IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, May 2022
Impact Factor: 6.252

Classification Autistic Spectrum Disorder with Selected Features

Khushboo Verma[1] and Abhinav Shukla[2]
Research Scholar, Department of IT and CS[1]
Associate Professor, Department of IT and CS[2]
Dr C V Raman University, Bilaspur, Chhattisgarh, India
Corresponding Author: vermakhushboo208@gmail.com[1]

Abstract: *In this paper, feature selection technique (FST) namely Chi-Square (CS) has been used for feature selection. The filter based CS is a ranking method. The FST key goals of improving classification efficiency and reducing feature counts. Naive Bayes (NB), K-Nearest-Neighbour (K-NN) and Support Vector Machine (SVM) with RBF kernel considered the classification methods on Autistic Spectrum Disorder (ASD) children dataset. Comparison to the non-reduced features and reduced feature of ASD datasets the reduced feature give up enhanced results in all classifiers NB, K-NN and SVM. Finally, minimum feature with high accuracy based classification model is proposed.*

Keywords: Autistic Spectrum Disorder, Chi-Square (CS), Classification, Feature selection technique (FST), Naive Bayes (NB)

I. Introduction

The Autistic Spectrum Disorder (ASD) is a group of Neuro-developmental disabilities disorders that cannot be cured but may be ameliorated by early detection with the help of data mining. To deal with the ASD problem, data mining structure used for ASD screening applying a class of learning NB, K-NN and SVM. Access to and use of ASD screening tools is critical as they can reduce wait times for a formal clinical evaluation of and provide people on the spectrum and their families with understanding. Know better about the resources and support services needed (special education speech therapy environmental work etc.). However most of the screening tools available today are based on diagnostic methods containing a large number of items for which a parent cares give or individual is required for verification. Recently a number of researchers in the field of ASD research have investigated data mining to improve the classification time of ASD diagnosis or to detect the factors that most influence the diagnosis of ASD. The Chi-square is a filter based ranking method is used to rank the ASD important features. The main goals of FST are to improve classification efficiency and reduce the number of features. Naive Bayes (NB), K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) with RBF kerenl evaluation classification methods on the Autistic Spectrum Disorders (ASD) Children dataset. The proposed classification model and with FST CS assist in a quick identification and early detection of ASD disease.

II. Literature Review

Tha bath et al.(2020) presented classification method like PRISM, CART, AdaBoost, Bagging, Nnge, RIDOR, Rules Machine Learning (RML), RIPPER and C 4.5 to classify 3 ASD datasets (children adolescents and adults). The result shows that RML performs higher compared to other methods (Thabtah & Peebles, 2020). Wang et al. (2019) worked on the feature engineering and feature encoding techniques, along with a deep learning classifier for ASD screening. The experimental result revealed that 99% sensitivity and 99% specificity (Wang, Li, Chi, & Zhao, 2019). Diabat used classifiers C4.5, PART, RIPPER and Voted Perceptron and Ensemble Classification for Autism Screening (ECAS). Thee Ensemble Classification obtained the highest accuracy 100% compared to other models (Diabat & Al-shanableh, 2019). Akyol et al. proposed the models Fuzzy Rule (FR) and Combination of Logistic Regression wit Fuzzy Rule (LR-FR) for classification of ASD problems. The experimental result clear that the LR-FR obtained the accuracy 97.33% was better than FR (Akyol, Gultepe, & Karaci, 2018). Vaishali et al. (2017) used the Binary Firefly Algorithm (BFA) FST wrapper and classifiers Naive Bayes,J48 Decision Tree, SVM, K-NN and MLP used for ASD classification. The final result verify that the MLP and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, May 2022
Impact Factor: 6.252

SVM have obtained the highest accuracy 99.66% with selected feature by BFA method compared to method (Vaishali & Sasikala, 2017).

III. Methodology and Data Set

3.1 Classification Techniques

Classification is a technique to arrange the dataset in a category or class when the pre-information already knows about the consequence such as "disease" and "no disease".

- **Naive Bayes**: Class for a Naive Bayes classifier using estimator classes. Numeric estimator precision values are chosen based on analysis of the training data. For this reason, the classifier is not an Updateable Classifier (which in typical usage are initialized with zero training instances) -- if you need the Updateable Classifier functionality, use the Naïve Bayes Updateable classifier. The Naïve Bayes Updateable classifier will use a default precision of 0.1 for numeric attributes when build Classifier is called with zero training instances.
- **Support Vector Machine:** Support Vector Machines (SVMs) are most frequently used for solving classification problems, which fall under the supervised machine learning category. In classification problem, the radial basis function kernel (RBF kernel), or RBF kernel, is accepted kernel function used in various kernel based learning classifiers. It is commonly used in SVM classification (Shrivas, Sahu, & Hota, 2018).
- **K-Nearest-Neighbour:** Nearest-neighbour classifier. Uses normalized Euclidean distance to find the training instance closest to the given test instance, and predicts the same class as this training instance. If multiple instances have the same (smallest) distance to the test instance, the first one found is used.

