ABSTRACT

Autism Spectrum Disorders are not hard to detect, but it requires that physicians have proper training and experience. At the moment however Autism Spectrum Disorder is detected much later than is actually possible. Early detection of Autism Spectrum Disorder improves the overall mental health of the child. In this research we use machine learning to determine a set of conditions that together prove to be predictive of Autism Spectrum Disorder. This will be a of great use to physicians, helping them detect Autism Spectrum Disorder at a much earlier stage. This will be done through literature review, data exploration and evaluation. Predicting if a child has Autism Spectrum Disorder proved possible by using developmental delay, learning disability and speech or other language problems as attributes and also include physical activity, premature birth and birth weight to improve the accuracy. Using the 1-away method it was also possible to predict the severity of Autism Spectrum Disorder quite reasonably. The 1-way method improved the accuracy from 54.1% to 90.2%, which is a significant increase. This and the fact that the severity was based on input from just the caretakers of the children, prompts the need for further research in this matter.

Keywords

Autism Spectrum Disorder, ASD, National Survey of Children’s Health, NSCH

1. INTRODUCTION

“Autism Spectrum Disorder is a neuro-developmental disorder characterized by persistent deficits in social interaction and communication and restricted, repetitive patterns of behavior, interests or activities” [10]. Autism Spectrum Disorder (ASD) prevales in many forms, from very mild ASD to very severe ASD, depending on the severity of the symptoms. In their paper, K et al. also stated there is a need for earlier recognition, evaluation and intervention of Autism Spectrum Disorder. Johnson et al. [9] also stressed the fact that early intervention programs are very beneficial for the overall well-being of a child. Previous research has shown that in children as young as 18 months ASD can be detected reliably when using 2-stage screening strategies [19]. That study also showed that the average age at first diagnosis is between 3 to 6 years. This gap between possible age at first diagnosis(18 months) and the average age at first diagnosis is a gap that should be closed as much as possible. As stated before, physicians can use certain reliable formal screening tools to increase the accuracy of the estimate regarding the developmental status of children. However, only a minority of physicians use those tools that are available[22]. One way to improve the accuracy would be routine developmental screenings for all children[11]. These kind of routine screenings might prove to be costly.

An alternative approach is developing a decision support system, which can be used by physicians to help them estimate the developmental status of children more accurately. Recent research has predicted cases of child abuse using structured data with reasonable performance[5][7]. The question that rises is if it would be possible to use structured data to predict ASD in an earlier stage. Using structured data we propose a decision support based approach to help physicians improve the accuracy of their estimates regarding the developmental status of children and specifically in the case of ASD.

The structured data that is used, is gathered from the 2011-2012 National Survey of Children’s Health (NCSH), which was conducted in the United States of America[17]. This survey includes children from the age of 2 to 17 across every state in the United States of America and contains answers from primary caretakers of these children.

1.1 Problem Statement

Machine learning has been used to predict cases of child abuse using structured data and textual data[2]. This has however, not been done often and hardly ever been done for ASD and developing a decision support system that helps physicians with the detection of ASD, has scarcely been done, which is proven by the lack of literature on the use of it.

1.2 Research questions

The problem statement above leads to the following main research question:

How can Machine Learning be used to detect Autism Spectrum Disorder?

This question can be divided in the following research questions:

1. Which conditions co-occur with Autism Spectrum Disorder?

2. Does data from the National Survey of Children’s
Health contain co-occurring conditions with Autism Spectrum Disorder?

3. Are the co-occurring conditions predictive of Autism Spectrum Disorder, using machine learning?

2. RESEARCH METHODS
The research performed in this paper contains several steps. Firstly, a literature review will be conducted to determine which conditions co-occur with ASD and to determine which machine learning methods should be used. With this literature review we will create an overview of co-occurring conditions with ASD and select the machine learning methods that are to be used. Secondly, the NSCH data will be explored and it will be determined if those co-occurring conditions are represented in the data and if any other interesting facts arise from the data. Finally, the indicators determined from the previous step will be applied to the data using machine learning to determine whether they truly are predictive of ASD. When the indicators are not sufficient enough we will go back to step one and find more conditions and keep repeating these steps until we have proper results.

