sensitivity rates and surprisingly lower than expected specificity rates when compared with other rule-based data mining models.**

Keywords—*Autistic traits; data mining; fuzzy rules; statistical analysis*

I. INTRODUCTION

Autism is a type of developmental condition initially listed under the umbrella of Diagnostic and Statistical Manual 4th edition text revised version (DMS-IV-TR) [1] as a type of Pervasive Developmental Disorder [PDD] [2]. Autism spectrum disorder [ASD] is defined as 'the challenges in social, communication, interaction, stereotyped movements, sensory and imagination skills, which significantly affect the behavioural performance of an individual'. According to the 2014 figures from the Disease Control and Prevention Centre [CDC], one child out of every 68 children is chronicled as a case of autism (1 per cent of the entire world population) [3]. By 2014, 3.5 million people in the USA had been diagnosed as cases of autism; the number of cases identified in the United Kingdom has risen 119.4 per cent from 2008 to 2014.

ASD screening is the process by which the autistic symptoms of an individual can be determined [4]. This is a crucial phase of ASD diagnosis as autism can't be identified by conventional clinical methods such as blood tests or body check-ups. There are various types of autism screening tools that involve direct observation, structured and semi-structured questionnaires and interviews [5]. Due to a lack of reliable measures in screening children for autism, in many situations the symptoms become visible only after they become adults. Therefore, the role of a viable screening instrument for identifying the risk of ASD at the preliminary stage is huge.

Existing ASD screening techniques rely on a simple domain expert, as well as a large number of questions that respondents have to answer, so these techniques have been criticized by scholars for being lengthy and subjective [5]-[9]. Therefore, developing detection systems that can be extracted using automated methods could be a promising direction. This approach of learning is called data mining and typically utilizes an historical dataset to discover effective hidden patterns for improving planning and the decision process [10], [11]. Recent initial studies in autism research, particularly ASD diagnosis, for example, [12]-[17] and others, indicated that data mining and machine learning techniques could enhance accuracy and efficiency of the diagnostic phase. However, there has been little headway in investigating data mining techniques within autism screening due to the unavailability of datasets. With the advancement of mobile technology, a recent dataset related to behavioural characteristics of autism has been proposed by [18].

This paper investigates fuzzy data mining models to detect autistic symptoms for cases and controls of children between the ages of 4-11 years. The proposed model learns If-Then rules based on different independent variables related to behaviour, i.e. AQ-10-Child [4], and other demographic features such as age, gender, and ethnicity. The dataset used in this research project consists of over 24 variables that have already been screened using a mobile application called ASDTests which was developed in 2017 [19]. A fuzzy rule based on data mining has been learnt using a Fuzzy Unordered Rule Induction algorithm (FURIA) [20]. The rules derived have been adopted to successfully distinguish individuals with ASD. In addition, these rules can be utilized to replace existing domain expert rules and possibly assist clinicians in referring individuals with ASD symptoms for further evaluation; additionally parents can now understand the relationship between autistic traits.

This paper is structured such that Section 2 discusses recent research developments related to the use of data mining in autism research, Section 3 presents data, features, the experimental setting, and results analysis. Finally, a conclusion is given in Section 4.

II. Literature Review

Investigated claims proposed by two other research studies regarding shortening the autism diagnosis related to the utilization of machine learning techniques in discriminating autism in the clinical context [21], i.e. [9], [22]. The researchers used 1949 instances [9], [22], that were obtained from the Autism Genetic Resource Exchange [AGRE] and Balance Independent [BID] datasets [23], [24]. Prior to experimentation, the dataset had been modified as [9], [22] eliminated instances that were not clear ASD cases. Then, the same machine learning techniques (tree-based algorithms) were used to classify individuals. The results of the [21] study revealed severe methodological and conceptual problems and, more importantly, no significant time reduction was found as claimed by the previous studies.

The authors in [17] explored the use of Twitter messages to feed into a data mining tool in order to obtain useful knowledge related to challenges, concerns and practices of autism, therefore raising awareness among people in the community. The ASD-related tweets and messages were collected by typing various keywords using the Twitter search engine to obtain the necessary data. The data was then analyzed in terms of Zipf's law criteria including message length, content, word frequency, hash tag frequencies, and parts of speech frequencies [6]. A further analysis was conducted to test whether the ASD tweets and non-ASD tweets could be automatically classified. The findings of the study concluded a number of common and differential characteristics related to ASD and non-ASD categories could be used to develop an automated mechanism to monitor the behaviours of the ASD community on social media.

