



## Development and Diagnosis of Autism Spectrum Disorder (ASD) Screening: Leveraging Machine Learning Techniques

Viplav Soliv<sup>1</sup>, Dr. Raghavendra Patidar<sup>2</sup>

<sup>1</sup> Electronics and Telecommunication, Rajiv Gandhi Institute of Technology, Andheri (w),  
Mumbai, India

<sup>2</sup> Electrical Engineering, Global Institute of Technology, Jaipur, India

Email: <sup>1</sup> [viplav.soliv@mctrigit.ac.in](mailto:viplav.soliv@mctrigit.ac.in), <sup>2</sup> [raghvendra.patidar@gitjaipur.com](mailto:raghvendra.patidar@gitjaipur.com)

---

### Abstract

The social and relational abilities of an individual with autism spectrum disorder (ASD) are impeded until the end of their lives. In this paper, we endeavor to explore the practicality of utilizing Naive Bayes, Backing Vector Machine, Strategic Relapse, KNN, Brain Organization, and Convolutional Brain Organization for foreseeing and dissecting ASD. There are 292 cases and 21 qualities in the first and the second dataset for screening grown-ups for autism spectrum disorder, and it comprises of 704 cases with 21 qualities. The juvenile autism spectrum disorder (ASD) screening dataset has 21 qualities and 104 cases. Information for grown-ups, youngsters, and teenagers were evaluated for autism utilizing an assortment of machine learning strategies, and the outcomes emphatically recommend that CNN-based expectation models work best. Their precision was 99.53%, 98.30%, and 96.88%, separately. Index Terms—Autism Spectrum Disorder (ASD), Computer-Aided Diagnosis, Autism Screening. Machine Learning, Screening.

---

### 1. Introduction

Autism spectrum infection (ASD) is a developing scourge that influences individuals, everything being equal. The subject's personal and actual prosperity might be impressively aided by early diagnosis of this neurological condition. Early distinguishing proof of human sicknesses in view of a great many wellbeing and physiological boundaries presently seems doable, because of the rising utilization of machine learning-based models for such expectations. There are various different psychological instabilities whose somewhat couple of side effects are strikingly like those of ASD, detecting the condition might challenge. Formative issues in the mind are at the base of autism spectrum disorder. An individual with Autism Spectrum Disorder frequently misses the mark on interactive abilities and correspondence capacities important to take part in significant associations with others. This frequently lastingly affects an individual's life. It is fascinating to think about how conceivable it is that this ailment has both ecological and genetic causes.

This condition frequently first appears around the age of three and may last an individual's entire life. Patients with this condition can't be relieved, but early location might assist with postponing the sickness' movement for some time. Low birth weight, having kin with autism spectrum disorder, having old guardians, and so on is all chance variables for the improvement of the disorder. In any case, there are issues with social contact and

correspondence, for example: Unsuitable chuckling and giggling, Absence of pain perception, Poor ability to maintain eye contact, Inadequate reaction to auditory cues, it's possible they don't want to be cuddled, Unable to show emotions via body language Isolation from other people, adhering unsuitable items, want to be on your own, Echoing phrases, words, etc.

Individuals on the autism spectrum often struggle with narrowed interests and rigid routines. Here are some concrete instances of each kind of conduct.

Engaging in a pattern of conduct, such as saying the same thing again and over.

