# Prediction of Attention Deficit Hyperactivity Disorder (ADHD) in Regional Locations Using Novel Support Vector Machine Algorithm Over the Naive Bayes Algorithm for Better Accuracy

1st NV Midhun Sai
*Department of Computer Science and Engineering,*
*Saveetha School of Engineering,*
*SIMATS, Chennai 602105.*
Tamilnadu. India.
midhunsainv18@saveetha.com

2nd Saravanan. M.S
*Corresponding Author, Department of Computer Science and Engineering,*
*Saveetha School of Engineering,*
*SIMATS, Chennai 602105.*
Tamilnadu. India.
saravananms@saveetha.com

3rd P. Subramanian
*Department of Computer Science and Engineering,*
*Saveetha School of Engineering,*
*SIMATS, Chennai 602105.*
Tamilnadu. India.
subramanianp.sse@saveetha.com

2022 International Conference on Knowledge Engineering and Communication Systems (ICKECS) | 978-1-6654-5637-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICKECS56523.2022.10059721

***Abstract*—The study's focus is to enhance the precision and accuracy of novel ways for anticipating attention-deficit/hyperactivity disorder at local sites utilizing the machine learning Support Vector Machine algorithm. Materials and Methods: The Novel Support Vector Machine classifier is applied to an ADHD dataset containing 1395 records, 119 characteristics, 104 samples A framework comparing SVM and Naive Bayes classifiers was presented and created for the purpose of predicting the regional distribution of ADHD in the medical area. The classifiers were evaluated and their precision and accuracy were noted. Results: On the dataset provided, the Naive Bayes classifier predicts the regional location of ADHD disorder with an accuracy of 89.12%, while the Support Vector Machine classifier predicts the same with a 96.75% accuracy. There is a statistically significant difference in the groupings between these two. has a 95% confidence interval of (p=0.010; p0.05). Conclusion: In comparison to Naive Bayes, Novel Support Vector Machine appears to perform better and produce more accurate and reasonable results.**

**Keywords—Attention Deficit Hyperactivity Disorder, Naive Bayes Algorithm, Novel Support Vector Machine, Regional location, Statistical analysis, Machine learning.**

## I. Introduction

The novel developmental disorder ADHD is characterised by inadequate performance independent of satisfactory academic performance and appropriate educational openness. ADHD affects regions of the brain. A typical functioning of the left-sided neural network in reading, which comprises the frontal, parietal, and temporal lobe regions, is what distinguishes ADHD in youngsters (Biederman 2020). This study will be possible to forecast patients' chances of quick disease recovery at an earlier stage. But it's critical to compare and analyse a number of categorization techniques that provide more accuracy. To compare the accuracy of SVM and NB algorithms in locating the regional location of ADHD disorder in daily life is the goal of this project [1], ADHD makes it easier to use helpful mobile applications to analyze potential fixes and present a working prototype that can be used to give the technician input on ADHD [2].

There are about 124 research papers published in IEEE and Google Scholar. Most of the results of the proposed research are an optimal study of the regional location in ADHD disease, primarily impacts those who are younger than the older ones. The phonological, orthographic, and semantic processes are especially linked to regional hypo activation in these readers. Increased subcortical - cortical connectivity between regions that exhibit regional reading deficits [3]. The convolution of the auditory cortex was more variable in children. Combining neural and behavioural data reliably predicted novel ADHD before school [4]. The neurobiology of ADHD provides a novel and heuristic theoretical framework for a new understanding of the disability (Kuhl et al. 2020; Kershner 2019). Novel ADHD children showed left hemispheric differences in cortical Gyrification. Gyrification differences correlated with neurite orientation dispersion and density imaging (NODDI) derived from neurite organisation [5]. Reducing neurophysiologic adaptation in adults and Children's ADHD (attention-deficit/hyperactivity disorder). Language manipulation of a constant voice was significantly reduced, which is the most cited article in the literature selected for this study of the research [6].

