

ENSEMBLE LEARNING MODEL FOR SCREENING AUTISM IN CHILDREN

Mofleh Al Diabat¹ and Najah Al-Shanableh²

^{1,2}Department of Computer Science, Al Albayt University, Al Mafraq- Jordan

ABSTRACT

Autistic Spectrum Disorder (ASD) is a neurological condition associated with communication, repetitive, and social challenges. ASD screening is the process of detecting potential autistic traits in individuals using tests conducted by a medical professional, a caregiver, or a parent. These tests often contain large numbers of items to be covered by the user and they generate a score based on scoring functions designed by psychologists and behavioural scientists. Potential technologies that may improve the reliability and accuracy of ASD tests are Artificial Intelligence and Machine Learning. This paper presents a new framework for ASD screening based on Ensembles Learning called Ensemble Classification for Autism Screening (ECAS). ECAS employs a powerful learning method that considers constructing multiple classifiers from historical cases and controls and then utilizes these classifiers to predict autistic traits in test instances. ECAS performance has been measured on a real dataset related to cases and controls of children and using different Machine Learning techniques. The results revealed that ECAS was able to generate better classifiers from the children dataset than the other Machine Learning methods considered in regard to levels of sensitivity, specificity, and accuracy.

KEYWORDS:

Artificial Neural Network, Autism Screening, Classification, Ensemble Learners, Predictive Models, Machine Learning

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a developmental condition that affects an individual's social and communication skills.¹ ASD traits are often observed in children of young age, i.e. 2-5 years. Recently, a number of scientists in the behavioral science and computational intelligence research fields, have investigated intelligent methods such as Artificial Intelligence (AI) and Machine Learning to detect the autistic traits at the diagnosis level.^{2,3-6} However, little research has been conducted on the use of AI and Machine Learning methods to improve ASD pre-diagnosis.^{7,8-11} To address this problem, this research develops a new Machine Learning framework of ASD screening using a new class of learning called Ensemble Learners.

The current research aims to improve the ASD screening process by incorporating a new Ensemble Learning method based on Artificial Neural Network (ANN). ANN models have proved their superiority in several classification domains such as image classification, pattern recognition, speech recognition, and medical diagnosis.^{12,13-14} The power of these models is gained for several reasons, for instance, their ability to learn, their ability to generalize, and their fault tolerance.^{12,15} ANN models are normally created using an exhaustive trial-and-error

approach; to avoid this issue we adopt a self-structuring mechanism to intelligently amend the outcomes without relying on the trial-and-error approach.¹³ This study also aims to reduce the role of domain expert rules within current screening tools by utilizing classifiers derived by Machine Learning to improve performance. This is helpful for instances that are difficult to predict (instances unclearly exhibiting autistic traits). The main research question that this thesis will answer is:

- Can Ensemble-based Learning screening methods improve ASD screening of children?

The Ensemble-based Learning approach combines a set of learned classifiers from a training dataset, then individual decisions of these classifiers are utilized to produce the class of a test data instance. This learning approach reduces biased decisions and improves performance.¹² The Ensemble-based approach learns classification models incrementally using a base classification algorithm, does not forget the previously learned models, and does not require access to the previously seen training data sets. This ensures that a collective decision is sought for test data class assignment. Boosting and Bagging are two common Ensemble Learners¹⁶ and will be discussed briefly in Section 2.

In this article, an ensemble-based classification model has been proposed and is called Ensemble Classification for Autism Screening (ECAS). The base classifier for ECAS is based on ANN which keeps amending the models' structures and adds them to ECAS. This base learning algorithm simplifies deriving classifiers with good generalization ability. The algorithm does not randomly amend the neural network model, rather it adjusts it intelligently. The ECAS model was evaluated on a real dataset related to infants that had recently been collected.¹⁷ The dataset was downloaded and then extensive experiments conducted using ECAS and other known Machine Learning techniques including decision trees, voting, and rule induction among others¹⁸. The main criteria for comparison are error rate, specificity, and sensitivity (Section 4 gives more details).

The paper is organized as follows:

Section 2 presents the problem and related literature on the adaptation of AI and Machine Learning techniques for autism screening. In Section 3, the methodology of the research is discussed. The dataset and features are presented in Section 4, and experimental analysis is discussed in Section 5. Lastly, conclusions and future work are given in Section 6.

2. THE PROBLEM AND LITERATURE REVIEW

2.1 The Research Problem

ASD screening does not require a clinical setup and usually includes a set of behavioral questions, i.e. social, repetitive behavior, communication, etc., looking for any symptoms of autism in an individual.^{19,20-21} Often, the individual's medical professional, caregiver, parent, or teachers answer these questions on behalf of the child during the screening process then a final score is produced to pinpoint if the child is potentially exhibiting autistic traits and if he requires any additional medical check-ups. Since the aim of the screening process is to assess any possibility of autistic symptoms based on questions (attributes) and a dependent variable

(existence of autistic traits) then we can deal with this problem as a predictive supervised learning task (Figure 1).

Based on Figure 1, cases and controls are collected (training dataset) using a screening method such as AQ or Q-CHAT.^{22,23} The training dataset is then processed to build a predictive model, also known as a classifier, using a classification method. The classifier is then used to assign the target class (ASD traits or not) to new individuals (test cases that the model has never encountered during the training phase) accurately. We think that if Machine Learning is utilized instead of conventional scoring functions to screen individuals for the autism application, this can improve the performance of the screening in terms of predictive accuracy, sensitivity, and specificity. More importantly, the decision will be totally objective, rather than subjective, as models are derived from historical cases and controls; this can assist clinicians and medical specialists to better diagnose individuals.