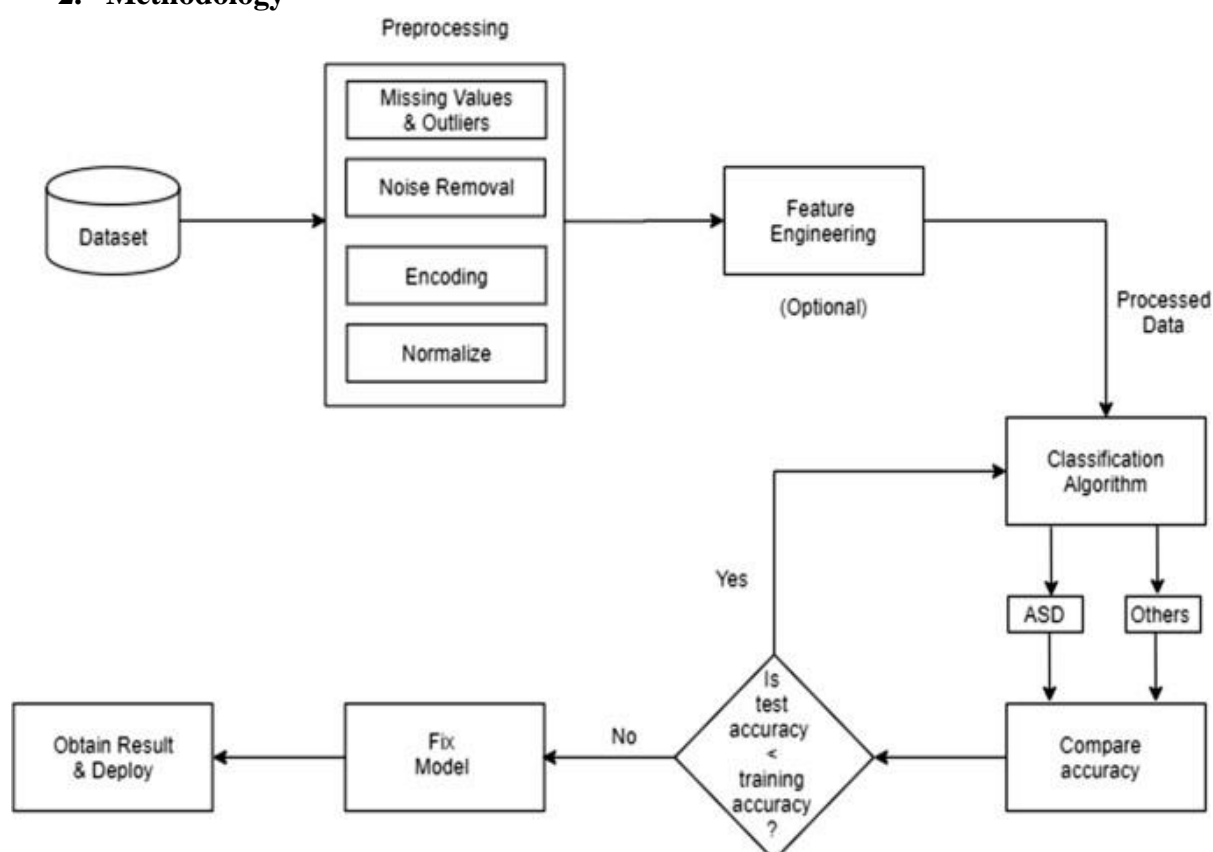
When a modification is made to the Person's usual routine, they get agitated.

Taking a passing fascination with particulars of the subject, such as figures, facts, etc.

Somewhat less reactive to external stimuli (light, sound, etc.).

The symptoms of autism spectrum disorder (ASD) may be mitigated and the quality of life for those who suffer from ASD can be greatly enhanced with early identification and treatment. However, no medical process or test exists to identify autistic individuals. Signs of autism spectrum disorder are often detectable via casual observation. Typically, parents and teachers are the ones who notice the signs of autism spectrum disorder in school-aged children and teenagers. The school's special education staff then conducts an assessment of the student's ASD symptoms.

## 2. Methodology



**Fig 1:** Methods for Identifying Autism Spectrum Disorder

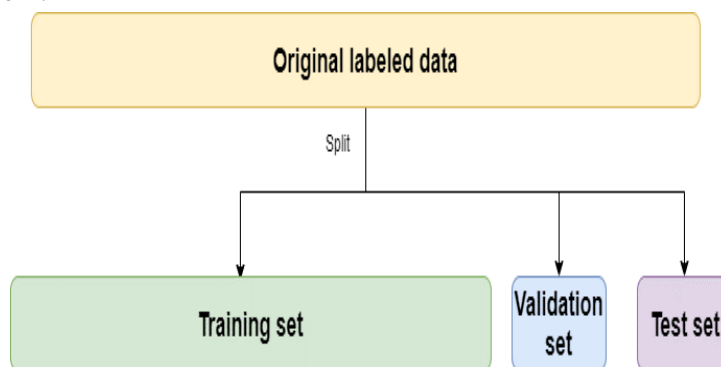
Pre-handling of information, preparing and testing with determined models, appraisal of results, and ASD expectation are completely displayed in Figure 1 as a feature of the proposed work process. Python is utilized for this venture's backend.