The authors in [25] studied how data mining techniques can be used to enhance the impact of behavioural therapy on autistic individuals. Data was collected through videotaped sessions of approximately nine hours each from eight different autistic children who were receiving treatment therapy. During each session, the therapists recorded the four appropriate and inappropriate child behavioural types: their own playroom behaviour, behaviour with parents, behaviour with therapists, and behaviour with strangers. The findings of the research, based on data obtained through data mining techniques, indicated that behavioural therapy can increase appropriate behaviours and reduce any inappropriate behaviour of the autistic children. The rules discovered confirmed that the likelihood and frequency of appropriate and inappropriate behaviours can be predicted more accurately with more data.

The authors in [3] investigated nuroimaging patterns of autistic individuals to establish an effective mechanism to discriminate autism without the involvement of a long adminstration process that requires exclusive training and expertise. Functional Magnetic Resonance Imaging (fMRI) [26], [27] is used to capture the brain images of the subject when he is resting or idle. A total of 1035 fMRI instances were obtained from Autism Brain Imaging Data Exchange (ABIDE) [28] and then analysed to discover the pattern that could help to diagnosis autism. Deep learning techniques are used to classify and understand the unique features of neuro images of autistic individuals' brains and the functionality that can be used to diferentiate cases of autism from controls. The findings of the research suggested that autism can be differentiated 69 percent more accurately through neuroimaging patterns of the brain, than the conventional diagnosis methods, by using deep learning techniques like denoising autoencorders.

The authors in [29] examined the temporal variability of the functional connections (FC) using machine learning techniques and brain neuroimaging techniques for ASD classification. The node variability of the subject's brain is obtained to train different machine learning models on a large resting mind fMRI [26] data of ASD and non-ASD individuals obtained from ABIDE [28]. Machine learning classifiers such as Naive Bayes [30], Random Forest [31], Support Vector Machines [32] and the Multilayer Perceptron algorithm [33] were applied on 147 cases and 146 controls of autism obtained from ABIDE using Weka, open source marchine learning tool kit [34]. According to the results of the study, the machine learning models trained on different functional variabilitiy connections of the brain can achieve an accuracy of 62 percent in classifying and distingusing autism with a sensitivity of 60-65 per cent and specificity of 60+ percent.

The authors in [26] investigated whether machine learning can be an effective mechanism to diagnosis autism and Attention Deficit Hyperactive Disorder (ADHD). To ahieve the objective, the authors tested six different machine learning techniques on 2925 Social Responsive Scale [SRS] data obtained from Simons Simplex Collection version 15, Boston Autism Consortium and Autism Genetic Resource Exchange [35], [23]. The data relevant to 65 SRS items was stratified into 10 folders each comprising 10 percent of both ASD and ADHD data to perform cross validation. For each cross validation session, a minimal redundancy-maximal relevance [mRMR] feature selection method [36] was performed to rank all 65 items. The six machine learning algorithms including Support Vector Machines [32] linear discriminant analysis [21], Categorical lasso [16], tree-based algorithms (Decision Tree and Random Forest) [31], [35] and Logistics Regression Model [37] were tested on all the 65 rankings using the package Scikit-learn [38]. The results of the experiments showed that the majority of the machine learning techniques improve the accuracy of autism diagnosis. Particularly, a combination of Support Vector Machines, Logistics Regression, linear discriminant analysis and Categorical Lasso techniques produced the optimum level of performance in classifying autism and ADHD test instances.

The authors in [37] suggested a machine learning-based system to forecast ASD symptomology through the eye movement patterns of individuals. Initial experiments were carried out on two target groups of Chinese children. A total of 20 ASD children, 21 age-matched typically developing (TD) children, 20 IQ matched TD children (1st group), and 19 ASD, 22 IQ matched Intellectually Disabled (ID), and 28 age matched TD young adults and adolescents (2nd group). The eye movements and gazing patterns were captured through a Tobii T60 eye tracker. The images captured were analyzed using k-means [39] to identify the eye gaze coordinates on the spatial domains and to divide the face into different regions. ASD cases are anticipated to be distinguished based on the

magnitude and directions of both the eye gaze coordinates and eye motions. A model similar to "bag of word" (BoW) is used to document the sequence of coordinates per image per person. The prediction models are developed using the Support Vector Machine (SVM) algorithm to avoid negative data and to identify linear decision boundaries. The subject level predictions with a global threshold are enabled as a scoring context to interpret functional boundaries and decision boundaries. The results of the experiments presented a greater potential and effectiveness in the proposed system for identifying symptoms of ASD.