2.1 Literature review
The literature review in this paper will be done by applying the five-stage process for reviewing literature as proposed by Wolfswinkel et al[26]. A summary of this process is shown in Table 1.

<table>
<thead>
<tr>
<th>Number</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. DEFINE</strong></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>Define the criteria for inclusion/exclusion</td>
</tr>
<tr>
<td>1.2</td>
<td>Identify the fields of research</td>
</tr>
<tr>
<td>1.3</td>
<td>Determine the appropriate sources</td>
</tr>
<tr>
<td>1.4</td>
<td>Decide on the specific search terms</td>
</tr>
<tr>
<td><strong>2. SEARCH</strong></td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Search</td>
</tr>
<tr>
<td><strong>3. SELECT</strong></td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Refine the sample</td>
</tr>
<tr>
<td><strong>4. ANALYZE</strong></td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Open coding</td>
</tr>
<tr>
<td>4.2</td>
<td>Axial coding</td>
</tr>
<tr>
<td>4.3</td>
<td>Selective coding</td>
</tr>
<tr>
<td><strong>5. PRESENT</strong></td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Represent and structure the content</td>
</tr>
<tr>
<td>5.2</td>
<td>Structure the article</td>
</tr>
</tbody>
</table>

This literature review will provide us with a basis to perform the rest of the research on and will enable us to investigate the validity of our research questions.

2.2 Data exploration
Since the data set contains 367 different attributes and over 95577 records, the data has to be cleaned before the data can be used for our machine learning process. After cleaning the NSCH data, the data will be explored to determine if there are co-occurring conditions present in the data. It will also be investigated if there are any clusters present that might help us predict ASD. An attribute selection algorithm will be used on the data set, to determine if there are attributes that seem strongly correlated to ASD. The method that will be used to determine the relevant attributes is correlation based feature selection, which has been proven to be very good for this task[8][14].

2.3 Evaluation
To determine whether the co-occurring conditions found in the literature review and the data exploration process, are predictive of ASD, machine learning will be applied. Instead of using a common even split method, we will use 10 fold cross validation to determine if we can accurately predict Autism Spectrum Disorder. 10-fold cross-validation is a great way to make sure the tests are less biased and will better predict the error estimation [21][12]. This k-fold cross-validation method is illustrated in Figure 1. 10-fold cross-validation is more reliable than a common even split (50/50) method, because it will be less biased. As can be seen in Figure 1, instead of randomly splitting the data in half and using one half as training set and the other half as test set, 10-fold cross-validation will split the data set in 10 pieces and run the algorithm 10 times. Each time a different piece of the data is the test set and the other 90 percent of the data is the training set. Afterwards the average of all 10 runs of the algorithm is your result. This result is a lot less biased than the even split method's result.

![Figure 1. K-Fold Cross-Validation](image)

3. LITERATURE REVIEW
This literature review will look into co-occurring conditions with ASD and machine learning methods that can be used to analyze structured data sets. The review will be done based on the earlier introduced research questions and will follow the review method as proposed by Wolfswinkel et al[26].

3.1 Co-occurring conditions with ASD
As shown in the introduction, there are good ways to detect ASD. Physicians however often don’t use those tools as early as they could [22]. A physician might diagnose a child with some minor developmental delay and then wait for a few months to see how it further develops. They don’t want to jump to conclusions too early. To be able to answer our main research question, we will propose a list of co-occurring conditions with ASD based on previous research. These co-occurring conditions can then be used to help us predict ASD.

After reviewing many papers, a few of them stood out regarding how strongly they reported the co-occurring conditions to be with ASD. Setoh et al. [23] show that developmental delay of children is something that often co-occurs with ASD. Using the 2011-2012 NSCH, McCoy et al. determined that children with ASD are more likely to be obese and are physically less active[16]. A research performed in Norway showed similar results regarding lower...
physical activity of adolescents (13-18 years old) with ASD[15]. Another study into the 2011-2012 NSCH showed that children with very low birth weight (<1500 g) had 3.2 times higher odds of autism/ASD than normal birth weight children [24]. They also determined that children born prematurely had 2.3 times higher odds of autism/ASD than term children. Yates and Couteur propose that difficulties and delay in social interaction are often the earliest features in ASD, but that they are also easily missed, because they are often subtle[28].