### A. Data Pre-Processing

To figure out a lot of crude information, a strategy called “information pre-handling” is utilized. Because of many slip-ups and missing numbers, true information is frequently untrustworthy and capricious. Great results are generally the result of well-preprocessed information. To manage missing qualities, exception recognizable proof, information discretization, information decrease (aspect and numerosity decrease), and so on, different information pre-handling strategies are utilized. The attribution approach was utilized to settle the issue of missing qualities in these datasets.

### B. Training and Testing Model

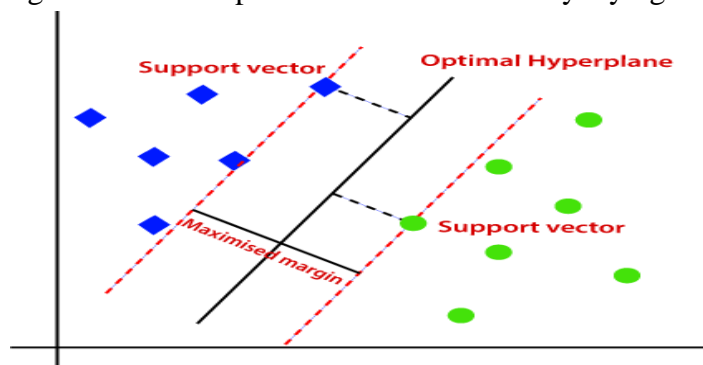
The entire dataset has been isolated into equal parts, with 80% devoted to the preparation stage and 20% to the testing stage. Once more, preparing information has been isolated in half for cross- approval. The datasets are parted 80:20 with the preparation dataset being the bigger of the two. Grouping results for the last preparation, testing, and approval sets are displayed in Figure 2.



**Fig. 2:** Complete Data Sets for Training, Testing, and Validation

#### i. Support Vector Machine (SVM)

Support vector machines (SVMs) are a sort of direct regulated machine learning utilized in order and relapse. The framework is an example acknowledgment instrument. It doesn't prompt the overfitting issue. SVM separates between classes by laying out a limit.



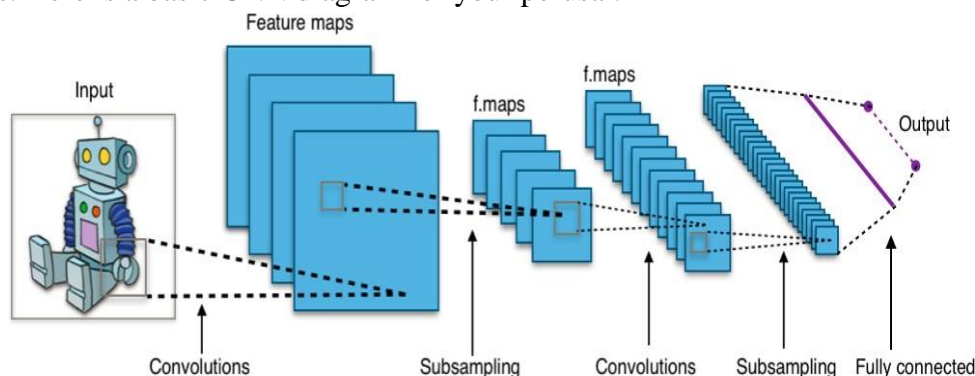
**Fig. 3:** An SVM classifier

#### ii. Naïve Bayes (NB)

One sort of managed learning calculation is the credulous Bayes classifier. This model creates new data and depends on the hypothesis of joint likelihood. Gullible Bayes is a deduction strategy that underestimates freedom. When contrasted with the SVM and ME displays, it calls for considerably less investment to prepare. It takes in the earlier likelihood and the probability for a dataset and yields the back likelihood.