Our team's diversity of experience and research skills have produced articles of the highest calibre [7–15]. In this existing system, all datasets have fewer values for attributes and data are lower. Finding the algorithm with the best accuracy is difficult.

The proposed system is to find out in which region ADHD is most common and what are the causes of this disease. The SVM and Naive Bayes algorithms are employed. Naive Bayes methods outperform other classification algorithms in terms of performance. Therefore, the proposed Novel Support Vector Machine) technique uses statistical analysis of machine learning algorithms to increase accuracy.

## II. Materials and Methods

SIMATS School of Engineering given opportunity to conduct this research work at AI Laboratory. The SVM and NB algorithm are the two groups used in this investigation. Using power analysis was measured for each group. The g-power tool was used to predict the sample size and the results showed that the sample size is 104, with an 86% g-power value.

The dataset used in this study includes 1395 rows and 119 attributes. A download of it was made from the Kaggle website. Age, naive and other languages, ADHD, and gender were a few of the characteristics. To find the method with the highest accuracy for the sex ratio in ADHD, two groups of algorithms were implemented using the provided information. The ratio of training to testing datasets is typically 80:20. The

978-1-6654-5637-1/22/$31.00 ©2022 IEEE

accuracy of the findings is anticipated using the training data, which are used to build the data model using the SVM algorithm.

### *A. Support Vector Machine (SVM) Algorithm*

Novel SVM algorithm act as a separative hyperplane used to formally develop the discriminative classifier known as SVM as in fig 1. It is a representation of events as mapped points in space with as much space as feasible separating the points of various categories. Because of the incredibly effective results achieved by SVM employing the supplied below, it is currently widely used.

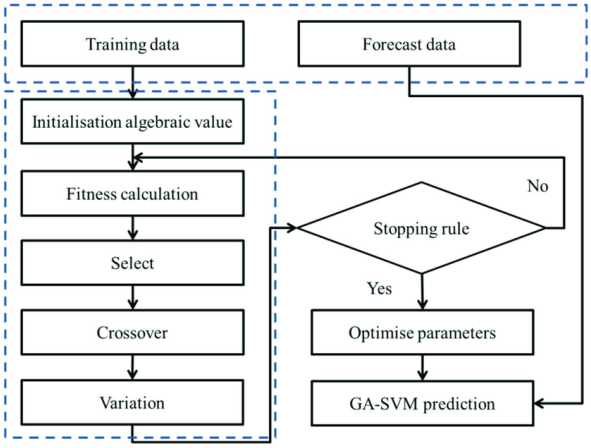


Fig. 1. Flow Chart of SVM Algorithm

### *B. Naive Bayes Algorithm Pseudocode:*

The NB algorithm is a method for supervised machine learning. To function, this method depends on probability. The Naive Bayes method has undergone a number of revisions. The Bernouli Naive Bayes and Multinomial Naive Bayes are what they are and Gaussian Naive Bayes algorithms. A Guassian Naive Bayes algorithm model was selected among other models. It is appropriate for typical classification issues. To forecast the outcome as shown below, naive bayes ignore the superfluous aspects of the datasets that are provided.

### *C. Statistical Analysis*

The dependent variable (Adhesive Deficit Hyperactivity Disorder) employed in the analysis is accuracy in locating the regional location. Age, gender, native language, and foreign language are the independent variables. The accuracy results from the two groups are then processed. On the suggested model as in fig 2, Anaconda tests have been run. Spyder©. The used computer features a 256 HDD, 8GB of RAM, and a Windows 10 CPU.