The author in [38] evaluated the machine learning techniques used in prevailing ASD screening and diagnosis tools to identify their pitfalls to provide recommendations and guidance for future developments. Most of the previous research works on the similar topic of interest have addressed the quality, accuracy, technology usage and many other areas related to the computerised ASD diagnosis, but no study has yet addressed the different conceptual, implementation, and other data issues associated with various ASD tools. Most importantly many of the ASD tools have not integrated machine learning techniques into their screening and diagnosis process. Therefore [5], highlighted the machine learning techniques used in large prevailing ASD diagnostic instruments along with their conceptual issues and data and features issues like data imbalances, and provides a series of promising recommendations for future developers to overcome those issues.

III. EMPIRICAL ANALYSIS

A. Data and Features

Controls and cases related to children (aged 4-11 years) have been collected using an ASD screening mobile application called ASDTests [40]. ASDTests was developed in 2017 to expedite ASD screening for different target groups including toddlers, children, adults and adolescents. In this paper, the focus is on instances related to the children category which have been collected based on the AQ-10-child ASD screening tool [4] using the ASDTests mobile application. Therefore, individual experiments were conducted on the children dataset only, which consists of 509 instances and 24 variables. The dataset has been obtained from its prospective author and covers the period between September 2017 and February 2018. Initially, the dataset was published in December 2017 with 292 instances at UCI data repository [41], but we were able to obtain from the dataset's owner the updated dataset with 227 child instances. The dataset contains 252 instances not on the spectrum (No ASD traits) and 257 instances with ASD traits; thus the dataset is somewhat balanced in regard to the target class variable. Initially, there were 24 independent variables including the target class. Most data instances relate to male participants with a ratio of 71.31 per cent (363 out of 509 instances). Moreover, 125 instances in the dataset were born with jaundice and 438 instances have been collected from parents. Table I depicts the primary variables that we have utilized prior to the data processing step. A number of variables have been discarded and not included in the table including: Country_of_Residence, Case_ID, Language, Screening_Type, Used_App_Before, since they have no added value and do not influence the classification of control and cases.

Independent variables A1-A10 shown in Table I correspond to the questions in the classic AQ-10-child screening tool and have been embedded within the ASDTests app. For simplicity, the authors of the dataset assigned these variables either "0" or "1" based on the answer given during the screening test by the participant. In particular, for questions 1, 5, 7, 10, "1" is assigned to the feature when the participant answers "Definitely" or "Slightly Agree" whereas, "1" will be given for "Definitely" or "Slightly Disagree" for questions 2, 3, 4, 6, 8 and 9. The dependent variable, which represents whether individuals have ASD traits, is associated with two possible values (Yes or No). This variable was assigned values based on the score obtained by individuals in the ASDTests app and was generated by the AQ-10-child tool. For a score larger than 6, "Yes" was assigned to the target variable for the instance, otherwise "No" was assigned. The process of assigning the values to the target variable was automated using the ASDTests app.

B. Settings

In this section, we investigate the performance of the fuzzy data mining algorithm called FURIA in detecting ASD traits for children and compare the performance with respect to different evaluation measures. To generalize the performance of FURIA, different data mining algorithms have been contrasted to reveal the upsides and the downsides of FURIA. In particular, we used JRIP, RIDOR and PRISM algorithms [25], [29] due to the fact they generate rules in the form of If-Then, as does FURIA, for fair comparison. In addition, these are rule-based data mining algorithms that have proved their merits in different classification applications, i.e. [42]-[44].

PRISM is a Covering algorithm that was developed to discover easy interpretable rules for decision-making by using a simple and effective metric called Expected Accuracy (EA). JRIP is a more advanced algorithm than PRISM that develops an optimization method and uses two subsets of data (growing, pruning) during the learning phase in order to reduce the number of rules generated. JRIP usually generates fewer rules than PRISM due to the pruning method implemented on the pruning set of data. RIDOR is a rule induction algorithm that generates exception in the format of rules. Lastly, FURIA is an extension of JRIP (RIPPER algorithm) which generates a fuzzy unordered set instead of classic ordered rules sets as JRIP. FURIA employs growing and pruning sets as JRIP in the process of rule learning and extraction. It learns rules sets per target class in a conventional strategy and then applies a stretch procedure to evaluate the rules sets derived. The outcome of FURIA is chunks of knowledge that can be used for decision-making especially in applications such as medical diagnosis. This is the primary reason for adopting FURIA to construct ASD classification models in order to detect ASD traits during the process of screening.