### iii. Convolutional Neural Network (CNN)

As mentioned in CNN is a deep learning approach that may be used to construct models for a wide range of situations. It is an analogous feed-forward neural network to the human brain. In addition to the input and output layers, a typical CNN model also includes many additional types of layers, such as convolution, max pooling, fully connected, and normalization layers. Matrix multiplication followed by a bias offset may be used to calculate their activation functions. Here is a basic CNN diagram for your perusal:



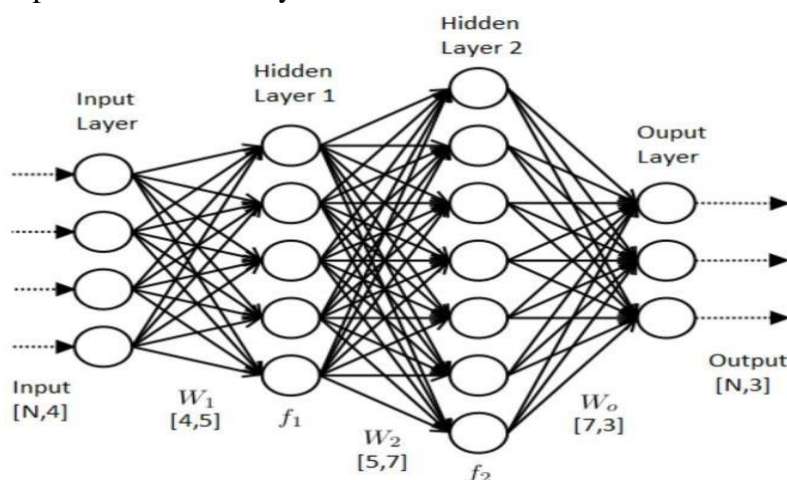
**Fig. 4:** Modeling a CNN's Primitive Components

### iv. K- Nearest Neighbor (KNN)

KNN is the easiest directed learning strategy. It very well might be utilized to grouping and relapse issues. Accepts encompassing information is equivalent. The 'K' part determines the expected number of seeds focuses. Choosing it appropriately will assist with limiting mix-ups. Accordingly, it is predicated on the idea of nearness, which might allude to either physical or applied closeness. Euclidean distance is the most frequently utilized distance metric.

### v. Artificial Neural Network

An ANN is an organization of neurons that is associated here and there. A bunch of information values and loads are doled out to every neuron cell. Forward brain networks are the most well- known sort of fake brain organization. There is just a single way for information to go across this organization, and it's forward. The information layer, the secret layer, and the result layer are the three essential parts of a multi-facet organization. The organization incorporates no sort of cycle or circle.



**Fig. 5:** Artificial Neural Network

### vi. Logistic Regression (LR)

LR is a relapse strategy for concentrating on subordinate factors that take only two potential qualities. It can give paired results, either 0 or 1. For the dataset with consistent qualities, it is used. It depicts the association between one paired subordinate variable and one ostensible or normal free factor. The sigmoidal capability is a decent image for it.

### 3. Result and Discussion

The explicitness, responsiveness, and exactness of the outcome are assessed utilizing the disarray grid and arrangement report. How well the model is prepared will decide the result.

#### *Performance Evaluation metrics*

The viability of a grouping model in gathering a goal might be assessed by following its exhibition measurements. Measurements for checking execution are utilized to the test dataset to evaluate how well the characterization model performed. Measurements like disarray frameworks, accuracy, review, and bogus positive rates ought to be in every way thought about while evaluating a model's viability. The presentation pointers are determined utilizing the accompanying formulae:

**Table 1:** Components of an Ambiguity Matrix

Predictive ASD values		
Actual ASD values	True positive (TP)	False Positive (FP)
	False Negative (FN)	True Negative (TN)

Most prediction models use a four-class distribution for their data:

**True positive (TP):** We were right in our diagnosis of autism spectrum disorder for this person.

**True negative (TN):** We were right in our assessment; this person does not have autism spectrum disorder.

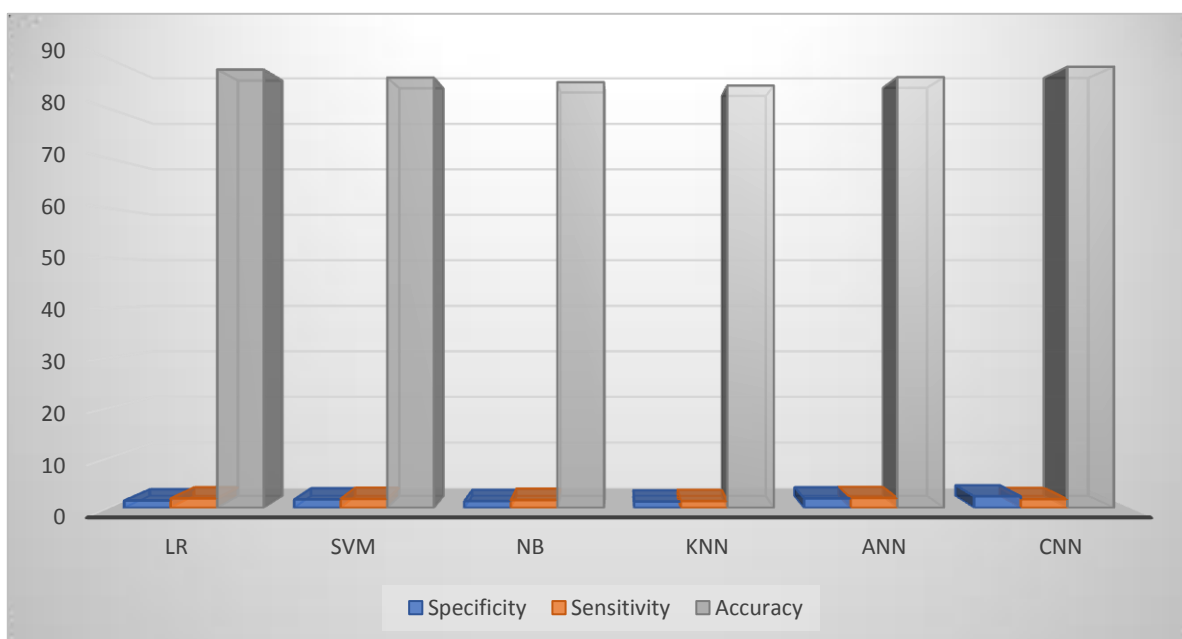
**False positive (FP):** We made an inaccurate diagnosis of autism spectrum disorder for this person. Type 1 mistake describes this situation.

**False negative (FN):** We were wrong in our assumption that this person did not have ASD; instead, they do. A Type 2 error has occurred.

For ASD screening information spreading over grown-ups, youngsters, and teenagers, exploratory discoveries of an assortment of machine learning strategies utilizing all elements determination have been introduced. Here, each of the 21 qualities are utilized to decide the prescient model's exactness, responsiveness, and particularity. The Gullible Inclination technique has been executed utilizing Gaussian NB. The RBF Bit with a gamma worth of 0.1 was utilized in the SVM examination. The norm for KNN is  $N=5$ . Adam Enhancer was utilized in ANN with a 0.01% learning rate and 100 cycles. The two-fold cross-entropy misfortune capability, 16 and 32 channels, and 0.5 dropouts more than 150 cycles. Definite beneath are the general presentation measurements for each machine learning classifier across every one of the three datasets:

**Table 2:** Data from Adult Screenings for Autism Spectrum Disorder

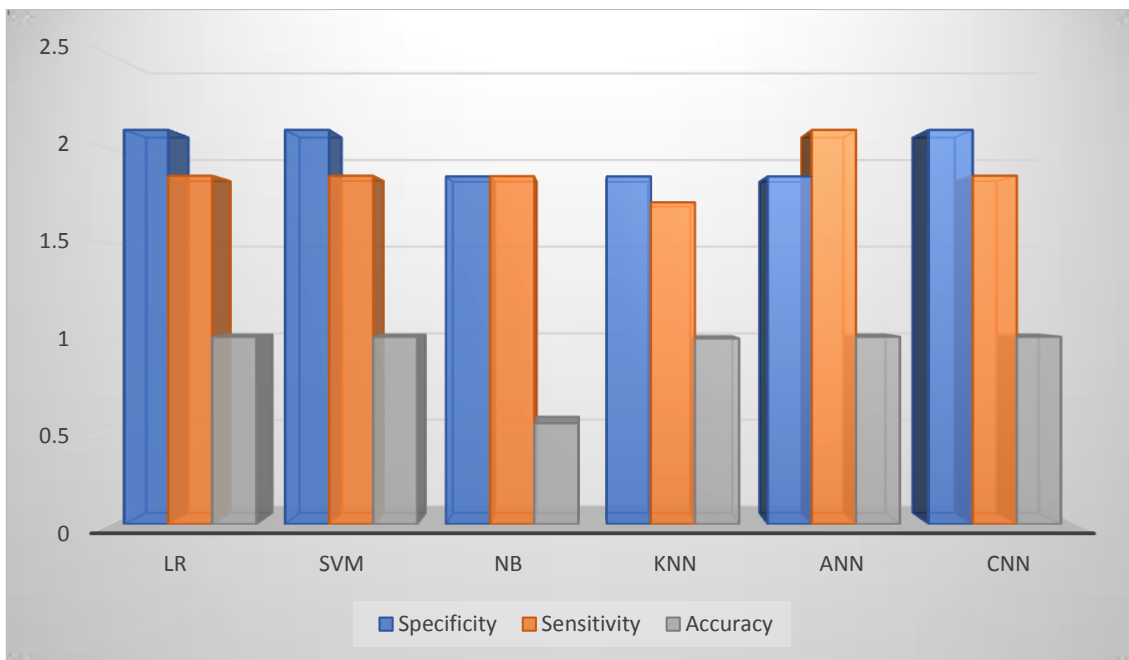
Classifier	Specificity	Sensitivity	Accuracy
LR	1.4578	1.8585	87.85
SVM	1.6583	1.7777	86.23
NB	1.3245	1.5545	85.33
KNN	1.2369	1.3259	84.64
ANN	1.8745	1.9476	86.32
CNN	2.3	1.7349	88.42



Machine learning models on a dataset of pediatric ASD analyze, specialists found that the first dataset's precision went from 99.24% to 100.20%. With K=5, the K-NN classifier accomplished a most minimal exactness of 99.24%. CNN, SVM, ANN, and LR all accomplished an expectation exactness of 100.20%.

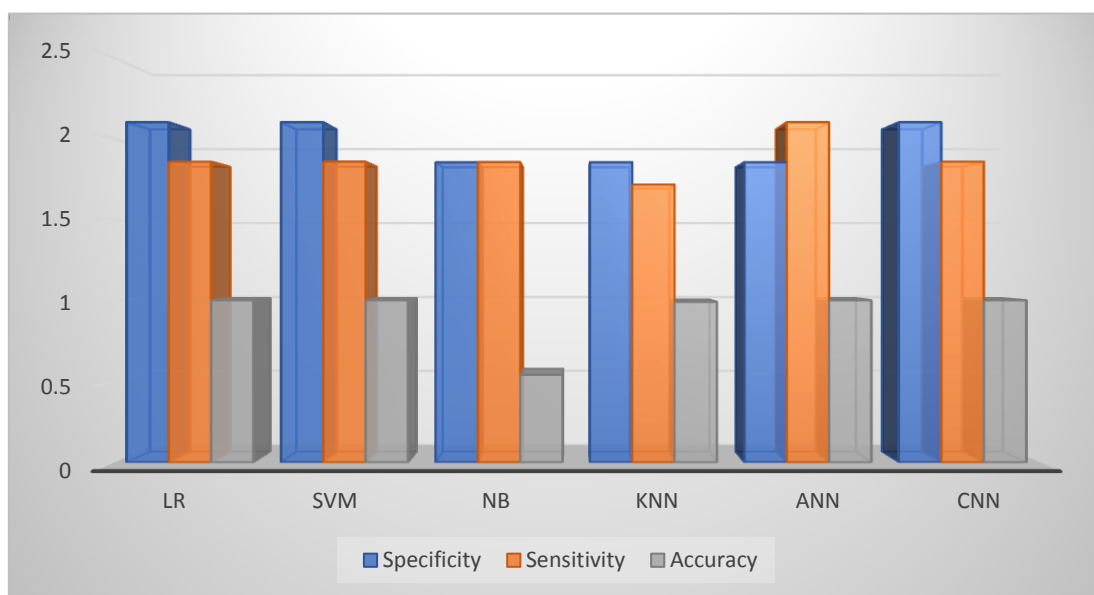
**Table 3:** Summary Data from Screening for Autism Spectrum Disorder in Children