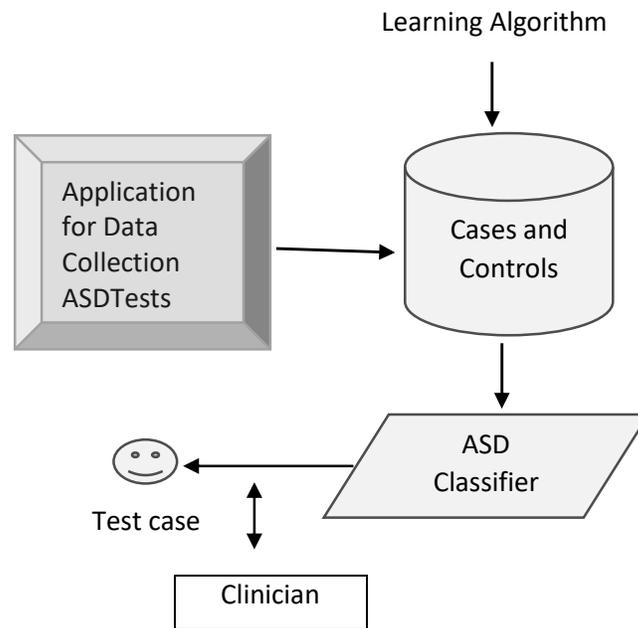


Fig. 1: Autism Screening as a Predictive Problem in Machine Learning

2.2 LITERATURE REVIEW

There are many types of Ensemble Learning models such as Bagging and Boosting.^{16,24} In the Bagging Learning model, the user must identify the size of the sample used to train the local classifier. This sample is often drawn from the training dataset, randomly based on the size specified by the end user. For example, if the user assigned the sample size to 50%, then new random instances with a size equal to 50% of the training dataset size will be created randomly. Moreover, the parameter of how many bags (classifiers) must be specified by the user. Normally, this parameter will be dynamic and continue to increase until no further improvement on the resulting classifiers is observed. However, continuing to increase the number of bags can lead to higher overhead costs in terms of processing time and computing resources.²⁴

Boosting learning models such as AdaBoost were disseminated to overcome the cost associated with wrong predictions during the class assignment process (classification phase).²⁵ They work by constructing an initial classifier using a base classification method such as decision trees.²⁶ Then, the initial classifier will be applied to the training dataset and each wrongly classified training data will be assigned a new weight. Once the training instances have been assigned weights, then a new classifier is derived from the updated training dataset and the process is repeated until no further improvement can be achieved in terms of predictive accuracy. Finally, whenever a test data is about to be predicted, then models learned with different data weights are used to decide its class in a collective manner.²⁵

The existing clinical procedures for diagnosing autism demand long waiting times.^{5,6,17} In an attempt to reduce this and facilitate early detection of autism amongst children and toddlers, researchers have been attempting to develop new initial screening tools using Machine Learning. While a few researchers have resorted to acoustic analysis of vocal production to detect autistic traits,²⁷ others are conducting non-verbal or behavioral analysis using video recordings.²⁸ A few other researchers have adopted the traditional approach of a clinical questionnaire that feature selection techniques to develop a prediction model using the most effective features.^{7,8-10}

A study recently in Bangladesh analyzed which individual regional features are an indication of autism.²⁹ This research targeted children aged between 16 and 30 months who were residents of various regions in Bangladesh. The data was submitted by the parents using Autism Barta apps which were developed using the M-CHAT standard and contain 23 questionnaires.³⁰ These entries, combined with the fieldwork conducted, resulted in a dataset of 642 instances. The raw data was cleansed, and pre-processing techniques were used to make the data suitable for analysis using Machine Learning tools such as WEKA (The Waikato Environment for Knowledge Analysis).³¹ The analysis was carried out using 10-fold cross validation on tree-based classifiers.³²⁻³³ Performance metrics such as accuracy, precision, recall, f-measure, and kappa statistics, among others, have been used in the comparison analysis. The C4.5 algorithm showed competitive accuracy, precision, recall, and an f-measure of 98.4%. In addition, the C4.5 algorithm was selected as the best model to extract 9 rules, 7 items, and 16 conditions about normal and autistic traits. The study concluded that only 8 out of the 23 features are needed to accurately predict if a child is autistic. These features also gave insight into the conditions required to detect autistic traits based on which region of Bangladesh the child came from.

A study aimed at answering if ML models built using minimum autistic features deliver better performance in the classification of autistic traits.¹⁰ The study used an ASD diagnostic dataset, which originally was comprised of 21 features.¹⁸ The noisy features in the dataset are eliminated by applying dimensionality reduction with feature selection. WEKA is used to build predictive models using original dataset and modified dataset post feature selection. Moreover, 10-fold cross-validation partition and various classifiers were employed in the experiments. The performance of the derived models was analyzed using evaluation metrics such as Feature Reduction Ratio, and accuracy, among others. For the original dataset, the support vector machine classifier reported the highest prediction accuracy of 99.66%. The Binary Firefly feature selection technique showed that a subset of 10 features of the dataset is optimum and has a feature reduction ratio of 0.48. The study concluded that feature selection improved the predictive performance of the model and 10 features were enough to predict autistic traits.

A recent research paper aimed to predict autism in children by using classification methods.⁹ The dataset used was obtained via the ASD Tests app,¹⁸ and the target age group of the child dataset

was 4-11-years. The dataset consisted of 292 instances and 19 different attributes. Various models were built using a 70 percent training set and 30 percent testing set, and models' evaluation was carried out using accuracy, sensitivity, specificity, precision, and f-measure. The study stated that overall, linear discriminant analysis performed well scoring acceptable accuracy, specificity, precision, and f-measure rates.

A mobile screening application for autism called ASDTests was developed.¹⁷ The author aimed at carrying out feature analysis on the collected data and then predictive analysis to study if this screening app increased efficiency and accuracy of ASD screening. The app can reach out to a wide audience as it is provided in 11 languages. Moreover, it takes into consideration the age of the test-taker and provides four categories: toddlers, children, adolescents, and adults. The data collected through the app is then used for analysis. The most influential features among the collected data are determined for each age group by using Wrapping Filtering along with a Naïve Bayes probabilistic classifier.³⁴ The predictive analysis is carried out using a Naïve Bayes algorithm. The classification performance is determined using evaluation measures such as accuracy, sensitivity, and specificity rate. The study concluded that the features of the data collected using the proposed app resulted in the building of effective predictive models and that the app will be an important tool for data collection.