3.2 Feature Selection Techniques

Feature selection in data mining also famous as variable selection, attribute selection, or variable subset selection, is the procedure of selecting a subset of significant features (variables, predictors) for use in model creation (Verma, Awasthi, & Sahu, 2021).

- **Chi-Square (CS):** Chi-square test usually used in the selection of features. Chi-Square test have used in statistics to check the self-determination of two measures. The feature selection has the measures are like incident of the feature and occurrence of the category. Chi-Square score measures how many the expected counts E_c and observed Count O_c derivative. Given the data set regarding two measures, be able to get the observed count O and also the expected count E can be obtained. Chi-square feature estimation calculates the significance of a feature via computing the significance of the chi-square statistic with the relevancy of the class (Holmes, Donkin, & Witten, n.d.) (Holmes, Donkin, & Witten, 1994) It is capable of simply scale back the feature of Chi-square technology in feature selection. For example, assume that we have a targeted variable (i.e., class labels) and several other features (feature variables) that express each sample of the data. At this moment, calculate the Chi-square data between each feature variable and the target variable and observe the existence of associations among variables and targets. If the target variable is independent of the variable, then may discard that feature variable. If they are dependent, the feature variable is extremely necessary.
- **Dataset Used:** In this research work have used two kind of chronic disease dataset namely Autistic Spectrum Disorder Screening Data for Children Data Set.
- **Autistic Spectrum Disorder Screening Data for Children Data Set:** The Autistic Spectrum Disorder Screening Data for children data set is collected from UCI repository having autism screening of adults that contained 20 features with one class label and the total number Instances 292.

IV. Experimental Result

4.1 Performances Measurement Techniques

We calculated the performance of classifiers based on the following confusion matrix.

Table 1: Confusion matrix

Actual Vs. Predicted	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, May 2022
Impact Factor: 6.252

The exhibition of the order models was estimated utilizing three execution measures: Accuracy, Sensitivity, and Specificity. All three execution measures are calculated using equation (1), (2) and (3).

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

\

$$\text{Sensitivity}=\frac{TP}{TP+FN} \quad (2)$$

$$\text{Specificity}=\frac{TN}{TN+FP} \quad (3)$$

4.2 Classification Result of ASD Dataset with All Features

Table 2 presents a comparison of results using the ASD dataset. It gives a comparison of the results obtained from the classifiers NB, SVM and K-NN. It is clearly seen from fig. 1 that the highest accuracy of 98.9726 % is obtained by the proposed NB model compared with other models. Sensitivity analysis techniques measure the rate of change at the output of a sample due to adjustments in the input of a variable. The greatest sensitivity of 98.6754% is got by the proposed NB model contrasted and different models. Explicitness investigation is the extent of real negatives that are accurately recognized. The most extreme specificity of 99.2907%is acquired by the proposed NB model contrasted and different models.

Table 2: Shows the performance of Classifiers with WFS

Algorithm	Accuracy	Sensitivity	Specificity
NB	98.9726	98.6754	99.2907
SVM	96.5753	94.7019	98.5815
K-NN	88.6986	85.4304	92.1985

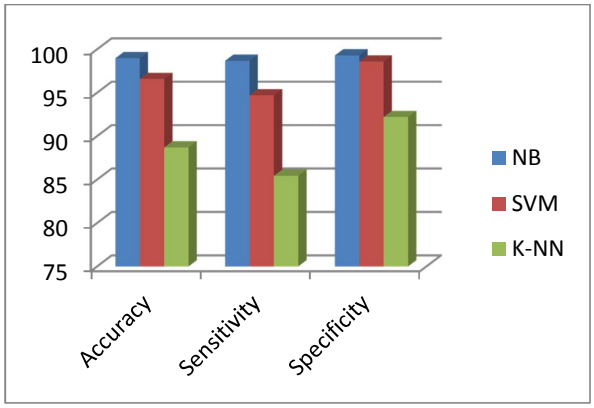

Figure 1: Comparative Analysis of Classification Results

4.3 ASD Dataset Feature Ranked by Chi- Square Methods

Now here we have applied the chi-square as ranking method for selecting relevant feature with the expectation to increase the accuracy while classification.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, May 2022
Impact Factor: 6.252

```
Attribute Evaluator (supervised, Class (nominal): 21 Class/ASD):
       Chi-squared Ranking Filter

Ranked attributes:
292       18 result
 94.3729   4 A4_Score
 74.5011  16 contry_of_res
 69.0155   9 A9_Score
 56.5163  10 A10_Score
 56.118    8 A8_Score
 50.8417   6 A6_Score
 45.6852   3 A3_Score
 45.2217   1 A1_Score
 42.1335   5 A5_Score
 21.914    7 A7_Score
 15.3188   2 A2_Score
 14.0443  13 ethnicity
  4.602   20 relation
  0.6954  15 austim
  0.6509  17 used_app_before
  0.4392  12 gender
  0.1832  14 jundice
  0       11 age
  0       19 age_desc

Selected attributes: 18,4,16,9,10,8,6,3,1,5,7,2,13,20,15,17,12,14,11,19 : 20
```

For the feature selection process we selected the top 15 features. Selected features are again applied the classifiers.