The findings in this literature review into co-occurring conditions with ASD, provides us with the following list of co-occurring conditions:

- Developmental delay
- Obesity
- Less physical activity
- Very low birth weight
- Premature birth
- Social interaction delay

There were many more possible co-occurring conditions with ASD, but they didn’t make the terms of inclusion. By applying the five-stage grounded-theory method for reviewing literature by Wolfsink et al, the conditions mentioned above remained and will be used during the machine learning process if possible.

3.2 Machine Learning methods

There are many classification algorithms that can be used for the type of classification task we will perform in this research. The most popular algorithms [1] will be used for this classification task. Although Aggarwal and Zhai propose that these are the most popular algorithms for text classification, they also state they are just as capable to classify the structured data we have in this research. The classification algorithms to be used are: Naive Bayes [13], Random Forest[3] and Support Vector Machine[4]. Besides these three popular algorithms, we will also use an older method J48 [20], which is a java implementation of a well known algorithm called C4.5[27], to be able to compare the Random Forest method with a different tree algorithm. Using these diverse algorithms will ensure our results are more reliable and help us determine whether a certain algorithm is not usable for this classification task.

4. DATA EXPLORATION

The NSCH data set contains 95577 records of children, with 367 variables. Because only a small percentage of the dataset was positively recognized as having ASD (around 2000 children), we used the random under-sampling approach[6]. Using this approach we created a data set which contains roughly 50 percent children with ASD and 50 percent children without ASD. By pre-processing this data set, we managed to reduce the amount of variables from 367 to 256, which still is a large amount to use as input for the classification algorithms.

4.1 Attribute selection

The next step consisted of applying the correlation based feature selection method mentioned earlier in section 2.2. This method resulted in only 16 remaining attributes. Not all of these attributes were used for the evaluation step, because it was determined that some of the attributes were too closely related to the actual class of the child (either autistic or not). By excluding those leading attributes and actually including some attributes that we determined earlier by performing our literature review the following attributes remained:

- Learning disability
- Developmental delay
- Speech or other language problems
- Birth weight
- Prematurely born
- Physical activity
- Attendance to religious events
- Body Mass Index

We had to exclude social interaction delay from the conditions found through literature review, because this was not represented in the data set. The attributes are visualized in Figure 2. The blue color means they are autistic and the red color means they are not. A lower score means they do not have, for example, a learning disability. A higher score means they do.

4.2 Data sets

After exploring the data it became clear that there was a possibility to create a data set that had 2 classes (no ASD or ASD) and a data set that had 4 classes (no ASD, mild ASD, moderate ASD and severe ASD). Both sets will be used for our machine learning and the results will be compared to each other.

5. RESULTS

Below, the results from machine learning are described. First we divided our data set in only 2 classes, namely no ASD or ASD. In a further iteration we divided the ASD class in 3 separate classes, which caused our data set to have 4 classes: no ASD, mild ASD, moderate ASD and severe ASD.

5.1 Machine learning results using 2 classes

We applied our 4 previously determined classification algorithms to the data set when it was divided in only 2 classes, either a child had ASD or a child did not have ASD. This was done by using the attributes determined in the data exploration step. The results of these classification algorithms (Naive Bayes(NB), Support Vector Machines(SVM), J48 and Random Forest(RF)) are summarized in Table 2.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision</th>
<th>Accuracy</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>0.865</td>
<td>0.865</td>
<td>0.865</td>
</tr>
<tr>
<td>SVM</td>
<td>0.835</td>
<td>0.833</td>
<td>0.833</td>
</tr>
<tr>
<td>J48</td>
<td>0.871</td>
<td>0.871</td>
<td>0.871</td>
</tr>
<tr>
<td>RF</td>
<td>0.854</td>
<td>0.851</td>
<td>0.851</td>
</tr>
</tbody>
</table>

These results show that the chosen attributes prove to be very predictive of Autism Spectrum Disorder, with an average correctly predicted classes of roughly 86%. After fine
Figure 2. Overview of attributes
tuning the algorithms it became clear certain attributes where linked much stronger to ASD than others. When looking further at the breakdown of these attributes in regards to the 2 classes it can be concluded that Developmental delay, Learning disability and Speech and other language problems are strongly linked to ASD. Physical activity is also linked to ASD, but much less so than the aforementioned attributes. Premature birth, birth weight and attendance to religious events are even less valuable to our classification task and Body Mass Index is not used at all. These conclusions can be derived from examining the decision tree generated from the fine tuned J48 algorithm, which can be seen in Figure 3.

