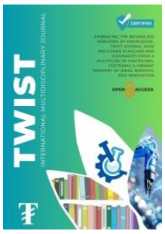
DOI https://doi.org/10.5281/zenodo.10049652#198

TWIST, 2024, Vol. 19, Issue 2, pp. 621-627

TWIST

Journal homepage: www.twistjournal.net



A New Model for Autism Diagnosis Using Meta-Classifiers

Hind Abdulrazzaq Mohammed Ali*
Civil Engineering Department, University of Technology-Iraq, Baghdad, Iraq
[**Corresponding author*]

Douaa Ibrahim Alwan Alsaadi
Software Engineering Department, Higher Health Institute, Najaf, Iraq

Abstract
The illness that has drawn the greatest attention from scientists recently is autism, which is also the most well-known. People of various ages might be affected by this condition. All currently available datasets are of poor quality for data analysis. Base classifiers are the main focus of most related research. In this study, we proposed a thorough comparison that makes use of efficient preprocessing and ensemble approaches to enhance the diagnosis of the autistic condition. Outliers and missing data are resolved during the preprocessing step. The ensemble methods include bagging, stacking, boosting, and voting. Two well-known data mining technologies are used to assess comparisons. The findings collected to demonstrate that our work improves classification performance in terms of precision, recall, accuracy, and F1 compared to the base approaches. The criteria greatest values are attained at 100%.

Keywords
Data mining, Pre-processing, Autism mellitus, Meta-classifiers, Preprocessing, Rapid miner

INTRODUCTION

Autism is a neurodevelopmental condition that damages neurons, and it has recently become more prevalent. Children, teenagers, and adults are typically affected. The patient develops a mental disability as a result of the disease destruction of nerve cells. Both the method of identifying and diagnosing autism spectrum disorders and the impetus for their global discovery are challenging issues. Long-lasting inefficient techniques that cost a lot of labor were employed. It is critical to creating these diagnosis procedures as soon as possible for simpler deployment and better results. We looked at the published papers on autistic disorder from 2010 to 2021. Missing values and classification accuracy are the two greatest problems with the autism dataset. The scientists proposed a model that accurately forecasts the autism condition and yields findings in 2020. Through a technique to extract data and determine if the patient has autism, they looked at the validity of the AC association categorization. The highest scores attained were 97.5% [1].

The scientists published a suggestion in 2021 to enhance categorization and forecast the status of autism spectrum disorder, which weakens the body, consumes nerve cells, and obliterates the patient intellect. In their article, they demonstrated a better outcome than the unnecessary effort. The values that the present writers obtained match this value exactly. Their Support Vector Machine (SVM) study has a 97% accuracy rate. This indicates that we do better when using more than seven algorithms and recordings 100% of the findings. This indicates that our work with all algorithms makes accurate predictions, produces better outcomes, makes predictions rapidly, and enhances the effectiveness of data categorization [2]. The present writers research the influence of ensemble approaches in addition to preprocessing phases to solve these issues. The present authors suggested a comparison utilizing ensemble approaches together with preprocessing techniques for an autism diagnosis. The aforementioned dataset may be downloaded from the UCI repository. To outperform their competitors, two programs called WEKA and Rapid Miner are used for data mining activities. Preprocessing methods were applied in the initial step to deal with missing results and find outliers. Ensemble classifiers including bagging, boosting, voting, and stacking are used during the classification step. Using stacking, the greatest accuracy rate of 100 % was attained. Then, we employed methods that did not involve preprocessing. Through bagging, the greatest accuracy rate of 97% was attained. After that, we compared the outcomes with the Rapid Miner tools using the WEKA tools. The accuracy rate was 100% across all experiments. Our research demonstrated that the outcomes were superior to those of earlier studies. As a result, our work dependence on ensemble algorithms proved to be a resounding success and offered good forecasts for enhancing classification performance in this setting. The submitted data had serious problems, which we fixed and made up for. Through the mean, the location of the missing data was inferred. Similar to this, the K-nearest neighbor (KNN) approach is used to identify outliers. The amazing and good enough goes above and above to solve these issues.