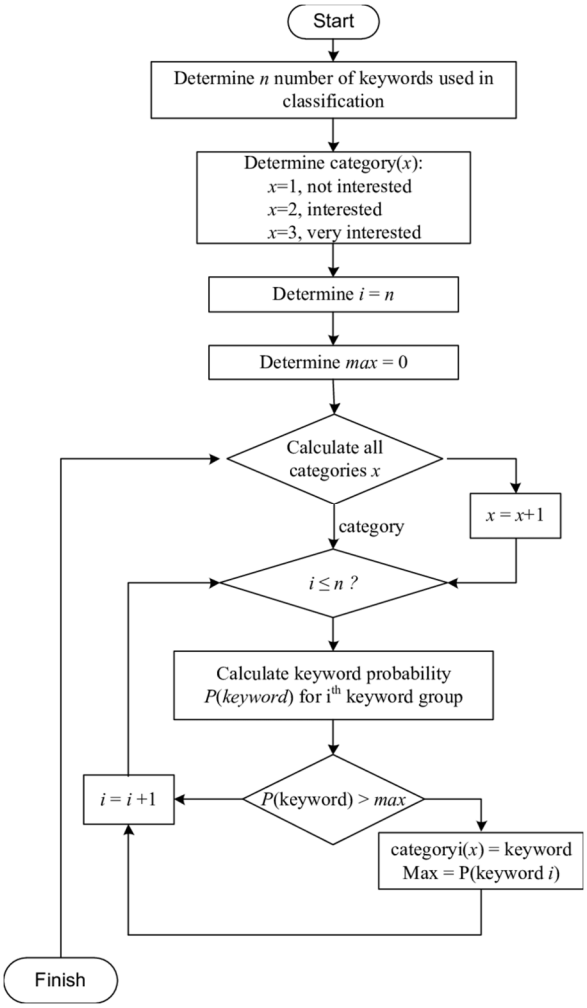


Fig. 2. Flow Chart of Analysis

## III. RESULTS

Data normalization comes after data pre-processing, allowing the data to be scaled. The next step is to clean up the data. Unwanted data is eliminated throughout this process. After that, the data are split into training and test sets as seen in Table 1. There is a statistically significant difference between the two groups with a 95% confidence interval ($p=0.010$; $p<0.05$). It was found that the detection accuracy of SVM for the attention deficit hyperactivity disorder dataset was significantly better than that of NB algorithm as shown in Table 1. The SVM method has the best accuracy of 96.75 percent, while the other has the lowest accuracy of 89.12 percent in determining the regional location.

The accuracy of the SVM and Naive Bayes algorithms are statistically analyzed in Table 2 utilizing the average, both the standard and standard error and deviation. The accuracy rates differ statistically significantly having difference. The average value of the SVM method is 89.89%, compared to the NB algorithm's mean value of 83.54%.

TABLE I. SVM ACHIEVED THE MAXIMUM ACCURACY OF 96.75 PERCENT WHEN COMPARED TO NAIVE BAYES, WHICH HAD THE HIGHEST ACCURACY OF 87.12 PERCENT.

| Trail No | Dataset size | Training Data % | Testing Data % | Naive bayes | SVM |
|---|---|---|---|---|---|
| 1 | 1200 | 70 | 30 | 79.13 | 84.21 |
| 2 | 1250 | 70 | 30 | 80.02 | 85.02 |
| 3 | 1300 | 70 | 30 | 81.14 | 86.04 |
| 4 | 1350 | 70 | 30 | 82.05 | 87.12 |
| 5 | 1480 | 70 | 30 | 83.56 | 88.23 |
| 6 | 1500 | 70 | 30 | 84.45 | 89.15 |
| 7 | 1530 | 70 | 30 | 84.56 | 92.67 |
| 8 | 1580 | 70 | 30 | 85.13 | 94.15 |
| 9 | 1650 | 70 | 30 | 86.24 | 95.56 |
| 10 | 1680 | 70 | 30 | 89.12 | 96.75 |

TABLE II. ANALYSIS OF NAIVE BAYES AND SVM'S IS MORE ACCURATE (89.89 PERCENT ). FURTHERMORE, NAIVE BAYES IS LESS ACCURATE (83.54 PERCENT).

| Algorithm | Trail No. | Mean | Std Deviation | Std. Error Mean |
|---|---|---|---|---|
| SVM | 10 | 89.89 | 1.632 | .8165 |
| NB | 10 | 83.54 | 1.632 | .8165 |

With a value of 0.05, Table 3 compares the significance level for the SVM and Naive Bayes methods. There is a statistically significant difference between the two groups with a 95% confidence interval (p=0.010; p<0.05). This is demonstrated by contrasting the SVM and Naive Bayes accuracies over the four trails in Fig. 3.