A recent study proposed the use of support vector machine to detect autistic traits in individuals.³⁵ The underlying dataset for this study is the "Resting-state functional magnetic resonance imaging" (fMRI) data which is taken from the "Autism Brain Imaging Data Exchange" (ABIDE) database. It is comprised of 107 instances of which the authors have used only 84 instances in the actual study because the rest didn't adhere to the specifications of the study. In the final sample size, 45 are autistic patients and 39 controls. The study reported that the support vector machine achieved a predictive accuracy of 96%; it further concluded that this performance is maximized and stabilized when the optimal number of base classifiers is 500, and the first 272 features make up the optimal feature set.

A study attempted to combine the traditional structured questionnaire-based approach with semi-structured video analysis for early detection of autism in children.⁸ The authors initially trained two algorithms; the first algorithm used answers to a short, structured questionnaire, which was completed by the parents. The second algorithm was trained by tagging the key behaviors of the child recorded in a short, semi-structured video taken in the home environment. The training dataset is obtained from score-sheets of ADOS and ADI-R diagnostic tools for children.^{36,37} The target age was 18 to 84 months. Intense feature selection and encoding techniques were applied to the test data to overcome the problem of scarcity, sparsity, and imbalance that existed in the training dataset. Logistic regression was then used to combine the numerical response of each of the parent questionnaires and video classifiers to build a combination model. This model was validated by using a clinical sample of 230 autistic children from autism centers in the United States. The predictive accuracy of the new model was assessed using ROC curves. The study concluded that using Machine Learning techniques along with traditional autistic screening methods significantly improved the effectiveness of clinical screeners.

3. METHODOLOGY

The proposed methodology is depicted in Figure 2 in which different structures of neural networks are constructed from the input data. The derived classifiers are integrated to create a global classifier that consists of an ensemble of outputs. When a test instance requires a target class, the

prediction decision of assigning the class value to test data becomes a collective decision based on a voting mechanism. This reduces any possibility of a biased decision as the class assignment procedure considers all ANN structures built. In the next sub-sections, we explain the details behind each component of the proposed methodology.

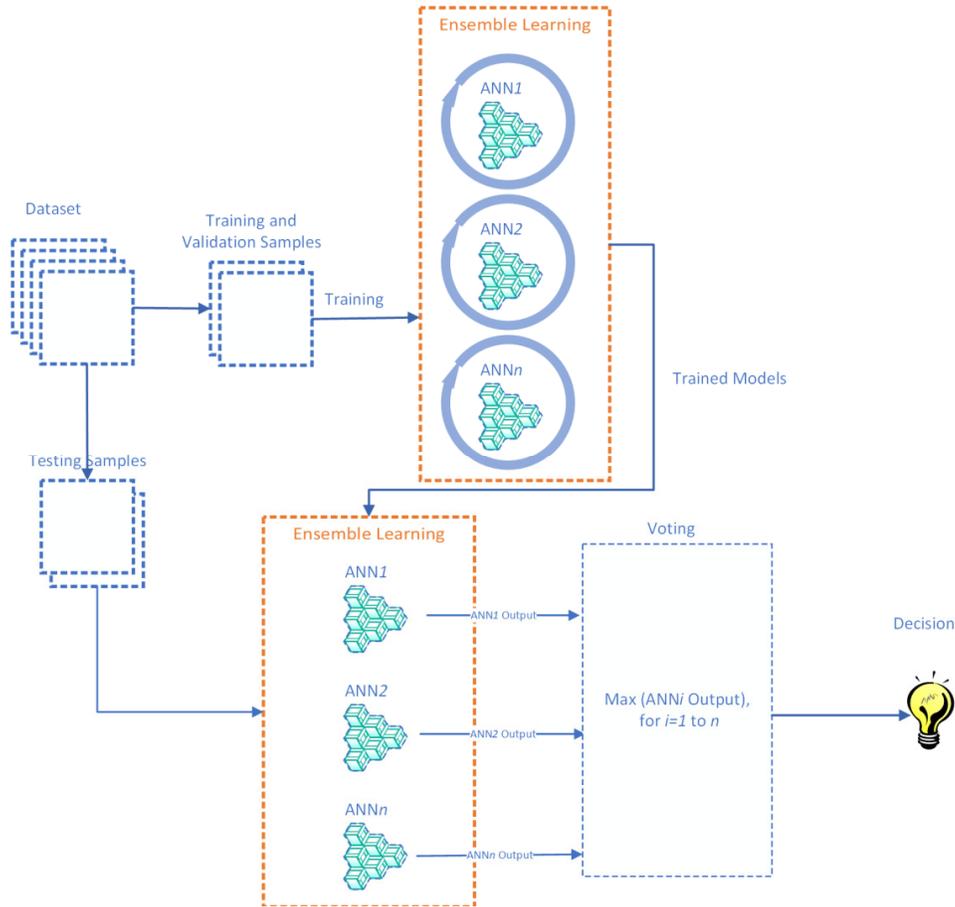


Fig. 2a: Proposed Methodology of ASD Screening

3.1 TRAINING PHASE

The mechanism of creating a new classifier in ECAS architecture is based on Bagging. The steps of the ANN algorithm used to build the local classifiers is shown in Figure 2a. Initially different parameters are set including number of neurons in the input layer and out layer respectively, connection weights, learning ratio, number of hidden layers, and maximum number of epochs. Details on the parameter's values are given in Section 4. The input dataset is split into training, validation, and testing sets. The training set will be employed to build the classifier and the validation set is utilized to validate the performance of the learned classifier from the training dataset samples. Finally, the testing set is used to evaluate the final classifiers and to measure their effectiveness in predicting test instances.

In learning the local classifiers (Figure 2b), we adopt an ANN approach in which the learning algorithm employs computed error rate and desired error rate to decide whether to update the structure of the classifier or to move ahead and generate the classifier. In this context, the learning algorithm determines the desired error rate to be accomplished in the upcoming epoch based on the computed error rate, therefore, it uses one sample of the dataset (randomly produced) to derive the computed error rate. When this computed rate is smaller than the desired error rate to be obtained before attaining the maximum allowed numbers of epochs, then the current classifier (network structure) can be enhanced. This ensures that the desired error rate is updated and set to the value of the current computed error rate. However, the algorithm will add a new hidden layer and continue the learning process when the desired error rate is not yet achieved. We used the Mean Square Error (MSE) (Equation 1) for the computed error rate.

$$MSE = \frac{1}{n} \sum_{n=0}^n (p_n - t)^2 \quad (1)$$

Where p_n is the forecasted value for instance n ; and t_n is the actual value of instance n in the training dataset.