RELATED WORK

According to several types of research, the diagnosis of autism may be expected between 2010 and 2021. A list of some of the most significant works has been put together by us. With two data mining tools, Rapid Miner and WEKA, the current authors used ensemble algorithms, which included boosting, bagging, stacking, and voting. These outperformed competitors and produced positive results by accurately predicting 100% of the time on the dataset used to model the autism disease. Authors suggested a way in 2011 to forecast better outcomes and portray favorable results more effectively than the work reported. They persisted and put a lot of effort into coming up with a plan to show a better treatment for autism. Using trusted robot software, this technique is used to identify the afflicted youngster and know how to respond favorably or unfavorably. They predicted positive outcomes, and their accuracy rate was 94%; nevertheless, this figure was lower than what we discovered in the current research, where our accuracy rate was 100%. The findings demonstrated that different children are influenced differently [3].

In 2012, authors proposed Screen Mobile Computers (TSMC) as a remedy for autism spectrum diseases. The TSMC gadget encourages autistic youngsters to exercise, gain experience, and strengthen their neurological and cerebral capacities, hence the authors suggested a laptop computer with a touch screen. These writers' experience is crucial for diagnosing and developing autism. Our work looks to be more precise and accurate in contrast to all prior work, making it superior in terms of improving classification performance [4]. The rise of technology and the increasing attention given to autism in many nations across the world have made it important to discover answers. This has led researchers to look for methods and submit research papers that enhance autism cases and the functionality of the workbook. Authors in 2016 employed strategies and tactics to enhance autism to be able to gauge children social behavior and develop original and distinct forecasts. They demonstrated a tool that uses electromyography of the facial muscles to identify and investigate smiles. They also worked with the human-robot for improved prediction since it produced acceptable outcomes. It produced better outcomes than our studies and accurately predicted how well people with autism would do [5]. For children with autism, one of the most common disorders in the world, scientists recommended employing distributive semantic models to find unexpected terms in stories in 2013. Researchers are forced to look for information, which leads to more accurate results 46.3% and 74.0%, respectively, were the accuracy and recall percentages. Additionally, their work measuring value forecasts favorable developments that will benefit people on the autism spectrum and attempt to foster such. The F1 score was 57.0%, which represents satisfactory outcomes. It demonstrates how much better our counterpart work ours in the present publication. All of our values were 100%, and the outcomes show that we may expect improved outcomes from our efforts to enhance categorization performance [6].

The authors offered a recommendation in 2017 to enhance categorization performance. They worked on autism prediction well enough to have good outcomes. Based on behavioral data, they presented a research study on diagnosing and tracking treatment progress among kids with difficult-to- treat autism. The authors made several observations on computer game players with autism who also employ computer science, computer procedures, and unique study strategies. They created this method for improved prediction due to the complexity of the disease and the difficulties in forecasting it. They had an accuracy rating of 80% and yet got good outcomes. Each of precision, recall, and F1 was 81.05%, 86.12%, and 83.51% respectively. They used several algorithms to demonstrate their findings [7]. A rule-based language programming performance that might describe the DSM BED was proposed by authors in 2018. A description of the temperament and conduct mentioned in the documents was what the writers aimed to glean. The findings revealed that the precision was 76%. The recall value was 42%, and the accuracy was 76%. They were successful in locating the parser for around 5,000 documents. They compared how many differences were there throughout the course of these years. They were able to work on and enhance autism performance to discover the most notable outcomes.

We demonstrated that our results are more accurate and recallable than existing results, and we provided them more quickly for more accurate forecasting [8]. A new model was presented by authors in 2019 to enhance the performance of the classifier. Supervised machine learning was utilized (ML). With the methods utilized in their article, such as SVM, boosting, and decision trees, their study sought to achieve high accuracy. Additionally, they used these algorithms to provide improved predictions to fill in the gaps in the data. Well-defined outcomes from their trials were attained during these stages. The best results were 97.1%, 97%, 98%, and 0.98% for accuracy, precision, recall, and F1. Although the results of their tests were promising, our method in this research is superior and provides reliable predictions for the data related to autism [9].

THE PROPOSED MODEL

Here, we looked into and discovered the aggregation approaches using two different techniques. Pre-treatment is done in the initial stage. The ensembles algorithm, which combines boosting, bagging, stacking, and voting, is then used to get high-quality features. To fill in the missing data in the second step, we created and employed the mean and k-nearest neighbors by stacking, bagging, voting, and boosting. All ensemble algorithms achieve high accuracy at 100%. Figure 1 depicts the suggested model.