TABLE III. THERE IS A STATISTICALLY SIGNIFICANT DIFFERENCE (P=0.010; P<0.05) WITH A 95% CONFIDENCE INTERVAL BETWEEN THE TWO GROUPS WHEN USING THE P=0.010 SIGNIFICANCE CRITTER FOR THE NAIVE BAYES AND SVM ALGORITHMS.

| | F | sig. | t | df | sig.(2-trail) | Mean difference | Std. error difference | 95% confidence lower | 95% confidence upper |
|---|---|---|---|---|---|---|---|---|---|
| Equal variances assumed | 0 | 0.010 | 4.000 | 6 | .007 | -8.00 | 1.000 | 1.55 | 6.44 |
| Equal variances not assumed | 0 | 0.010 | -4.000 | 6 | .007 | -89.00 | 1.000 | -6.44 | -1.55 |

The Y-trails and X-axis axes are used to plot the accuracies. Neither a continuous can be seen in the line plots. This figure's line graphs each have a distinct trend. The size of the data has no correlation between the number of groups or any other factor and accuracy.

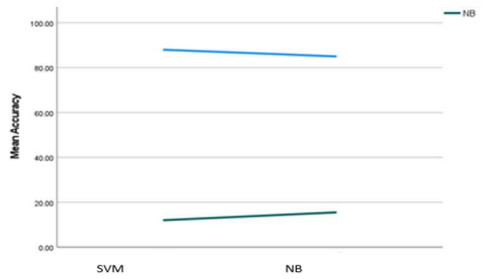


Fig. 3. Comparison of the trail accuracy of the NB (blue) and SVM (black) (X-axis :Group, and Y-axis :Mean)

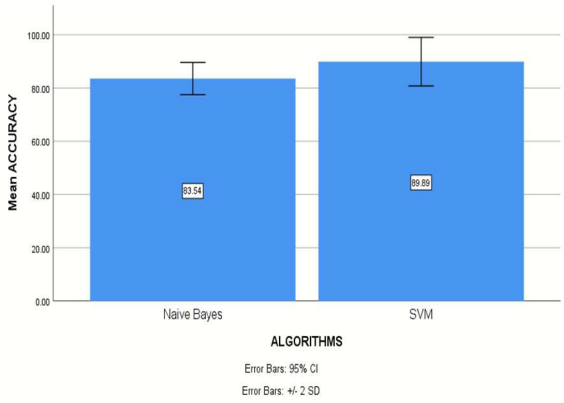


Fig. 4. Comparison of the NB and SVM algorithms' mean accuracy. The NB algorithm has a higher mean accuracy than the SVM. Boosting vs. SVM method Mean Regional Location Accuracy + 1 SD on the X-axis.

The accuracy of a model of the SVM method is 89.89%, while that of a model of the Naive Bayes algorithm is 83.54%, as shown in Fig. 4. Comparing the suggested SVM method to other algorithms, the Naive Bayes algorithm performed better. When four samples were subjected to statistical analysis. The SVM method produced 1.257 standard deviations and a SE of 0.912, but the NB approach produced identical results for the table's standard deviation and variance 2. A number less than 0.05 is significant and suggests that our hypothesis is correct. This also holds true both the dependent variable's output values and the independent variable's input values, to which they are related, for changes.

When the two algorithms' accuracy was tested using an independent t-test and found $p < 0.05$. The SVM model had a 96.75% accuracy rate. The accuracy of the NB algorithm is 89.12%, whereas the accuracy is 80%, while compared to the other methods.