Each training pass (epoch) initiates setting the learning rate to a value equivalent to the computed error rate obtained in the previous training pass, and therefore a dynamic amendment process of the learning rate each epoch. This methodology ensures improvement in performance and classifiers obtained are competitive.

Upon the completion of a training pass, the learning algorithm evaluates the computed error rates obtained at two consecutive time frames (T and T-1) to observe any error rate increase or decrease and based on this information, the learning rate can be adjusted. In particular there are two scenarios:

- a) When the error rate at time T is lower than that of Time T-1. In this case, the learning rate will be increased and the algorithm may converge to a potential good network structure
- b) When the error rate at time T is larger than that of Time T-1. In this case, the learning rate will be decreased and the algorithm may investigate the possible network structure obtained before moving forward to seek any potential improvement.

This evaluation of computed error rates at different times in the training pass, and the amendment of the learning rate, will improve the quality of the training and building of the classifier processes; thus the learning algorithm quickly updates the weights and efficiently derives a network structure. Overall, the network learning algorithms always seek when the condition of computed error rate has been reached or the maximum number of epochs is met.

To evaluate the ANN classifiers obtained from the training sets, a validation set is used which measures the effectiveness of the obtained computed error rate from the training dataset sample. Therefore, after a training pass, the learning algorithm computes the error rate on the validation set. When the error on the validation data set starts to show increasing patterns, this is a sign of overfitting and thus the learning for that model should be stopped. However, to avoid local minimum point (terminating the learning too early), the ANN algorithm keeps the error rate obtained from the validation set. When the last obtained error rate is smaller than the minimum recorded error rate, this can be considered a good sign as the algorithm results can be generalized.

As soon as a new individual is about to be classified for the possibility of autistic traits, each ANN classifier decides a class then all decisions are utilized to finalize the outcome using voting. Note that we don't have the concept of "model selection" in Deep Learning the same context that we have it in traditional Machine Learning. Deep Learning is the art of identifying what types of neurons we require, assembling them, and then selecting the hyperparameters for each layer. Given a robust representation of knowledge by selecting the proper types of neurons and hyperparameters in the previous layers, the number of dense neurons or learning rate is not significant to handle. For dense layers, we usually go with small numbers of neurons, and increase these gradually until desirable performance is achieved, considering overfitting is monitored as well. For learning rate, it's usually 0.001 when we start the training, and then a smaller learning rate is chosen to further smooth the classifier. It also depends on how much time and resources we must train the DNN.

4. DATA DESCRIPTION

Input

Samples of Independent and Dependent Variables as *Dataset*

Output

A set of classifier models trained on dataset as *Trained Models*

Initialize

Divide the Dataset into *Training, Verification, and Testing Samples*

Training

1. Create n number of *ANN Ensembles* with different hyperparameters, loss functions, etc.
2. Train the *ANNs* on *Training Samples*
3. Use *Verification Samples* to check of *ANNs'* training performance and overfitting
4. Adjust the hyperparameters if necessary
5. Repeat 2-4 until the networks achieve the desired performance
6. Return the *Trained Models*

Testing

1. Feed the Testing Samples to the *ANN Ensembles*
2. Collect the *ANN Outputs*
3. Apply a *voting function* (such as *max*) on the *ANN Outputs*
4. The *voting function* outputs the ensemble decision

Fig. 2b: Learning Algorithm Steps of ECAS

The dataset in the study was collected through a mobile application called ASDTests^{17,18} which implemented a screening method called AQ-10-Child³⁸ to collect primary data related to autism screening. The ASDTests application contains ten questions related to behavioral traits besides other features such as age, gender, ethnicity, and family history of autism. The participants undergo autism screening using the ASDTests app and, at the end of the screening, a score is calculated for the individual based on the answers and using the AQ-10-child scoring function. The score is utilized to screen ASD in which if the score obtained is greater than 6, this indicates that the child should take a further clinical assessment as he potentially exhibits autistic traits. Often parents, caregivers, teachers, or medical professionals take the test on behalf of the child. The dataset features are depicted in Table 1 in which features A1 - A10 are answers to the AQ-10-child questions. It should be noted that each question in the AQ-10-child (Questions

correspond to A1-A10 features) has four possible answers: “Definitely Agree”, “Slightly Agree”, “Definitely Disagree”, “Slightly Disagree”.³⁸

Table 1: The Dataset Features and Their Description

| Feature | Description |
|------------------------|--|
| A1 | Answer to Question 1 in AQ-10-child screening method; Child usually notices small sounds when other do not |
| A2 | Answer to Question 2 in AQ -10-child screening method; Child concentrates less on the small details of picture than the picture as a whole |
| A3 | Answer to Question 3 in AQ -10-child screening method; It is easy for the child to keep account of conversations from many different people in a social group |
| A4 | Answer to Question 4 in AQ -10-child screening method; It is easy for the child to switch between activities and go back to the previous one |
| A5 | Answer to Question 5 in AQ -10-child screening method; Child finds it difficult to continue a conversation with his/her peers |
| A6 | Answer to Question 6 in AQ -10-child screening method; Child is fine with social chit-chat |
| A7 | Answer to Question 7 in AQ-10-child screening method; It is not easy for the child to figure out the intentions or feeling of the characters in a story they read |
| A8 | Answer to Question 8 in AQ-10-child screening method; At preschool, child loves playing pretending games |
| A9 | Answer to Question 9 in AQ-10-child screening method; It is easy for them to understand what someone is feeling or thinking by just looking at their face |
| A10 | Answer to Question 10 in AQ-10-child screening method; They find it difficult to make new friends |
| Age | Age of the child under study (4-11 years) |
| Sex | Biological gender of the child (Male or Female) |
| Ethnicity | Which ethnic group the child belongs to |
| Jaundice | Whether or not the child was born with jaundice |
| Family_ASD | Whether or not there is family history of ASD |
| Who completed the test | The person who took the test (Medical staff, parent, caregiver, etc) |
| Class | ASD screening based on the score derived from the test (Yes = Scores above 6; No = Scores less than or equal 6) |