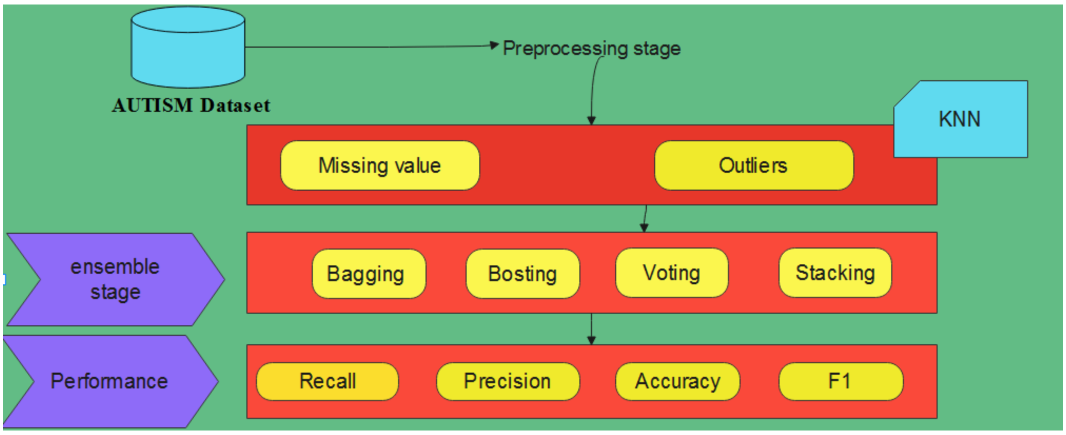


Fig. 1 The proposed model

PRE-PROCESSING STAGE FOR MISSING VALUES

The data were acquired from UCI. The Adult Autism ML webpage has outliers and missing values. Pre-processing was used to process these missing data and replace them with fresh ones. The KNN approach was also used to find missing values and replace them. The k variable has the value of ten when this is taken into account. The values put the KNN in the position of outliers, and good results were produced to boost this data's performance and predictability.

The Ensembles Stage

Ensemble techniques are one of the ML classifiers that improve classifier performance and yield better predictions. This stage also includes the methods used to get good results. The research revealed that classifiers that used stacking, bagging, voting, and boosting outperformed their rivals by up to 100%. We discuss each suitable method separately after the preprocessing stage. These classifiers are among the best and provided findings that were remarkable. Then, in order to improve forecasting, we attempt to forecast and improve this dataset while simultaneously building a model to advance better results.

Boosting

A descriptive algorithm called AdaBoost improves the performance of many-based classification algorithms to get better results. The reinforcement process is a method for combining several weak learners into one strong learner who can make improved predictions. It is an associative system that brings together several models to get superior results. Because it gave excellent results with 100 percent accuracy when used with the two programs WEKA and Rapid Miner, it was used in this investigation both before and after pretreatment. This outcome is superior for improving the classifier's performance and forecasting successful outcomes for data on adults with autism spectrum disorders.

Voting

There are at least two base classifiers in the voting component of the voting technique, which is a sub-process. This technique creates a regression model where the majority chooses the classification, a classification model based on the learners, or a model with high accuracy. Without preprocessing, it produced results at a rate of 97%, but preprocessing in Rapid Miner produced results at a high rate of 97 %. As these datasets need to be improved and work on better prediction, it was a high outcome for enhancing classification performance and performing the prediction process successfully.

Stacking

As an additional option, we used the dataset to apply the stacking approach. The training of foundational learners is the initial phase. The procedure of testing the foundational learners is the second component. We used it in the Rapid Miner tool, and the accuracy in both data mining tools achieved 100% as a consequence of our efforts. It was beneficial for raising the classifier's and predictor's effectiveness as well. The base learners and the usual agglutinative sub-learner processes were the two sub-processes of the stacking technique, a hypothetical process.

Bagging

As an alternative, we used the stacking approach and put it into practice on the dataset. The training of basic learners makes up the initial phase. The procedure of testing the fundamental learners makes up the other component. When we used it with the Rapid Miner tool, we achieved good results with both data mining tools' accuracy reaching 100%. Additionally, it helped to anticipate better and boost classifier performance. The stacking approach consisted of two sub-processes: the base learners and the standard agglutinative sub-learner process.