4.4 Classification Result of ASD Dataset with Selected Features

For a comparative study of the classification models, the performances of the proposed classification models are shown in Table 3.

Table 3: Shows the performance of Classifiers with FS

Algorithm	Accuracy	Sensitivity	Specificity
NB	99.3151	98.6754	100.000
SVM	96.9178	95.3642	98.5815
K-NN	94.5205	92.7152	96.4539

Tale 3 shows a comparison of the results using the ASD dataset. It allows comparison of results obtained from NB, SVM and KNN classifiers. It is clearly seen from Fig. 2 that the NB model has the highest accuracy of 99.315% achieved compared to other models. Sensitivity analysis techniques measure the rate of change in the output of a sample due to adjustments in the input of a variable. The highest sensitivity of 98.675% is achieved by the NB model providing contrast and different models. The explanatory investigation is the extent to which the negation is actually correctly perceived. The highest specificity of 100% was obtained by the NB model provided by contrast and differential models.

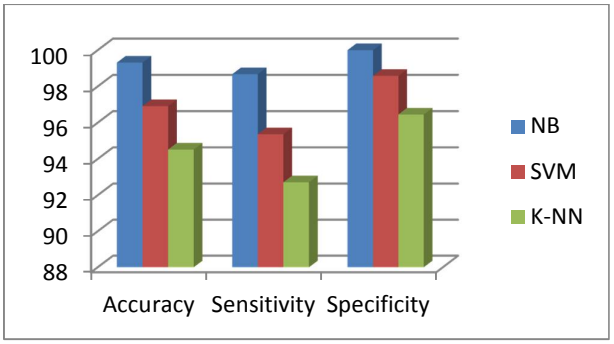

Figure 2: Comparative Analysis of Classification Results with Feature Selection

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, May 2022
Impact Factor: 6.252

V. Conclusion

One of the key issues in ASD screening research is improving the screening process so that individuals and their families can benefit from faster and more accurate service. This can be done using automated data mining based methods to generate accurate classification systems from historical controls and cases. This study applies the ASD dataset to three classification methods NB KNN and SVM. We also use FST as CS to rank the characteristics of the ASD dataset. The performance of all algorithms was compared in terms of accuracy sensitivity and specificity. The results demonstrate that in all cases in the ASD dataset the proposed classifiers with 15

References

[1]. Akyol, K., Gultepe, Y., & Karaci, A. (2018). "A Study on Autistic Spectrum Disorder for Children Based on Feature Selection and Fuzzy". International Congress on Engieneering and Life Science, pp. 804–807.

[2]. Diabat, M. Al, & Al-shanableh, N. (2019). "E NSEMBLE L EARNING M ODEL FOR S CREENING". International Journal of Computer Science & Information Technology (IJCSIT), vol.1, no. 2, pp.13–14. https://doi.org/10.5121/ijcsit.2019.11205

[3]. Holmes, G., Donkin, A., & Witten, I. H. (n.d.). "WEKA: a machine learning workbench. Proceedings of ANZIIS". '94 - Australian New Zealnd Intelligent Information Systems Conference, pp. 357–361. https://doi.org/10.1109/ANZIIS.1994.396988

[4]. Shrivas, A. K., Sahu, S. K., & Hota, H. S. (2018). "Classification of Chronic Kidney Disease with Proposed Union Based Feature Selection Technique". SSRN, vol. 5, pp. 649–653. https://doi.org/10.2139/ssrn.3168581

[5]. Thabtah, F., & Peebles, D. (2020). "A new machine learning model based on induction of rules for autism detection". Health Informatics Journal, vol. 26, pp. 1, 265–286. https://doi.org/10.1177/1460458218824711

[6]. Vaishali, R., & Sasikala, R. (2017). "A machine learning based approach to classify Autism with optimum behaviour sets". International Journal of Engineering & Technology, vol. 5, pp. 1–6.

[7]. Verma, P., Awasthi, V. K., & Sahu, S. K. (2021). "An Ensemble Model With Genetic Algorithm for Classification of Coronary Artery Disease." International Journal of Computer Vision and Image Processing, vol.11, no. 3, pp.70–83. https://doi.org/10.4018/ijcvip.2021070105

[8]. Wang, H., Li, L., Chi, L., & Zhao, Z. (2019). "Autism Screening Using Deep Embedding Representation". In Lecture Notes in Computer Science. Springer, Cham.