TABLE. I. FEATURES IN THE DATASETS

Variable No	Variable	Description
1.	A1	First question in AQ-10-Child screening tool
2.	A2	Second question in AQ-10-Child screening tool
3.	A3	Third question in AQ-10-Child screening tool
4.	A4	Fourth question in AQ-10-Child screening tool
5.	A5	Fifth question in AQ-10-Child screening tool
6.	A6	Sixth question in AQ-10-Child screening tool
7.	A7	Seventh question in AQ-10-Child screening tool
8.	A8	Eighth question in AQ-10-Child screening tool
9.	A9	Ninth question in AQ-10-Child screening tool
10.	A10	Tenth question in AQ-10-Child screening tool
11.	Age	Age of individual in numeric (years)
12.	Gender	Male or Female
13.	Ethnicity	Chosen from a list of predefined values
14.	Jaundice	Yes or No
15.	Family History	Whether any family members diagnosed with autism
16.	User	Who has taken the test (parent, self, relative, caregiver, etc)
17.	Target Class	The dependent variable (Yes/No). This variable was assigned based on the score obtained by individuals in the ASDTests app. If score larger than 6 "Yes" was assigned otherwise "No was assigned".

All experiments of the data mining algorithms and FURIA have been conducted on WEKA, a machine learning platform that contains useful data mining, pre-processing and learning techniques [32]. In addition, a ten-fold cross validation procedure was adopted to conduct the data processing experiments. Lastly, all experimental runs have been conducted on a personal computing machine with 2.3 GHz processor and 8 RAM of memory.

C. Results and Discussions

Different evaluation methods, such as predictive accuracy, specificity and sensitivity among others, have been utilized to report the learning algorithms performance in classifying ASD test instances from the child dataset. Predictive accuracy is a common performance measure in classification that reveals the percentage of test data that was correctly detected from the total number of test instances. On the other hand, sensitivity represents the percentage of the test instances that is truly positive, and specificity represents the test instances that are truly negative. The accuracy of FURIA and the considered data mining algorithms on the child dataset are shown in Fig. 1. The figure pinpoints that classification models generated by FURIA are more accurate in detecting ASD traits than the remaining algorithm. In particular, the classification model of FURIA outperformed models produced by JRIP, PRISM and RIDOR by 3.14%, 7.66% and 0.98% on the child autism dataset. A principal reason for the superiority of FURIA is the rules fuzzification process and the stretching procedure that takes into account the order of the rule's antecedent during the process of rule evaluation. This increases the rule's purity and possibly data coverage making FURIA favours a more general rule than those that are specific. The sensitivity rate obtained by the considered data mining algorithms on the child dataset is shown in Fig. 2. The sensitivity rates derived are consistent with the predictive accuracy results in which FURIA outperformed the considered

data mining algorithms. The sensitivity rate of FURIA is higher by 3.2%, 1.0% and 3.0% than JRIP, RIDOR and PRISM algorithms respectively. To evaluate the behaviour of FURIA we looked at the confusion matrix results obtained by its classification model. The confusion matrix results showed that only 14 instances with ASD traits have been incorrectly classified by FURIA as being without ASD traits, which is indeed a low number when compared with the remaining algorithms. To be specific, 42, 27, and 44 instances which are with ASD traits were misclassified by JRIP, RIDOR and PRISM algorithms. These numbers explain the higher predictive rate obtained by FURIA.