EXPERIMENTS AND RESULTS

In the first experiment, we looked at how the autism data set was affected by collecting techniques without any preprocessing. We used a combination of algorithms, such as stacking, boosting, voting, and bagging. On Weka Program, we used these algorithms. The precision of Bagging was 100%, whereas it was the highest. Therefore, the stacking method outperforms the other methods and also enhances classification performance using ensembles and Weka program without preprocessing. The bagging, voting, and boosting methods yielded the best accuracy, precision, recall, and F1 at

95.77%, and 97.16% and 99.77 % respectively. The stacking value is one of these values with the highest value. Accuracy, precision, recall, and F1 values were all 99.77 % for every measurement, and this meant that every measurement was accurate. The classifier's performance and prediction accuracy, in general, can be improved with the help of these results. The ensembles without preprocessing combined with KNN, and RF utilizing the Rapid Miner Tool yield the best results, as shown in Table 1.

Table 1 The obtained results through ensemble without pre-processing in conjunction with DT and RF through Weka Program

Classifier	Precision	Recall	Accuracy	F1
Bagging	100 %	100 %	100 %	100 %
Bosting	95.77 %	95.77 %	95.77 %	95.77 %
Voting	97.16%	97.16%	97.16%	97.16%
Stacking	99.77 %	99.77 %	99.77 %	99.77 %

Table 1 shows the findings for the ensemble with stacking, bagging, voting, and boosting without preprocessing as a good predictor of significantly better classification performance for adult autism spectrum data. The maximum degree of accuracy was attained at 100%. By analyzing promising outcomes for prediction and performance enhancement. In the second trial, we used adult autism data and ensemble techniques without any preparation. The outcomes showed a strong and clear prediction for performance improvement. A high rate of 100% was achieved for accuracy, precision, recall, and F1, which is regarded as the best since it will produce accurate predictions when bagging, boosting, voting, and stacking are used. Table 2 displays the measurement results that we applied using Rapid Miner's aggregation techniques, demonstrating how well it outperformed its competitors and prior knowledge.

Table 2 The obtained results through ensemble without pre-processing with DT and KNN with Rapid Miner tool

Classifier	Precision	Recall	Accuracy	F1
Bagging	100 %	100 %	100 %	100 %
Bosting	97.99 %	97.99 %	97.99 %	97.99 %
Voting	98%	98%	98%	98%
Stacking	100 %	100 %	100 %	100 %

We then conducted an experiment to treat missing values and outliers by applying ensemble algorithms with preprocessing techniques. See Table 3.

Table 3 The obtained results through ensemble with pre-processing with DT and KNN with Weka Program

Classifier	Precision	Recall	Accuracy	F1
Bagging	100 %	100 %	100 %	100 %
Bosting	99.99 %	99.99 %	99.99 %	99.99 %
Voting	98%	98%	98%	98%
Stacking	100 %	100 %	100 %	100 %

After entering the data into the Weka program, machine learning techniques were applied, but this time with pre-processing techniques, such as replacing the missing values with correct values, as well as replacing the shah values with good and correct values, as shown in the above table. The highest value obtained was 100. Then we conducted another experiment with the Rapid Miner program, but with pre-processing techniques, the techniques of replacing values with the nearest neighbors, and the technique of detecting abnormal values with correct values. See Table 4. We will explain the most important results that have been obtained.

Table 4 The obtained results through ensemble with pre-processing with DT and KNN with Rapid Miner Program

Classifier	Precision	Recall	Accuracy	F1
Bagging	100 %	100 %	100 %	100 %
Bosting	100 %	100 %	100 %	100 %
Voting	100 %	100 %	100 %	100 %
Stacking	100 %	100 %	100 %	100 %

Through the results above, it was found that the experience we worked on with the Rapid Miner program was better than the experience with the Weka program, and we got very high results with all the algorithms. We got 100%.

Evaluation Metrics

To evaluate, accuracy, precision, recall, and F1 were applied. These measures define in Table 3.

Accuracy	$(TP+TN)/(P+N)$
Precision	$(TP)/(TP+FP)$
Recall	TP/P
F1	$\frac{2\times precision\times recall}{precision+recall}$

Data Collection

The autism dataset for adults that was retrieved from the UCI ML data repository was the one we focused on and used for this work among the several datasets. The dataset includes 21 features as well as a collection of 704 entries pertaining to autism. It is used to find characteristics on datasets for autism and to work on creating, improving, and diagnosing autism cases generally. All numerical and nominal characteristics are included in the dataset. Although these are well-known data, there are some missing numbers. We will try to fix them and create strategies to raise the level of the data and produce better results.