## IV. DISCUSSION

The accuracy of the suggested system only slightly increased by 8%. In addition, it is shown that the SVM algorithm in classifying the gender ratio achieves better specificity. Between the two groups, also 95% confidence interval of (p=0.010), which is p <0.05. The SVM algorithm's accuracy has increased as a result of more accurate regional location classification. It produces an Area Under Curve (AUC) of 0.82, indicating that the method performed better on the gathered data. Although it may improve accuracy, the classification in the dataset is not examined in this study. The disease of ADHD primarily affects younger people than older people, according to an ideal assessment of geographical location.

There are similar finding presence over the problem identified and Most of the applications are suitable for children's brain care. We require proof of knowledge for diagnosing ADHD at a previous time in order to conduct advanced research on ADHD seizures [16,17]. Males have

less reading proficiency than females [18]. The accuracy of diagnosing ADHD using computer systems is 81.06 percent when taking into account people's issues with reading, writing, and speaking. We can forecast more accurately thanks to software algorithms.

This research is more effective than the other research. Particularly Regional hypoactivation in ADHD is linked with processing of phonological, orthographic, and semantic information. Increased subcortical - regional reading deficiencies shown by cortical connectivity across regions [19]. So that if any problem occurs they will rectify at the starting stage itself [20]. Therefore, the cost reduction and the early detection is achieved. In the opposing findings they haven't performed early detection which leads to the effect of brain seizure. ADHD still has some errors in finding the regional location. In this paper, we reinforced regional location prediction to save time, cost, and the lives of children.

The Limitations and the future of an algorithm may be improved to forecast the factors affected in ADHD. At which age group and the symptoms of the disease of people the ADHD will occur. This model performs different computational tasks faster than the traditional systems. In this study, normal standardiser is used to train and evaluate the models using datasets. The same method can be used with other algorithms, but this one is more effective. The number of trained samples directly relates to the model's accuracy, therefore if there are more samples, we will get the required result.

## V. Conclusion

Using the Naive Bayes procedure, there is a statistically significant difference between the two groups. the SVM algorithm obviously outperforms the NB algorithm in this regard within the confines of this study. With an accuracy of 96.75 percent, the SVM algorithm has the highest precision has a 95% confidence interval of ($p=0.010$; $p<0.05$). The packages such as Keras are used for eliminating attributes from the dataset and the machine learning algorithms are employed are the elements that contributed to this boost in accuracy. As a classifier, the SVM algorithm provides decent accuracy when compared to the Naive Bayes algorithm.

## References


[1] Charach A, Harrison J. Empowering The Parents Of Preschoolers: Approaches To Promoting Positive Parenting. Journal of the American Academy of Child & Adolescent Psychiatry. 2020. p. S106. doi:10.1016/j.jaac.2020.07.431

[2] Ko M. 6.25 Medication Adherence Monitoring in Pediatric and Young Adult Populations on Attention-Deficit/Hyperactivity Disorder (ADHD) Stimulant Medications. Journal of the American Academy of Child & Adolescent Psychiatry. 2017. pp. S284–S285. doi:10.1016/j.jaac.2017.09.370

[3] Ghanizadeh A, Jafari P. Risk factors of abuse of parents by their ADHD children. European Child & Adolescent Psychiatry. 2010. pp. 75–81. doi:10.1007/s00787-009-0067-y

[4] Charach A, Yeung E, Volpe T, Goodale T, dosReis S. Exploring stimulant treatment in ADHD: narratives of young adolescents and their parents. BMC Psychiatry. 2014. doi:10.1186/1471-244x-14-110

[5] Anderberg E. Persistent reading difficulty despite improved focus on a stimulant medication. Child and Adolescent Psychiatry. 2021. pp. 11–16. doi:10.1093/med/9780197577479.003.0002

[6] T. Hassan, M. U. Akram, M. F. Masood, and U. Yasin, "Deep structure tensor graph search framework for automated extraction and characterization of retinal layers and fluid pathology in retinal SD-OCT scans," Comput. Biol. Med., vol. 105, no. February 2019, pp. 112–124, 2019.