We investigated the false positives rates by deriving the specificity figures. Specificity (true negative rates) shows the percentage of participants who are without ASD and have been identified without ASD by the learning algorithm. Fig. 3 displays the specificity rates derived by the considered algorithms on the child dataset. Surprisingly, FURIA achieved lower specificity rates when compared with the remaining algorithms. We then investigated the false positive rates since they contribute largely in computing the specificity rate. From 252 instances, 33 which are actually without ASD have been misclassified by FURIA as being with ASD. In other words, there were 33 false positive instances generated by FURIA, compared with 18, 22 and 12 false positive instances generated by JRIP, RIDOR and PRISM algorithms respectively. These figures show that the specificity rate of PRISM is the highest, and the specificity rate of FURIA is the lowest, which is surprising. One possible reason for the higher false positive rates by FURIA and JRIP is the inability of this algorithm to differentiate among instances with limited ASD traits. These are instances that may show some autistic traits yet they are not classified to be on the spectrum by the screening tool. This shows a clear shortcoming of rule induction and fuzzy data mining algorithms, at least on the child data set considered in this paper.
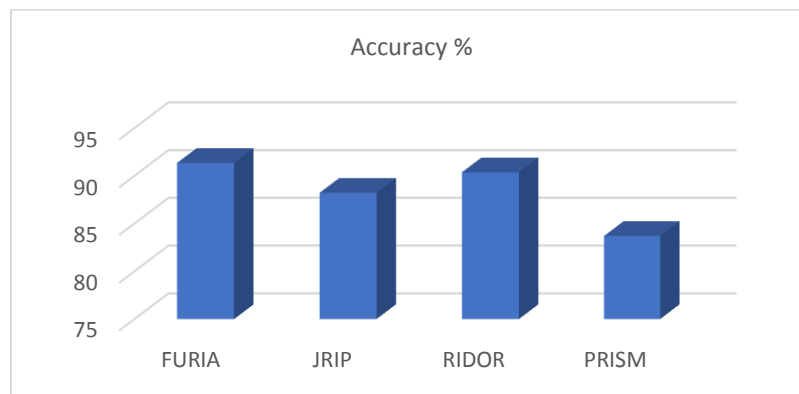
Fig. 1.   Predictive accuracy derived by FURIA and the other Considered Data Mining Algorithms.
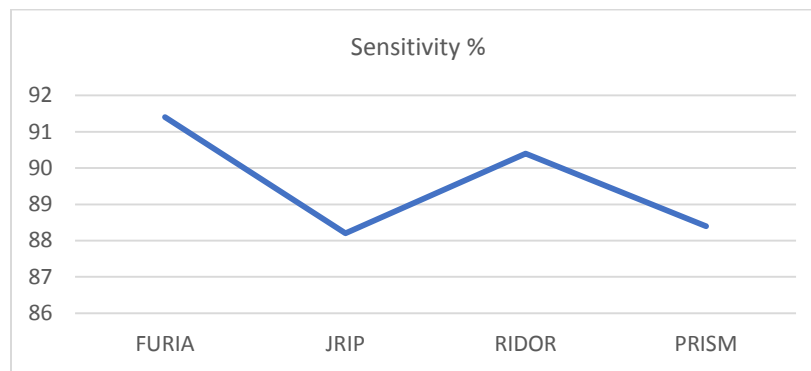

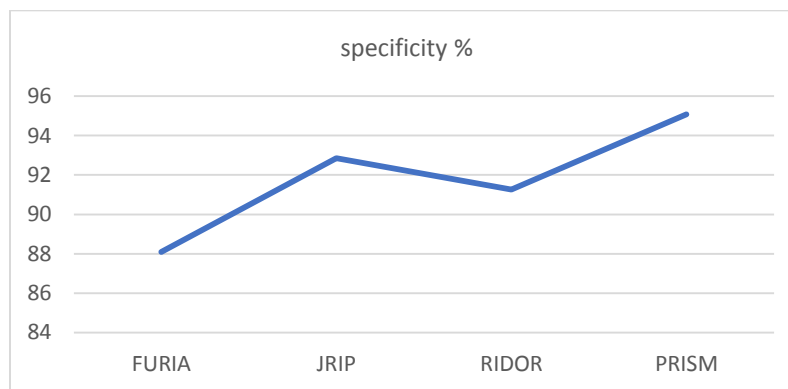Fig. 2.   Sensitivity rate derived by FURIA and the other Considered Data Mining Algorithms.

Fig. 3.   Specificity rate derived by FURIA and the other Considered Data Mining Algorithms.

The fuzzy sets produced by FURIA are shown below: 29 fuzzy rules were derived by FURIA from the child autism dataset in which 11 rules are connected with target class "yes" and the remaining rules with class "no". Based on the rules generated, the features related to AQ-10-child screening methods proved to be influential in detecting autistic traits particularly features such as A4, A7 and A9 appearing largely in the fuzzy rules sets. Specifically, features named A4, A7, A9, A2, A1, A10, A5, A3, A6 and A8 have appeared in the fuzzy rules sets 14, 11, 10, 10, 10, 12, 9, 9, 9, 9, respectively. This indicates that these features have high impact on detecting ASD traits and more important than demographic features in the child autism dataset.

Overall, FURIA produced useful chunks of knowledge that can be exploited by clinicians, parents, caregivers, and teachers among others, in understanding autism traits of children for better screening. When FURIA is integrated within screening tools of autism we expect that the automated fuzzy rules to be highly influential in detecting cases of autism for further referral and possibly to replace existing static domain expert rules.