5.2 Machine learning results using 4 classes

After seeing the positive results when using 2 classes, we were keen to see if we could use this data set in other ways. The data set contained information about the severity of the child’s ASD, which allowed us to divide the data set in 4 classes. A child had either no ASD, mild ASD, moderate ASD, severe ASD. The severity that is reflected in the data set is based on a question asked to the caretaker of the subject child. The same attributes as those used for the machine learning process when using 2 classes, were used to predict the severity of ASD (4 classes). After applying the same algorithms (NB, SVM, J48 and RF) it looked like these attributes were not sufficient enough to predict the severity of ASD. This is clearly shown in Table 3, which shows that it only correctly predicted the class in about 50 percent of the cases.

We tried to increase this number by including other attributes, but this didn’t prove fruitful. The accuracy went up to about 70 percent, but it was evident that this happened because leading attributes were included.

We went back to the original attributes that proved predictive for 2 classes and applied the 1-away method [18]. By applying this method to our most successful algorithm J48 we increased the accuracy from 54.1% to 90.2%. The results are shown in Table 4.

The overall percentages were calculated as follows:

**Exact match** = 54.1%  
Number of hits / total cases = 1216 / 2247  
1-away match = 90.2%  
(number of hits + number of 1-away hits) / total cases = 2027 / 2247

These results indicate that the used machine learning process is actually quite good at predicting the severity of ASD.

### Table 3. Performance scores for classifier using 4 classes

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision</th>
<th>Accuracy</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>0.479</td>
<td>0.512</td>
<td>0.512</td>
</tr>
<tr>
<td>SVM</td>
<td>0.475</td>
<td>0.493</td>
<td>0.493</td>
</tr>
<tr>
<td>J48</td>
<td>0.524</td>
<td>0.541</td>
<td>0.541</td>
</tr>
<tr>
<td>RF</td>
<td>0.489</td>
<td>0.507</td>
<td>0.507</td>
</tr>
</tbody>
</table>

6. DISCUSSION & CONCLUSION

ASD is a neuro-developmental disorder that has many (small) symptoms and therefore is hard to detect on first sight. From literature and data exploration, a set of conditions that co-occur with ASD could be identified and then be used in our machine learning process. After taking several iterations in our machine learning process, it became clear that only a few attributes were necessary for our classification algorithms, which rises the question whether these attributes are indeed leading towards ASD. Our machine learning algorithms did however show significant promise and warrants further research.

6.1 Predicting ASD or no ASD

When the population was split in 2 almost equal size groups, where one half had some form of ASD and the other half had no ASD, we could apply our algorithms to good success. It showed that strongly co-occurring conditions are developmental delay, learning disability, speech or other language problems and it also showed that some co-occurring conditions were also influential but to a lesser degree, such as physical activity, premature birth and birth weight. The other conditions, Body Mass Index and attendance to religious events didn’t help predicting ASD. Our machine learning process generated a simple tree, which shows that developmental delay, learning disability and speech or other language problems are really strongly linked to ASD and if a child has a certain combination of those conditions, he or she should definitely do a formal screening for ASD.