DISCUSSIONS

We utilized two data mining tools in this post. The first one was Rapid Miner version 9, and the second one was WEKA versions 3.5 and eight, which were used to process the data and apply algorithms for a collection of patient records relating to autistic patients. We used the bagging, boosting, voting, and stacking ensemble algorithms. Additionally, these aggregates were applied to the DT, KNN, and RF. In the two tools, the implementation of these algorithms with and without preprocessing is assessed. These methods produced excellent results for all bagging combinations when used in conjunction with preprocessing. High findings were received, estimated at 100%. We also used the Rapid Miner program to apply these algorithms without any preprocessing, and the results were successful. Recall, precision, F1, and accuracy were also considered four crucial assessment criteria. Since the accuracy is determined by the ratio of true positives to total positives, true positives over false positives, and true negatives over false negatives, the results were achieved using ensembles with KNN and RF with Rapid Miner without preprocessing. The outcomes of using these algorithms, which included bagging, boosting, voting, and stacking, are displayed in the table below. We obtained accurate estimations, as shown in Table 2, where the outcomes were likewise given with stacking at 100%. As seen in Table 2. The ensemble algorithms, which include bagging, boosting, voting, and stacking, were implemented in the ensembles using DT and RF with the Rapid Miner tool, and the results were obtained. The table below displays the good and high results for a better prediction of 100%. We utilized the outcomes from ensembles with preprocessing, DT, and RF with Rapid Miner in the third table. The four measures of bagging, boosting, voting, and stacking are shown in this table. They all had strong outcomes and are good indicators of an autistic improvement. The sum of the four measures in the table was 100%. The findings in the fourth table were obtained using ensembles by preprocessing, KNN and RF.

A comparison of our work with past attempts at this issue is shown in Table 6. Since our paper had a maximum accuracy of 100 %, performance needs to be improved. In comparison to previous models, ours improves classification performance while still requiring more development and forecast labor. Our results are compared to those of their equivalents in Figure 2.