The dataset was collected by conducting tests on 1104 children (aged between 4 and 11 years) to screen for the presence of autistic traits at an early stage. The score obtained per test was used to

Table 2: Sample of the Five instances

| A 1 | A 2 | A 3 | A 4 | A 5 | A 6 | A 7 | A 8 | A 9 | A 10 | Age | Sex | Ethnicity | Jaundice | Family ASD | Who completed the test | Class |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|----------------|----------|------------|------------------------|-------|
| 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 4 | m | Middle Eastern | yes | no | family member | NO |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 4 | m | Middle Eastern | yes | no | family member | YES |
| 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 5 | m | White European | no | no | family member | YES |
| 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 4 | m | Asian | yes | no | family member | NO |
| 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 5 | m | White European | no | no | family member | NO |

detect early autism symptoms in a child; a class attribute was created (last column of Table 2) during the screening. Any child having a total score of less than or equal to 6 out of 10 was assigned “No ASD Traits”; otherwise the class attribute was assigned “Yes Autistic Traits”. Table 2 shows a sample of five instances from the dataset. The binary values assigned to features A1-A10 (“0” or “1”) have been transformed based on the actual answers given to these questions during the screening process. In particular, for questions 1,5,7 and 10, if the participant responded with “Definitely Agree” or “Slightly Agree” we assigned “1”; otherwise “0”. For questions 2,3,4,6,7,8 and 9, if the participant’s answer was either “Definitely Disagree” or “Slightly Disagree” then “1” will be assigned; otherwise “0”. The transformation process to binary representation of the original answers will ease the process of data processing.

We performed a descriptive analysis on the dataset to reveal possible useful patterns. The child dataset is balanced with respect to the class label in which there are 575 instances without autistic traits and 529 instances with autistic traits. There are 740 male instances and 364 females. Figure 3a depicts the number of male and female instances with and without autistic traits. The number of male instances that exhibit autistic traits is more than that of the female; this supports recent research as more male are on the ASD spectrum. Figure 3b shows the distribution of two age categories with respect to autistic traits. We divided the instances into two age groups; Group 1: 4-7 years and Group 2: 8-11 years. The dataset is dominated by instances of Group 1 (715 instances) which shows that more parents are concerned about autistic traits early. Based on the collected dataset, there were 336 children with autistic traits from Group 1 and 193 children from Group 2.

Among the 1104 instances, there are 273 instances born with jaundice and 831 born without jaundice. Figure 3c depicts the distribution of instances with and without ASD traits that have been born with and without jaundice. As seen in the figure, 51.28% of individuals born with jaundice were identified with autistic traits, whereas 53.19% children born without jaundice were identified without any autistic traits. More specifically, among the 273 instances born with jaundice, there was 140 who exhibited autistic traits and 133 without any traits. Among the 831 instances born without jaundice, 389 exhibited autistic traits and 442 without autistic traits. Lastly, for all cases that exhibited autistic traits (529 children), 140 of them were born with jaundice, and from 575 instances without autistic traits, 133 were born with jaundice. All these figures show there is no clear indication that jaundice is an influential feature of the possibility of having autistic traits, and hence further study is recommended on the impact of this feature.

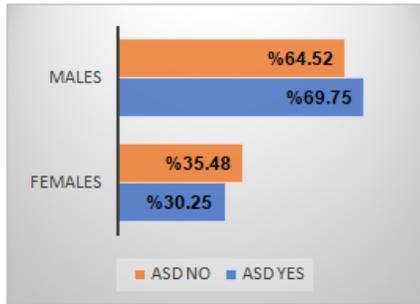


Fig. 3a: Numbers of male and female instances in the dataset with respect to autistic traits

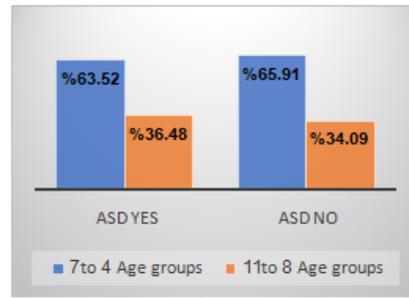


Fig. 3b: Age groups frequency in the dataset with respect to autistic traits



Fig. 3c: Numbers of instances with and without jaundice in the dataset and with respect to autistic traits

5. RESULTS ANALYSIS

5.1 EXPERIMENTAL SETTINGS AND EVALUATION MEASURES

Several experiments were conducted to analyze the children dataset using different Machine Learning algorithms and with respect to different evaluation metrics. The aim of this experiment was to reveal whether ECAS can do the screening process by detecting autistic traits as accurately as possible when compared with common Machine Learning techniques. The platform used to run the experiment is WEKA.³¹ WEKA is a Machine Learning tool that contains large collections of intelligent data mining and Machine Learning techniques for descriptive and predictive tasks. All compared algorithms are implemented in WEKA. The evaluation metrics used in the experiments are error rate, sensitivity, and specificity. These measures can be derived from the confusion matrix, a contingency table (Table 3) that shows True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) for a binary or multi-class classification problem such as for autism screening. TP, FP, FN, and TN are defined below:

- True Positive (TP): Number of children that are forecasted to have autistic traits and they do have autistic traits.
- False Positive (FP): Number of children that are forecasted to have autistic traits and they do not have autistic traits.

Table 3: Confusion Matrix for ASD Screening

| | Predicted Class Value | |
|--------------------|-----------------------|---------------------|
| Actual Class Value | ASD | No-ASD |
| ASD | True Positive (TP) | False Negative (FN) |
| No-ASD | False Positive (FP) | True Negative (TN) |

- False Negative (FN): Number of children that are forecasted not to have autistic traits and they do have autistic traits.
- True Negative (TN): Number of children that are forecasted not to have autistic traits and they do not have autistic traits.

Below are the evaluation metrics used in the results analysis.