6.2 Predicting severity of ASD

After proving that it is possible to predict whether someone is likely to have ASD or not, we proceeded to divide the data into 4 groups, namely No ASD, Mild ASD, Moderate ASD and Severe ASD. With this we tried to actually predict the severity of ASD. This scale from No ASD to Severe ASD was derived from our data set, but supporting documents of the data set did not describe what this scale of severity actually meant. Those classes are based on the judgment of the primary caretakers of the subject child, which raises questions to the validity. Only a small group of the children had severe ASD, which makes it harder to properly predict the severity. After applying the 1-away method though, we got some promising results for the 4 classes. This led us to the conclusion that our proposed co-occurring conditions are, besides useful to predict if a child has Autism Spectrum Disorder, also predictive of the severity of ASD. The problem however is that the 1-away method on only 4 classes is really influential in the results. Therefore we conclude that our proposed co-occurring conditions are not sufficient enough to determine the severity of ASD, but are well suited to determine if a child would need further formal screenings for ASD. With these screenings, physicians could then determine the severity of ASD and help in the child’s development.

6.3 Conclusion

Our goal was to determine a set of conditions that prove to be predictive of ASD, which then can be used by physicians to help them decide if they should do a formal screening for ASD. By using the co-occurring conditions developmental delay, learning disability, speech or other language problems and physical activity you can quite accurately predict whether a child has ASD or not and even how severe the ASD might be. We believe the decision making process for physicians will be easier and ASD in children can be detected much earlier than it is currently. This will increase the overall well-being of the children with ASD. To determine whether a child has certain problems a physician should also rely on the input of the parents [25] as they have much more insight in some of these conditions. It is also important to mention that this NSCH data set was a survey held under caretakers of children. As Ward, Sullivan and Gilmore (2016) mentioned, including
both parent and physician ratings of behavioural indicators would almost certainly improve the diagnostic classification.

7. FUTURE WORK
This research focused on determining a way to predict whether a child had ASD or not. We managed to do this, but new challenges and questions arose during this research. When splitting up the ASD group into 3 gradations of ASD (mild, moderate, severe) it showed that it is a lot harder to predict the severity of ASD. Predicting the severity of ASD was possible with this kind of data set after applying the 1-away method, but the big increase in accuracy does warrant further research. This research could verify whether the 1-away method is actually useful in this case, or should not be used. It could also look into other attributes that are linked better to the severity of ASD.

We also recommend testing the proposed attributes on data sets which include data of the same child over time to test the true predictive value of these attributes. This would require data sets from large health organizations or gathered from a large research.

Furthermore future work should focus on verifying our findings by using the proposed decision tree in practice or test it on other data sets. The proposed decision tree
could for example be tested on data sets that include textual data. Running the same algorithms on data sets generated from the input of physicians would be a nice way to compare both types of data.

Another important step would be testing these attributes to more classes than just ASD. It might very well be that developmental delay, learning disability and speech or other language problems are predictive of other disorders as well or more predictive of those disorders and therefore less predictive of ASD. We believe it’s important that these conditions are tested on this data set as well, but this was outside the scope of this research.

8. ACKNOWLEDGEMENTS
I would like to thank Chintan Amrit for his guidance and support during this research. I’d also like to thank Miha Lavric for his ideas and input. Hopefully my research can assist them on their future research into the subject.

9. REFERENCES

Figure 3. J48 decision tree

Table 4. J48 algorithm, 1-away method applied

<table>
<thead>
<tr>
<th>Classes</th>
<th>No ASD</th>
<th>Mild ASD</th>
<th>Moderate ASD</th>
<th>Severe ASD</th>
<th>Exact match</th>
<th>1-away</th>
</tr>
</thead>
<tbody>
<tr>
<td>No ASD</td>
<td>589</td>
<td>155</td>
<td>47</td>
<td>3</td>
<td>74.2%</td>
<td>93.7%</td>
</tr>
<tr>
<td>Mild ASD</td>
<td>161</td>
<td>484</td>
<td>146</td>
<td>17</td>
<td>59.9%</td>
<td>97.9%</td>
</tr>
<tr>
<td>Moderate ASD</td>
<td>50</td>
<td>291</td>
<td>131</td>
<td>9</td>
<td>27.2%</td>
<td>89.6%</td>
</tr>
<tr>
<td>Severe ASD</td>
<td>9</td>
<td>94</td>
<td>49</td>
<td>12</td>
<td>7.3%</td>
<td>37.2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>54.1%</td>
<td>90.2%</td>
</tr>
</tbody>
</table>