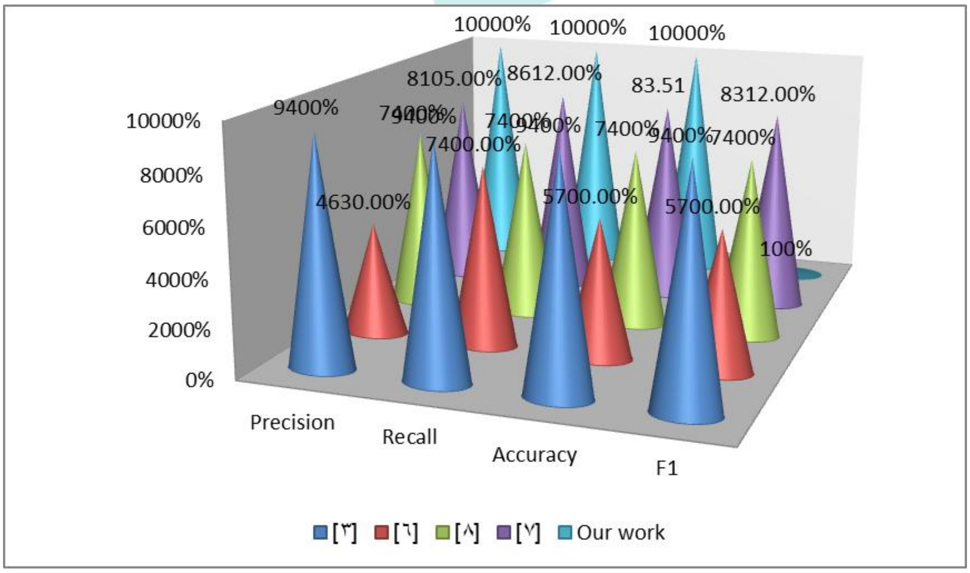


Fig. 2 The Comparison results among our work and others

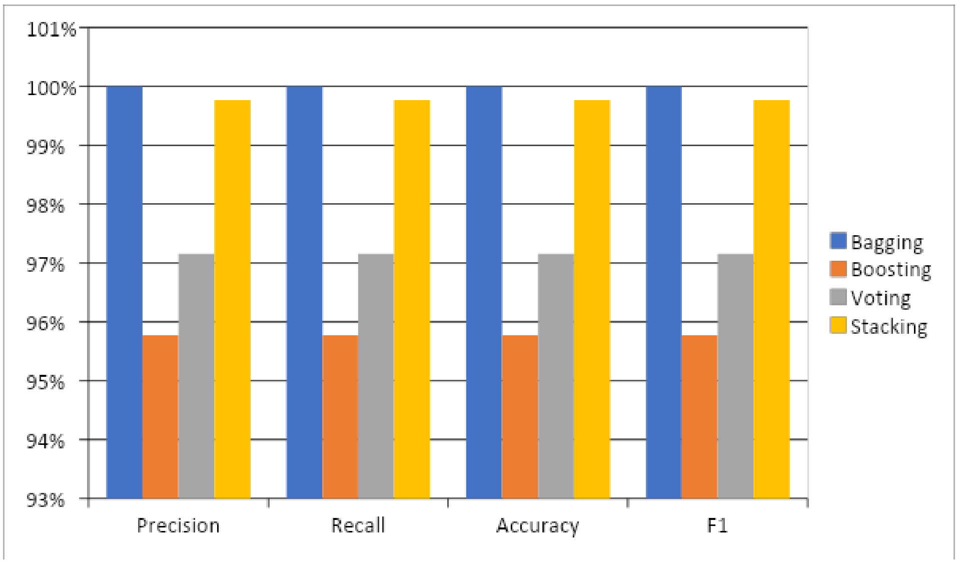


Fig. 3 The Comparison results among our work through ensemble without pre-processing in conjunction with DT and RF through Weka Program

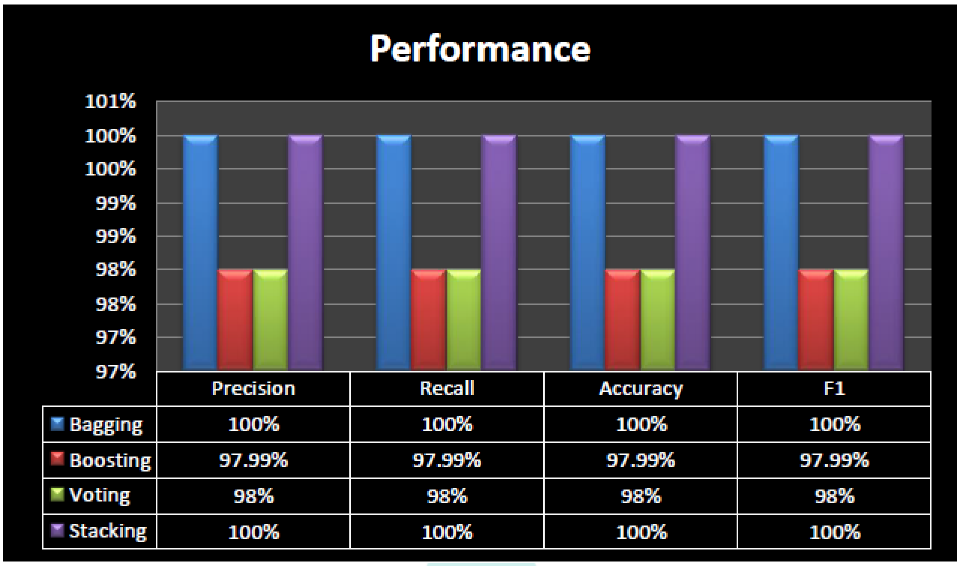


Fig. 4 The Comparison results among our work ensemble without pre-processing with DT and KNN with Rapid Miner tool

