Autism Screening using Deep Embedding Representation

Haishuai Wang[1]★, Li Li[2], Lianhua Chi[3], and Ziping Zhao[4]

[1] Fairfield University, Fairfield, CT 06825, USA
hwang@fairfield.edu
[2] Harvard Medical School, Boston, MA 02115, USA
Li_Li@hms.harvard.edu
[3] La Trobe University, Melbourne, Austrilia
L.Chi@latrobe.edu.au
[4] Tianjin Normal University, Tianjin, China
zhaoziping@tjnu.edu.cn

Abstract. Autism spectrum disorder (ASD) is a developmental disorder that affects communication and behavior. An early diagnosis of neurodevelopmental disorders can improve treatment and significantly decrease associated healthcare cost, which reveals an urgent need for the development of ASD screening. However, the data used for ASD screening is heterogenous and multi-source, resulting in existing screening tools for ASD screening are expensive, time-intensive and sometimes fall short in predictive accuracy. In this paper, we apply novel feature engineering and feature encoding techniques, along with a deep learning classifier for ASD screening. Algorithms were created via a robust deep learning classifier and deep embedding representation for categorical variables to diagnose ASD based on behavioral features and individual characteristics. The proposed algorithm is effective compared with baselines, achieving 99% sensitivity and 99% specificity. The results suggest that deep embedding representation learning is a reliable method for ASD screening.

Keywords: Autism Spectrum Disorder · Deep Learning · ASD Screening · Categorical Embedding.

1 Introduction

Autistic Spectrum Disorder (ASD) is a mental disorder characterized by difficulties with social interaction and communication, and by restricted and repetitive behavior [7]. In the United States, there are 3 million people who have ASD, and around 1 out of 68 children are diagnosed with ASD [24]. ASD is a neurodevelopment condition associated with significant healthcare costs, and early diagnosis can significantly reduce the cost and improve the quality of life of children with ASD. The increase in the number of ASD cases and the cost impact of ASD push forward the research of effective screening methods [15, 20, 14]. Most tools

---

★ corresponding author

for autism screening are based on score-sheets with questions for the parent or the medical practitioner [2], and the summation score is compared with predetermined thresholds to produce results. For example, the Modified Checklist for Autism in Toddlers (M-CHAT) [5] is a checklist-based screening tool for autism with children between the ages of 16 and 30 months. The Child Behavior Checklist (CBCL) [4] is a parent-completed screening tool. However, the diagnostic process for ASD is costly and time-consuming [12]. Recently, machine learning based approaches have been showing a great direction on objective evaluation of neuropsychiatric disorders [3, 26, 13]. Machine learning or deep learning based approaches might allow for detecting ASD diagnosis automatically and are able to provide a map of high-risk populations [19].

Due to the significance of ASD screening, we propose to develop, validate and assess efficient deep learning algorithms that identify ASD versus non-ASD. The ASD screening algorithm will be based on data consolidated from heterogeneous sources, including questionnaire and demographics. There are some existing forecasting algorithms/tools being used for ASD screening based on these kinds of data [22, 21, 23]. However, these approaches more focused on improving or applying machine learning algorithms whereas ignored to propose effective feature engineering methods for the heterogenous ASD data. For instance, the ASD data are usually a mix of numerical and categorical features, which pose a challenge for directly applying classifiers as they can only deal with numerical inputs by design. Nevertheless, most of the existing methods applied one-hot encoding to deal with categorical variables for the ASD data, and then input one-hot vectors (or dummy variables) to machine learning models or feed them into neural networks. Since one-hot encoding usually yields sparse vectors, it potentially limits model performance and screening accuracy. Although categorical data is very common in medical datasets, there are no existing works to effectively handle or represent categorical medical data. Thus, an effective feature encoding method is required to improve screening accuracy.

Nowadays, deep learning has been one of the most prominent machine learning techniques [25, 15], and it has the capability to model data with non-linear structures and learn a high-level representation of features. In this paper, we present deep embedding representation for categorical variables, along with neural network as a classifier for ASD screening. Specifically, we first learn the categorical feature representation from an embedding layer, then we combine the learned embeddings and continuous variables as input to a dense layer, followed by a non-linear activation layer. We use several fully connected layers on top of the embedding layer. We use DENN for short to denote the model name. The DENN model outperforms the existing methods on the ASD screening data.

2   Data and Analysis

2.1   Data Description and Data Exploration

The data we used is the Autism Screening Adult Data Set provided by the UCI Machine Learning Repository [8]. This is a new dataset related to autism screen-

Table 1. Basic descriptive statistics for the variables (we truncated the variable names of the A1 Score to A10 Score variables to A1 to A10).

|	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	age	result
mean	0.723	0.453	0.459	0.497	0.498	0.285	0.417	0.651	0.325	0.574	29.698	4.883
std	0.447	0.498	0.498	0.500	0.500	0.451	0.493	0.476	0.468	0.494	16.507	2.498
min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	17.000	0.000
max	1.000	1.000	1.000	1.00	1.00	1.000	1.000	1.000	1.000	1.000	383.00*	10.000