[7] Valsalan P, Sriramakrishnan P, Sridhar S, Latha GCP, Priya A, Ramkumar S, et al. Knowledge based fuzzy c-means method for rapid brain tissues segmentation of magnetic resonance imaging scans with CUDA enabled GPU machine. J Ambient Intell Humaniz Comput. 2020. doi:10.1007/s12652-020-02132-6

[8] Dhaya chandhran KS, Jothilakshmi M, Mohan A, Tholkapiyan M. Performance evaluation and R-value analysis for thermally insulated PCM wall with fluted sheets. Materials Today: Proceedings. 2020;22: 912–918.

[9] Saravanan A, Kumar PS, Yaashikaa PR, Karishma S, Jeevanantham S, Swetha S. Mixed biosorbent of agro waste and bacterial biomass for the separation of Pb(II) ions from water system. Chemosphere. 2021;277: 130236.

[10] Indiramma P, Sudharani C, Needhidasan S. Utilization of fly ash and lime to stabilize the expansive soil and to sustain pollution free environment – An experimental study. Materials Today: Proceedings. 2020;22: 694–700.

[11] Ezhilarasan D. Dapsone-induced hepatic complications: it's time to think beyond methemoglobinemia. Drug Chem Toxicol. 2021;44: 330–333.

[12] Chupradit S, Jalil AT, Enina Y, Neganov DA, Alhassan MS, Aravindhan S, et al. Use of Organic and Copper-Based Nanoparticles on the Turbulator Installment in a Shell Tube Heat Exchanger: A CFD-Based Simulation Approach by Using Nanofluids. J Nanomater. 2021;2021. doi:10.1155/2021/3250058

[13] Venu H, Appavu P. Combustion and emission characteristics of tamarind seed biodiesel-diesel blends in a compression ignition engine. Int J Ambient Energy. 2021;42: 1441–1446.

[14] Prabhu A, Venkata Ramanan M, Jayaprabakar J, Harish V. Experimental investigation of emission characteristics on ricebran biodiesel–alcohol blends in a diesel engine. Int J Ambient Energy. 2021;42: 1123–1128.

[15] Raja G, Chandramohan D, Gnanavel BK, Sathish T. Effect of inter-facial coupled contact forces in the multilayered pacemaker lead cable. Proceedings of International Conference On Recent Trends In Mechanical And Materials Engineering: ICRTMME 2019. AIP Publishing; 2020. doi:10.1063/5.0024971

[16] Bosch-Bayard J, Girini K, Biscay RJ, Valdes-Sosa P, Evans AC, Chiarenza GA. Resting EEG effective connectivity at the sources in developmental dysphonetic dyslexia. Differences with non-specific reading delay. International Journal of Psychophysiology. 2020. pp. 135–147. doi:10.1016/j.ijpsycho.2020.04.021

[17] Rello L, Baeza-Yates R, Ali A, Bigham JP, Serra M. Predicting risk of dyslexia with an online gamified test. PLoS One. 2020;15: e0241687.

[18] Yang, Yang Y, Tam F, Graham SJ, Sun G, Li J, et al. Men and women differ in the neural basis of handwriting. Human Brain Mapping. 2020. pp. 2642–2655. doi:10.1002/hbm.24968

[19] Paz-Alonso PM, Oliver M, Lerma-Usabiaga G, Caballero-Gaudes C, Quiñones I, Suárez-Coalla P, et al. Neural correlates of phonological, orthographic and semantic reading processing in dyslexia. Neuroimage Clin. 2018;20: 433–447.

[20] Charach A. Editorial: Time for a New Conversation on Stimulant Use. Journal of the American Academy of Child & Adolescent Psychiatry. 2020. pp. 929–930. doi:10.1016/j.jaac.2019.10.004