1) *Error rate: It is the proportion of incorrect classifications.*

$$\text{Error rate} = 1 - \frac{t_p + t_n}{t_p + t_n + f_p + f_n}$$

2) *Sensitivity: It is the True positive rate. It is the average probability of complete retrieval.*

$$\text{Sensitivity} = \frac{t_p}{t_p + f_n}$$

3) *Specificity: It is the True negative rate.*

$$\text{Specificity} = \frac{t_n}{t_n + f_p}$$

The intelligent algorithms considered in the comparison with ECAS to assess screening performance are C4.5 (J48), PART, Repeated Incremental Pruning to Produce Error Reduction (Ripper), Voted Perceptron and Ensemble ANN.^{38,39-41} These algorithms are chosen because they adopt distinct learning approaches in deriving the classifiers. C4.5 is a decision tree algorithm that employs the concept of information gain to build decision trees. In doing so, each node of the decision tree represents an attribute and the one with the largest information gain is chosen as a root node. Each path from the root node to the leaf denotes a rule.³² Moreover, PART works on combining partial decision tree and converts them into rules after applying certain pruning procedures to reduce error.³⁹ RIPPER is rule-based learning algorithm that employs the reduced error pruning method to trim unnecessary items during the process of constructing rules.⁴⁰ The algorithm builds a set of rules to minimize the number of misclassifications by using a validation set. Lastly, the Voted Perceptron algorithm uses a classification method that involves linearly separable data having large margins by combining the perceptron algorithm and online learning to batch learning algorithm.⁴¹ Ten-fold cross-validation was used to train the classifiers.^{42,43} The choice of these algorithms was based on their common use in practical applications and the different learning methods employed during the training process.^{44,45-46}

Below are the main parameters for learning algorithms:

1) ECAS Ensemble Learner

- L 0.3** (Learning rate for the backpropagation algorithm)
- M 0.2** (Momentum rate for the backpropagation algorithm)
- N 500** (Number of epochs to train through)
- V 0** (Percentage size of validation set to use to terminate training)
- S 0** (The value used to seed the random number generator)
- E 20** (The number of consecutive increases of error allowed for validation testing before training terminates)
- H a** (The hidden layers to be created for the network)

2) Voted Perceptron

- I 1** (The number of iterations to be performed)
- E 1.0** (The exponent for the polynomial kernel)
- S 1** (The seed for the random number generation)
- M 10000** (The maximum number of alterations allowed)

3) C4.5

- C 0.25** (Set confidence threshold for pruning)
- M 2** (Set minimum number of instances per leaf)

4) PART

- M 2** (Set minimum number of objects per leaf)
- C 0.25** (Set confidence threshold for pruning)
- Q 1** (Seed for random data shuffling)

5) RIPPER

- F 3** (Set number of folds for REP. One-fold is used as pruning set)
- N 2.0** (Set the minimal weights of instances within a split)
- O 2** (Set the number of runs of optimizations)
- S 1** (The seed of randomization)

5.2 EXPERIMENTAL ANALYSIS

The error rate results by percentage, obtained after obtaining the ASD screening classifiers of all considered Machine Learning algorithms on the child dataset, are depicted in Figure 4. The Ensemble Learner of ECAS shows ultimate superiority in deriving ASD screening classifiers when contrasted with all considered Machine Learning algorithms. In particular, the Ensemble Learner produced a classifier close to nominal classification error. The Ensemble Learner of ECAS derived classifiers with 9.93%, 6.59%, 8.95%, and 12.21% less error rate than C4.5, PART, RIPPER and Voted Perceptron algorithms respectively. One principle reason of achieving this low error rate by the Ensemble learner is the ability of this learning algorithm to continuously amend the structure of the classifier during the training phase to a performance level desired by the domain expert yet not reaching overfitting. Among the contrasted algorithms, the Voted Perceptron algorithm produced classifiers with the highest error rate of 12.86% and C4.5, RIPPER and PART algorithms generated classifiers 10%, 9.6%, and 7.24% respectively. Overall, the performances of the Machine Learning algorithms regarding error rate were all over 80% which is the acceptable level in applied behavioral science domains.

Figure 5 depicts the sensitivity and specificity rates derived by the different Machine Learning algorithms including the Ensemble Learner on the child dataset. The Ensemble Learner-derived classifiers with a 100% sensitivity and Spcificity show no misclassified instances using tenfold cross-validation. Voted Perceptron, C4.5, PART, and RIPPER algorithms produced classifiers with a good Sensitivity rate nearly or above 90%. For example, Voted Perceptron had 53 instances wrongly classified as No-ASD (FN) out of 529 instances. The same algorithm derived the highest error rate due to low specificity rate as it has the largest FP instances when compared with the considered algorithms.

Based on the confusion matrix (Table5), out of 575 positive instances, 89 instances were misclassified to be on the spectrum by the Voted Perceptron algorithm. In the case of the RIPPER and PART algorithms, similar results were observed in terms of sensitivity and specificity rates. When we look at FN for PART and RIPPER, 31 and 40 instances were misclassified as No-ASD out of 529 positive instances respectively resulting in a higher sensitivity. Considering the Specificity for the PART and RIPPER algorithms, a lower result was generated by RIPPER owing to the 66 wrongly classified instances (FP) when they actually belong to No-ASD group. Looking at C4.5 results, the sensitivity and specificity rates are quite close to 90%. Considering that there are 529 positive instances and only 54 instances were classified as No-ASD or FN, and 63 instances out 575 negative instances were predicted in the ASD class.

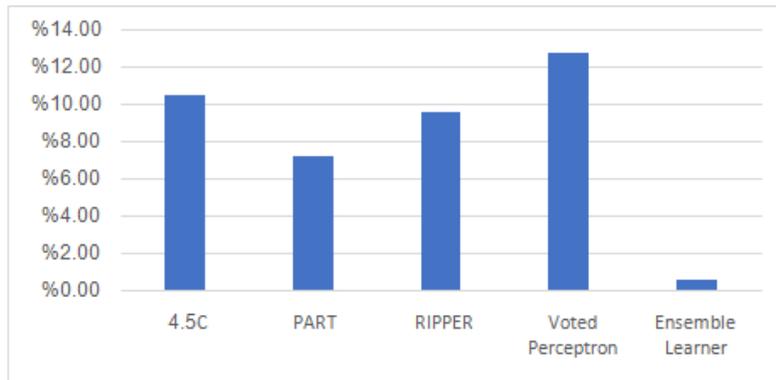


Fig. 4: Error Rate Derived by Machine Learning for ASD Screening

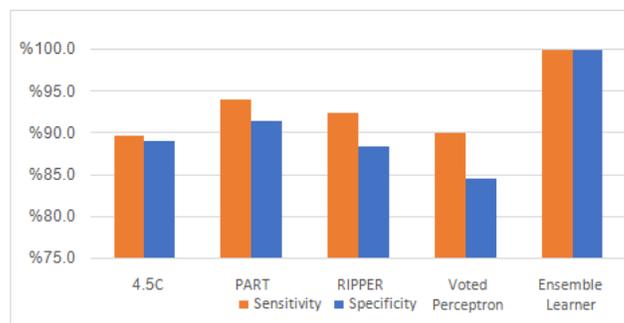


Fig. 5: Sensitivity and Specificity Figures Derived by Machine Learning Techniques for ASD Screening on the Child Dataset