FURIA rules:

==========

(A4 = 0) and (A8 = 0) and (A9 = 0) => Class=NO (CF = 0.99)

(A5 = 0) and (A10 = 0) and (A7 = 0) => Class=NO (CF = 0.98)

(A4 = 0) and (A1 = 0) and (A5 = 0) => Class=NO (CF = 0.98)

(A4 = 0) and (A10 = 0) and (Age in [-inf, -inf, 6, 7]) => Class=NO (CF = 0.98)

(A1 = 0) and (A2 = 0) and (A9 = 0) => Class=NO (CF = 0.98)

(A6 = 0) and (A5 = 0) and (A9 = 0) => Class=NO (CF = 0.98)

(A7 = 0) and (A3 = 0) and (A2 = 0) => Class=NO (CF = 0.96)

(A4 = 0) and (A7 = 0) and (A2 = 0) and (Family_ASD = no) => Class=NO (CF = 0.97)

(A10 = 0) and (A2 = 0) and (A1 = 0) => Class=NO (CF = 0.97)

(A8 = 0) and (A5 = 0) and (A3 = 0) => Class=NO (CF = 0.98)

(A6 = 0) and (A7 = 0) and (A10 = 0) => Class=NO (CF = 0.98)

(A4 = 0) and (A8 = 0) and (A1 = 0) => Class=NO (CF = 0.99)

(A9 = 0) and (A7 = 0) and (A8 = 0) and (A1 = 0) => Class=NO (CF = 0.98)

(A4 = 0) and (A6 = 0) and (A8 = 0) and (A2 = 0) => Class=NO (CF = 0.98)

(A9 = 0) and (A3 = 0) and (A4 = 0) => Class=NO (CF = 0.99)

(A9 = 0) and (A7 = 0) and (A5 = 0) and (Jaundice = no) => Class=NO (CF = 0.98)

(A10 = 0) and (A2 = 0) and (Family_ASD = yes) => Class=NO (CF = 0.94)

(A6 = 0) and (A3 = 0) => Class=NO (CF = 0.99)

(A4 = 1) and (A5 = 1) and (A9 = 1) and (A10 = 1) => Class=YES (CF = 0.99)

(A8 = 1) and (A1 = 1) and (A3 = 1) and (A5 = 1) => Class=YES (CF = 0.96)

(A4 = 1) and (A7 = 1) and (A3 = 1) and (A6 = 1) and (A2 = 1) => Class=YES (CF = 0.99)

(A8 = 1) and (A10 = 1) and (A1 = 1) and (A6 = 1) and (A4 = 1) and (A3 = 1) => Class=YES (CF = 0.99)

(A7 = 1) and (A9 = 1) and (A1 = 1) and (A6 = 1) and (A10 = 1) => Class=YES (CF = 0.99)

(A4 = 1) and (A10 = 1) and (A7 = 1) and (A3 = 1) => Class=YES (CF = 0.99)

(A9 = 1) and (A10 = 1) and (Age in [-inf, -inf, 5, 6]) and (A8 = 1) and (A7 = 1) => Class=YES (CF = 0.98)

(A4 = 1) and (A10 = 1) and (A2 = 1) and (Ethnicity = asian) and (Age in [-inf, -inf, 7, 10]) => Class=YES (CF = 0.94)

(A9 = 1) and (A2 = 1) and (A1 = 1) and (A3 = 1) => Class=YES (CF = 0.97)

(A4 = 1) and (A10 = 1) and (A2 = 1) and (A5 = 1) and (A1 = 1) and (A6 = 1) => Class=YES (CF = 0.99)

(A8 = 1) and (Jaundice = yes) and (A5 = 1) and (A7 = 1) and (A6 = 1) and (A10 = 1) => Class=YES (CF = 0.97)

IV. CONCLUSION AND FUTURE WORKS

Autism Spectrum Disorder (ASD) is one of the growing neurodevelopment conditions worldwide with many individuals undetected, making early screening crucial for individuals, family members and physicians. Most of the existing ASD methods consist of a large set of questions covering communication, social and repetitive behaviours and rely on domain expert rules with a basic scoring function to detect autistic traits. One promising approach that can automate the process of ASD screening and improve the accuracy and efficiency of the detection is the use of fuzzy data mining. In this paper, the Fuzzy Unordered Rule Induction algorithm (FURIA) has been evaluated for ASD traits detection. FURIA builds screening models in an automated way from historical controls and cases and then utilizes the models to detect the possibility of autistic traits in new individuals. The key strength of FURIA screening models is the fact that they contain useful chunks of knowledge (fuzzy rules) that not only clinicians and other medical staff can interpret but also family members, teachers and caregivers. These fuzzy rules are a source of information that can help different stakeholders understand the main influential factors for ASD and therefore proper individualized plans can be planned and developed to cater to the needs of people who fall within the spectrum. Empirical results based on real data collected recently from children between 4-11 years old using a mobile application called ASDTests, revealed that FURIA fuzzy rules were able to detect ASD traits with up to 91.35% classification accuracy and 91.40% sensitivity rate; these results were superior to other Greedy and Rule Induction techniques. Despite FURIA producing an acceptable specificity rate, i.e. 88.09%, other data mining techniques generated better specificity results.