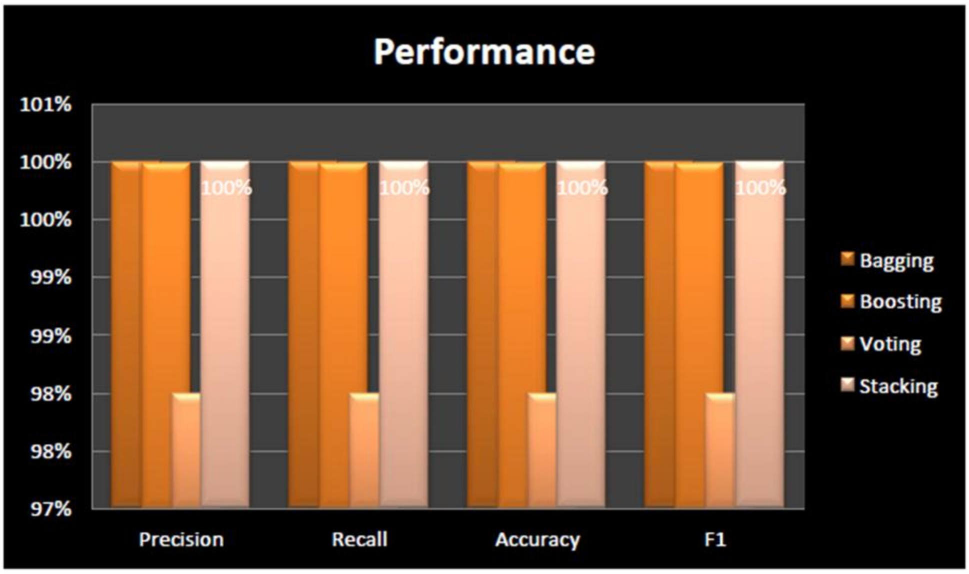


Fig. 5 The Comparison results among our work through ensemble with pre-processing with DT and KNN with Weka Program

Fig. 6 The Comparison results among our work results through ensemble with pre-processing with DT and KNN with Rapid Miner Program

CONCLUSION

When it comes to its effects on the psyche and mind of the patient, autism is one of the most common disorders and most vulnerable to people. When a child's speech is slow or unresponsive and we are unable to reply, worry regarding autism may develop because some people's mental health and anxiety limit the depth of their response and comprehension. In this post, we applied sound methods and achieved desirable outcomes. These methods demonstrate their aptitude. As DT methods, KNN technology, and RF were combined with the ensemble algorithm, which comprises bagging, boosting, voting, and stacking, the performance of autism classification and the prediction was significantly enhanced. Two data mining tools, one with and one without preprocessing, were used to apply these strategies. Has a 100% value across all metrics. Additionally, Rapid Miner and preprocessing reached 100% for all measures. The identical algorithm was also employed without any preprocessing, and the results indicated that bagging, AdaBoost, and voting all averaged 100%, while stacking achieved 100% across the board. These findings demonstrated a solid advancement, a high level, and a fantastic prediction to enhance classification performance and work on the quick prediction of autism. It also beats earlier research in terms of the following metrics: accuracy, recall, precision, and F1.

The suggested method in this article shows that it is a great method for enhancing classification performance and for accurately predicting autism more accurately than business assessment. Regarding the results that we demonstrated in this post, this strategy demonstrated tremendous success. In terms of future development, we'll focus on the genetic algorithm and other optimization methods. We are attempting to identify the best approaches to enhance performance and forecast autism in a high and outstanding manner. We strive to develop our work, achieve results, and have a very high future prediction. We have demonstrated the quality and success of our work, and we will do it again by citing examples from earlier projects.

REFERENCES

1. Chowdhury, K. and M.A. Iraj. Predicting Autism Spectrum Disorder Using Machine Learning Classifiers. in 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT). 2020. IEEE.
2. Emon, M., et al., A Comparative Analysis of Autistic Spectrum Disorder (ASD) Disease for Children using ML Approaches.
3. Feil-Seifer, D. and M.J. Matarić. Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features. in 2011 6th ACM/IEEE international conference on human-robot interaction (HRI). 2011. IEEE.
4. Rasche, N. and C.Z. Qian. Work in progress: Application design on touch screen mobile computers (TSMC) to improve autism instruction. in 2012 Frontiers in Education Conference Proceedings. 2012. IEEE
5. Hirokawa, M., et al. Design of a robotic agent that measures smile and facing behavior of children with Autism Spectrum Disorder. in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 2016. IEEE.
6. Rouhizadeh, M., et al. Distributional semantic models for the evaluation of disordered language. in Proceedings of the conference. association for computational linguistics. north american chapter. meeting. 2013. NIH Public Access
7. Kołakowska, A., et al., Automatic recognition of therapy progress among children with autism. Scientific reports, 2017. 7(1): p. 1-14.
8. Leroy, G., et al., Automated extraction of diagnostic criteria from electronic health records for autism spectrum disorders: development, evaluation, and application. Journal of medical Internet research, 2018. 20(11): p. e10497.
9. Raju, N.G., et al., prognostication of Autism Spectrum Disorder (ASD) using Supervised Machine Learning Models. International Journal of Engineering and Advanced Technology (IJEAT), 2019. 8(4): p. 2249-8958