Classifier	Specificity	Sensitivity	Accuracy
LR	2.1	1.8566	100%
SVM	2.1	1.8568	100%
NB	1.8531	1.8543	54%
KNN	1.8531	1.7153	99%
ANN	1.8531	2.1	100%
CNN	2.1	1.8567	100%



**Table 4:** Comprehensive Data from Adolescent Screening for Autistic Spectrum Disorder

Classifier	Specificity	Sensitivity	Accuracy
LR	2.1	1.7777	87.23
SVM	2.1	1.6666	71.45
NB	1.8531	1.9999	87.23
KNN	1.8531	1.4444	71.45
ANN	2.1	1.3333	87.23
CNN	2.1	1.7895	85.77



While assessing a few machine learning models on the ASD Juvenile diagnosis dataset, specialists observed that their expectations were inside the first dataset's mistake edges. The most reduced precision was accomplished utilizing a K-NN classifier with K=5. On the essential dataset, CNN classifiers had the best forecast precision.

#### **4. Conclusion**

A few machine learning and profound learning strategies were tried for their capacity to analyze Autism Spectrum Disorder in this review. Models created for ASD diagnosis utilizing non-clinical datasets from youngsters, teenagers, and grown-ups were assessed utilizing an assortment of execution evaluation pointers. Conversely, the CNN-based model outflanked any remaining model-building techniques on the other two datasets, emphatically showing that it very well may be utilized to identify ASD rather than the more conventional machine-learning classifier that had been proposed in past examinations.

#### **References**

- [1] Baihua Li, Arjun Sharma, James Meng, Senthil Purushwalkam, and Emma Gowen. (2017) Applying machine learning to identify autistic adults using imitation: An exploratory study. *PloS one*, 12(8): e0182652.
- [2] Fadi Thabtah. (2017). "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment." In *Proceedings of the 1st International Conference on Medical and Health Informatics*, pp. 1-6. ACM.
- [3] Thabtah, Fadi, Firuz Kamalov, and Khairan Rajab. (2018) "A new computational intelligence approach to detect autistic features for autism screening." *International journal of medical informatics* 117: 112-124.
- [4] Constantino, John N., Patricia D. Lavesser, Y. I. Zhang, Anna M. Abbacchi, Teddi Gray, and Richard D. Todd. (2007) Rapid quantitative assessment of autistic social impairment by classroom teachers. *Journal of the American Academy of Child & Adolescent Psychiatry* 46(12): 1668-1676.
- [5] Daniel Bone, Chi-Chun Lee, Matthew P. Black, Marian E. Williams, Sungbok Lee, Pat Levitt, and Shrikanth Narayanan. (2014). "The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody." *Journal of Speech, Language, and Hearing Research*, 57(4): pp.1162-1177.
- [6] Vaishali, R., and R. Sasikala. "A machine learning based approach to classify Autism with optimum behaviour sets. (2018) "International Journal of Engineering & Technology 7(4): 18.
- [7] Daniel Bone, Matthew S. Goodwin, Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, and Shrikanth Narayanan. (2015) "Applying machine learning to facilitate autism diagnostics: pitfalls and promises." *Journal of autism and developmental disorders* 45(5): 1121-1136.
- [8] Thabtah, Fadi. "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. (2018) "Informatics for Health and Social Care: 1-20.
- [9] J. A. Kosmicki, V. Sochat, M. Duda, and D. P. Wall. (2015) "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning." *Translational psychiatry*, 5(2): e514.