One of the limitations of this study is not extensively considering feature assessment on the dataset and not considering other target datasets such as infants, adolescent and adults.

In near future, we are going to apply the fuzzy rules on datasets related to infants and adolescents and seek whether the performance will be sustained. In addition, we will investigate features that are similar among different age categories.

REFERENCES

[1] American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders DSM-IV-TR (Text Revision). Washington: American Psychiatric Association.

[2] Biklen, D. (2005). *Autism and the myth of the person alone*. New York and London: New York University Press.

[3] Christensen, D., Baio, J., Braun, K., Bilder, D., Charles, J., Constantino, J., & Yeargin-Allsopp, M. (2014). *Prevalence and characteristics of autism spectrum disorder among children aged 8 years* — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. Surveillance Summaries, 1–23.

[4] Allison, C., Auyeung, B., and Baron-Cohen, S. (2012). Toward brief "Red Flags" for autism screening: The short Autism Spectrum Quotient and the short quantitative checklist for autism in toddlers in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child Adolescent Psychiatry 51*(2), 202–12Y.

[5] Thabtah, F. (2018) Machine learning in autistic spectrum disorder behavioural research: A review and ways forward. *Informatics for Health and Social Care 43* (2), 1-20.

[6] Duda, M., Ma, R., Haber, N., & Wall, D. (2016). Use of machine learning for behavioural distinction of autism. *Translational Psychiatry*, 221.

[7]   Liu, W., Yu, Z., Raj, B., Yi, L., Zou, X., & Li, M. (2015). Efficient Autism Spectrum Disorder Prediction with Eye Movement: A Machine Learning Framework. 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), 649-655.

[8]   Kosmicki, J., Sochat, V., Duda, M., & Wall, D. (2015). Searching for a minimal set of behaviours for autism detection through feature selection-based machine learning. *Translational Psychiatry*, 1-7.

[9]   Wall, D., Kosmicki, , J., DeLuca, T., Harstad, E., & Fusaro, V. (2012a). Use of machine learning to shorten observation-based screening and diagnosis of autism. *Traditional Psychiatry, 2*(4):e100.

[10]   Mohammad, R., Thabtah, F., McCluskey, L. (2016) *An improved self-structuring neural network.* Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, 35-47. Auckland, New Zealand.

[11]   Aburrous, M., Hossain, M. A., Dahal, K. & Thabtah, F. (2010) *Predicting phishing websites using classification mining techniques with experimental case studies.* Proceedings of the Information Technology: New Generations (ITNG). Las Vegas, Nevada, USA.

[12]   Tejwani, R., Liska, A., You, H., Reinen, J., & Das, P. (2017). *Autism classification using brain functional connectivity dynamics and machine learning*. California, Santa Barbara: Cornwall University Library.

[13]   Thabtah, F. (2017a). *Autism spectrum disorder screening: Machine learning adaptation and DSM-fulfilment.* Proceedings of the 1st International Conference on Medical and Health Informatics 2017, pp.1-6. Taichung City, Taiwan, ACM.

[14]   Heinsfeld, A. S. (2017). *Identification of autism disorder through functional MRI and deep learning.* Rio Grande Do Sul: Pontifical Catholic University of Rio Grande Do Sul.

[15]   Leroy, G., Irmscher, A., Charlop-Christy, M., Kuriakose, S., Pishori, A., Wurzman, L., . . . Stephanie, B. (2006). *Data mining techniques to study therapy success*. International Conference on Data Mining, 26 - 29 June 2006 (pp. 1-7). Las Vegas, USA: Institute of Electrical and Electronics Engineers (IEEE).

[16]   Angelo, G., Rao, D., & Gu, C. C. (2009). Combining least absolute shrinkage and selection operator (LASSO) and principal-components analysis for detection of gene-gene interactions in genome-wide association studies. *Genetic Analysis Workshop 16*, 25-63.