Table 4: Confusion Matrix Results

| Machine Learning Technique | Actual Class Value | Predicted Class Value | |
|----------------------------|--------------------|-----------------------|--------|
| | | ASD | No-ASD |
| C4.5 | ASD | 475 | 54 |
| | No-ASD | 63 | 512 |
| PART | ASD | 498 | 31 |
| | No-ASD | 49 | 526 |
| RIPPER | ASD | 489 | 40 |
| | No-ASD | 66 | 509 |
| Voted Perceptron | ASD | 476 | 53 |
| | No-ASD | 89 | 486 |
| Ensemble Learner of ECAS | ASD | 529 | 0 |
| | No-ASD | 0 | 575 |

6. CONCLUSIONS

One of the key issues in autism screening is that the validity and accuracy of the screening process rely on the items designed in the screening method and the scoring function used to compute the final score by domain experts. This research investigated these issues by developing a new ASD screening framework based on Ensembles Learning to improve the performance of the screening process outcome with respect to accuracy, specificity, and sensitivity measures. The new proposed framework is called ECAS and it incorporates a novel Ensemble Learner to construct multiple classifiers from cases and controls related to children; it then employs a voting mechanism to produce the prediction in a collective manner. This not only improves the classification performance of detecting ASD traits, but also reduces biased decisions. Experimental tests on a real dataset related to children with 1104 instances (cases and controls) report that the ECAS reduces the numbers of misclassified instances for both children that exhibit ASD and children without ASD traits. More importantly, sensitivity and specificity results produced show that the classification method of ECAS (Ensemble Learner) was able to derive more accurate classifiers than decision trees (C4.5), PART, ANN with Voted Perceptron) and rule induction (RIPPER) techniques. This can be attributed to the ability of this learning method to amend the classifier during the learning process. All Machine Learning techniques showed improvements in ASD screening when compared with conventional screening approaches and maintained acceptable levels of performance in terms of sensitivity, accuracy, and specificity rates.

The development of a new autism screening method may encourage the adoption of this powerful technology in clinical environments for autism. This not only provides enhancement to the validity, efficiency and accuracy of the diagnostic process, but also empowers clinicians with distinctive tools that hold useful knowledge for better decision-making. In the near future, we will investigate how machine learning, particularly Ensemble Learners, can be embedded within a formal diagnostic tool in autism.

REFERENCES

- [1] Pennington, M. L., Cullinan, D., & Southern, L. B. (2014). Defining autism: variability in state education agency definitions of and evaluations for Autism Spectrum Disorders. *Autism Research and Treatment*, 1-8.
- [2] Thabtah, F. (2018A) Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Informatics for Health and Social Care* 43 (2), 1-20.
- [3] Chu, K. C., Huang, H. J., & Huang, Y. S. (2016). Machine learning approach for distinction of ADHD and OSA. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on* (pp. 1044-1049). IEEE.
- [4] Lopez Marcano, J. L. (2016). Classification of ADHD and non-ADHD Using AR Models and Machine Learning Algorithms (Doctoral dissertation, Virginia Tech).
- [5] Duda M., Ma R., Haber N., Wall D.P. (2016). Use of machine learning for behavioral distinction of autism and ADHD. *Translational Psychiatry* 9(6), 732.
- [6] Bone, D., Goodwin, M. S., Black, M. P., Lee, C.-C., Audhkhasi, K., & Narayanan, S. (2016). Applying machine learning to facilitate autism diagnostics: pitfalls and promises. *Journal of Autism and Developmental Disorders*, 1121–1136.
- [7] Thabtah F., Kamalov F., Rajab K. (2018) A new computational intelligence approach to detect autistic features for autism screening. *International Journal of Medical Informatics*, Volume 117, pp. 112-124.
- [8] Abbas, H., Garberson, F., Glover, E., & Wall, D. P. (2018). Machine learning approach for early detection of autism by combining questionnaire and home video screening. *Journal of the American Medical Informatics Association*, 25(8), 1000-1007. doi:10.1093/jamia/ocy039
- [9] Altay, O., & Ulas, M. (2018). Prediction of the Autism Spectrum Disorder Diagnosis with Linear Discriminant Analysis Classifier and K-Nearest Neighbor in Children. 2018 6th International Symposium on Digital Forensic and Security (ISDFS). Antalya, Turkey: IEEE. doi:10.1109/ISDFS.2018.8355354
- [10] Ravindranath, V., & Ra, S. (2018). A machine learning based approach to classify Autism with optimum behaviour sets. *International Journal of Engineering and Technology*. doi:10.14419/ijet.v7i3.18.14907
- [11] Thabtah F., Peebles D. (2019) A new machine learning model based on induction of rules for autism detection. *Health Informatics Journal*, 1460458218824711.
- [12] R. M. Mohammad, F. Thabtah and L. McCluskey, "Predicting Phishing Websites using Neural Network trained with Back-Propagation," in ICAI, Las Vigas, 2013-C.
- [13] R. M. Mohammad, F. Thabtah and L. McCluskey, "Predicting phishing websites based on self-structuring neural network," *Neural Computing and Applications*, vol. 25, no. 2, pp. 443-458, 2013-B.
- [14] S. Madhusmita, S. K. Dash, S. Dash and A. Mohapatra, "An approach for iris plant classification using neural network," *International Journal on Soft Computing*, vol. 3, no. 1, 2012.
- [15] F. Amato, A. López, E. M. Peña-Méndez, P. Vañhara, A. Hampl and J. Havel, "Artificial neural networks in medical diagnosis," *Journal of Applied Biomedicine*, vol. 11, no. 2, p. 47–58, 2013.

- [16] M. Riley, J. Karl and T. Chris, "A Study of Early Stopping, Ensembling, and Patchworking for Cascade Correlation Neural Networks," IAENG International Journal of Applied Mathematics, vol. 40, no. 4, pp. 307-316, 2010.
- [17] Thabtah, F. (2018). An accessible and efficient autism screening method for behavioural data and predictive analyses. *Health Informatics Journal*, 1-17. doi:10.1177/1460458218796636
- [18] Thabtah F., ASDTests. A mobile App for ASD Screening, (2017) (Accessed 14 March 2019), www.asdtests.com.
- [19] Ventola, P., Kleinman, J., Pandey, J., Barton, M., Allen, S., Green, J., . . . Fein, D. (2006). Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders*, 839-47.
- [20] Vllasaliu, L., Jensen, K., Hoss, S., Landenberger, M., Menze, M., Schutz, M., . . . Freitag, C. M. (2016). Diagnostic instruments for autism spectrum disorder (ASD). *Cochrane Database of Systematic Reviews*, 1-27.
- [21] Thabtah F. (2017) Autism Spectrum Disorder Tools: Machin Learning Adaptation and DSM-5 Fulfillment: An Investigative Study. Proceedings of the 2017 International Conference on Medical and Health Informatics (ICMHI 2017), pp. 1-6. Taichung, Taiwan. ACM.
- [22] Baron-Cohen, S. (2001). Take the AQ test. *Journal of Autism and developmental disorders*, 5-17.
- [23] Allison, C., Baron-Cohen, S., Charman, T., Wheelwright, S., Richler, J., Pasco, G., & Brayne, C. (2008). The Q-CHAT (quantitative checklist for autism in toddlers): a normally distributed quantitative measure of autistic traits at 18–24 months of age: preliminary report. *Journal of Autism and Developmental Disorders*, 1414–1425.
- [24] Witten, I. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*.
- [25] Freund, Y. and Schapire, R.E., (1997) A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), p.119–139.
- [26] Quinlan, J. (1986). Induction of Decision Trees. *Mach. Learn.* 1(1): 81-106.
- [27] Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017, March). "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis". *Autism Research*, 10, 384-407. doi:<https://doi.org/10.1002/aur.1678>
- [28] Tariq, Q., Daniels, J., Schwartz, J. N., Washington, P., Kalantarian, H., & Wall, D. P. (2018, November 27). Mobile detection of autism through machine learning on home video: A development and prospective validation study. *PLoS Med*, 15(11). doi:<https://doi.org/10.1371/journal.pmed.1002705>
- [29] Satu, S., Sathi, F. F., Arifen, S., & Ali, H. (January 2019). Early Detection of Autism by Extracting Features: A Case Study in Bangladesh. *International Conference on Robotics, Electrical and Signal Processing Techniques*. Dhaka. Retrieved from https://www.researchgate.net/publication/330383730_Early_Detection_of_Autism_by_Extracting_Features_A_Case_Study_in_Bangladesh
- [30] Wong, V., Hui, L., Lee, W., Leung, L., Ho, P., Lau, W., . . . Chung, B. (2004). A modified screening tool for autism (Checklist for Autism in Toddlers [CHAT-23]) for Chinese children. *Pediatrics*, 166-76.

- [31] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
- [32] Quinlan, J. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.
- [33] Breiman L. (2001) Random forests. *Mach. Learning*, 45(1):5-32, 2001. 1300
- [34] Friedman, N., Geiger, D. and Goldszmidt, M. (1997) Bayesian Network Classifiers. *Machine Learning - Special issue on learning with probabilistic representations*, 29(2-3), pp.131-63.
- [35] Bi, X.-a., Wang, Y., Shu, Q., Sun, Q., & Xu, Q. (2018). Classification of Autism Spectrum Disorder Using Random Support Vector Machine Cluster. *Frontiers in genetics*, 9(18). doi:10.3389/fgene.2018.00018
- [36] Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., & Pickles, A. (2000). The Autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 205-223.
- [37] Schopler, E., Van Bourgondien, M. E., Wellman, J., & Love, S. R. (1980). Toward objective classification of childhood autism: childhood autism rating scale (cars). *Autism DevDisord*, 91-103.
- [38] Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief “red flags” for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(2), 202-212.
- [39] Frank, E., and, Witten, I. (1998) Generating accurate rule sets without global optimisation. *Proceedings of the Fifteenth International Conference on Machine Learning*, (p. . 144-151). Madison, Wisconsin.
- [40] Cohen, W. W. (1995). Fast effective rule induction. In *Machine Learning Proceedings 1995* (pp. 115-123). Morgan Kaufmann.
- [41] Freund, Y., & Schapire, R. E. (1999). Large margin classification using the perceptron algorithm. *Machine learning*, 37(3), 277-296.
- [42] Abdelhamid, N., Thabtah, F., and Abdel-jaber, H. (2017). Phishing detection: A recent intelligent machine learning comparison based on models content and features. 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 72-77. 2017/7/22, Beijing, China.
- [43] Abdelhamid N., Ayesha A., Thabtah F. (2013) Classification. *Proceedings of the International conference on AI '2013*, pp. 687-695. LV, USA. Associative Classification Mining for Website Phishing
- [44] Thabtah F., Hadi W., Abdelhamid N., Issa A. (2011) Prediction Phase in Associative Classification. *Journal of Knowledge Engineering and Software Engineering*. Volume: 21, Issue: 6(2011) pp. 855-876. WorldScinet.
- [45] Thabtah F., Mahmood Q., McCluskey L., Abdel-jaber H (2010). A new Classification based on Association Algorithm. *Journal of Information and Knowledge Management*, Vol 9, No. 1, pp. 55-64. World Scientific.
- [46] Thabtah F., Cowling P., and Peng Y. (2006): Multiple Label Classification Rules Approach. *Journal of Knowledge and Information System*. Volume 9:109-129. Springer.