[17]   Beykikhoshk, A., Arandjelovi´, O., Phung, D., Venkatesh, S., & Caelli, T. (2014). *Data-mining twitter and the autism spectrum disorder: A pilot study.* 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014) (pp. 349 - 356). IEEE Conferences.

[18]   Thabtah, F. (2017c). *ASD dataset for children*. https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Scr eening+Data+for+Children++ [accessed January 15th, 2018].

[19]   Biklen, D. (2005). *Autism and the myth of the person alone*. New York and London: New York University Press.

[20]   Hhn J., Hllermeier E. (2009) FURIA: An algorithm for unordered fuzzy rule induction. *Data Mining Knowl. Discovery, 19,* (3), pp. 293-319, 2009.

[21]   Bone, D., Goodwin, M., Black, M., Lee, C. C., Audhkhasi , K., & Narayanan, S. (2014). Applying machine learning to facilitate autism diagnostics: Pitfalls and promises. *Journal of Autism and Developmental Disorders, 1-48.*

[22]   Wall, D., Dally, R., Luyster, R., Jund, J., & Deluca, T. (2012b). Use of artificial intelligence to shorten behavioural diagnosis of autism. Plos One, 7(8).

[23]   Geschwind, P., Sowinski, J., Lord, C., Iversen, P., Shestack, J., Jones, P. Spence, S. (2001). The autism genetic resource exchange: A resource for the study of autism and related neuropsychiatric conditions. *American Journal of Human Genetics, 69*(2), 463.

[24]   Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The autism diagnosis observation schedule: Revised algorithm for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, 613-627.

[25]   Cohen W. W. (1995) *Fast effective rule induction*, Proceedings of the 12th International Conference on Machine Learning, pp. 115-123, 1995

[26]   De la Iglesia-Vaya, M., Molina-Mateo, J., Jose, M., & Marti-Bonmati, L. (2013). Brain connections - resting state fMRI functional connectivity. *Novel Frontiers of Advanced Neuroimaging*, 236-286.

[27]   Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P. A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 3371–3408.

[28]   Nielsen, J., Zielinski, B. A., Fletcher, P., Alexander, A. L, Lange, N, Bigler, E. D, . . . Anderson, J. S. (2013). Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in Human Neuroscience*, 1-12.

[29]   Cendrowska J., PRISM: an algorithm for inducing modular rules. *Int. J. Man. Mach. Stud., 27* (1987) 349-370.

[30]   Langseth, H., & Nielsen, T. (2006). Classification using hierarchical naïve Bayes models. *Machine Learning*, 135–159.

[31]   Breiman, L. (2001). *Random forests.* Machine Learning, 5–32.

[32]   Smola, A., & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing,* 199–222.

[33]   Chaudhuri, B., & Bhattacharya, U. (2000). Efficient training and improved performance of multilayer perceptron in pattern classification. *Neurocomputing*, 11-27.

[34]   Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. (2009). The WEKA data mining software: An update. *SIGKDD Explorations 11*(1).

[35]   Fischbach, G. D., Lord C. (2010) The Simons simplex collection: A resource for identification of autism genetic risk factors. *Neuron* (68), 192–195.

[36]   Peng, H. C., Long, F., & Ding, C. (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(8, pp. 1226–1238, 2005.

[37]   Barros, A., & Hirakata, V. (2003). Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology,* 569-598.

[38]   Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., (2011). Scikit-learn: machine learning in python. *J Mach Learn Res 2011; 12*: 2825–2830

[39]   MacQueen, J. B. (1967). *Some methods for classification and analysis of multivariate observations.* Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1. University of California Press. pp. 281–297. MR 0214227. Zbl 0214.46201. Retrieved 2009-04-07.

[40]   Thabtah, F. (2017b). *ASDTests. A mobile app for ASD screening.* www.asdtests.com [accessed March 10th, 2017].

[41]   Olanow, C., & Koller, W. (1998). An algorithm (decision tree) for the management of Parkinson's disease. *Neurology,* 50-69. Barros

[42]   Abdelhamid, N., Thabtah, F., & Abdel-jaber, H. (2017). *Phishing detection: A recent intelligent machine learning comparison based on models content and features.* 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 72-77. 2017/7/22, Beijing, China.

[43]   Mohammad, R., Thabtah, F., McCluskey, L. (2014). Intelligent rule-based phishing websites classification. *IET Information Security, 8*(3): 153-160.

[44]   Abdelhamid, N., and Thabtah, F. (2014). Associative classification approaches: Review and comparison. *Journal of Information and Knowledge Management (JIKM), 13*(3).