* Indicate a possible outlier value

ing of adults that contains 704 observations with 21 variables to be utilised for further analysis especially in determining influential autistic traits and improving the classification of ASD cases. The raw data contains ten binary variables
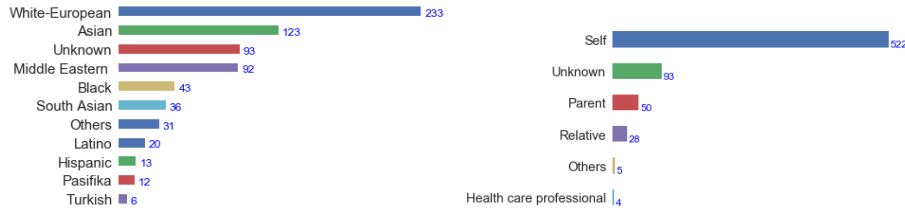
Fig. 1. Statistical distribution of ethnicity. **Fig. 2.** Statistical distribution of relation.
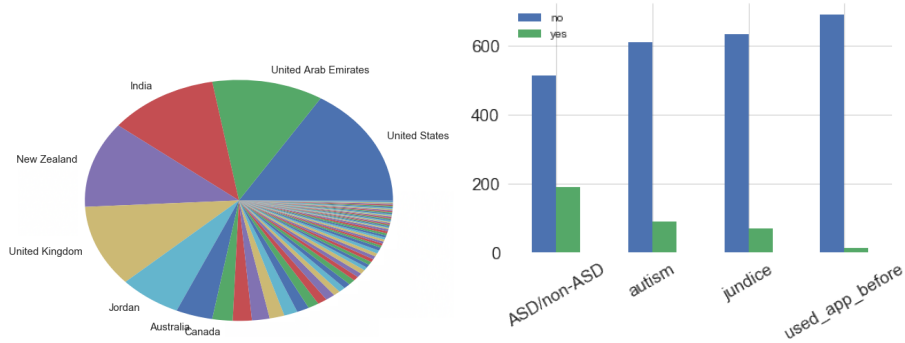
Fig. 3. Statistical distribution of country (only show names of major countries).

Fig. 4. Statistical distributions on three categorical variables (*autism, used_app_before, jaundice*), and the class label (*ASD/non-ASD*).
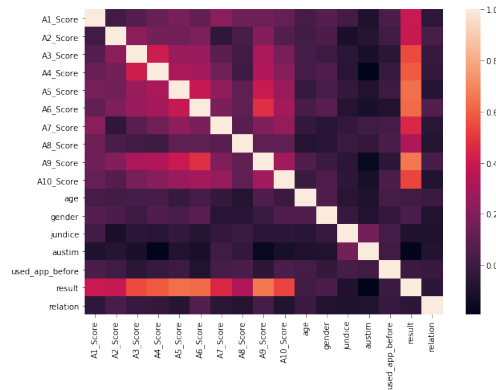
Fig. 5. Correlation heatmap for the dataset.

(AQ-10-Child) representing the screening questions, and the categorical variables of *gender, ethnicity, jaundice, autism, country_of_res, used_app_before, age_desc, relation* and *Class/ASD.* There are also two numeric variables named *age* and *result.* The 21 variables, representing behavioural features and individuals characteristics, that have proved to be effective in detecting the ASD cases from controls in behaviour science [22, 6, 23]. Some basic descriptive statistics for the variables are shown in Table 1 and Figures 1 - 4.

2.2   Data Cleaning and Preprocessing

After obtaining the statistic information of the variables from data exploration, we need to clean the data of unwanted information to prepare it as appropriate input features for our machine learning algorithms.

Cleaning missing values and outliers: In this data, there are missing values in the age variable. Since there are only two missing age values, we simply remove all of the observations containing the missing age values. As we can see from Table 1, there is an impossibly large maximum value of 383 in the age variable. Given that there are already many typographical errors present in the data, it is reasonable to assume that this is because of a typing error and the intended value should be 38. After preprocessing the missing values and outliers, we visualize the data to explore potential challenges and solutions for a learning model.

2.3   Data Visualization and Analysis

Before proceeding to apply any machine learning algorithm, we visualize and analyze the data to provide a better solution for ASD screening. In Figure 5, we compare the correlation among the variables to demonstrate how close two variables are to having a linear relationship with each other. From Figure 5, we can observe that there is a high correlation between AQ-10-Child score and
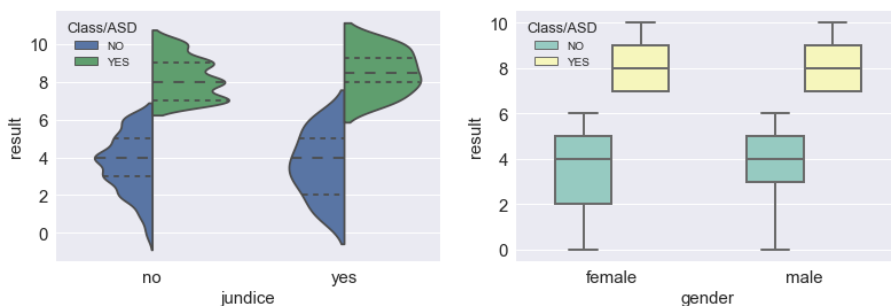
**Fig. 6.** The connection visualization between *jaundice* and *result*.
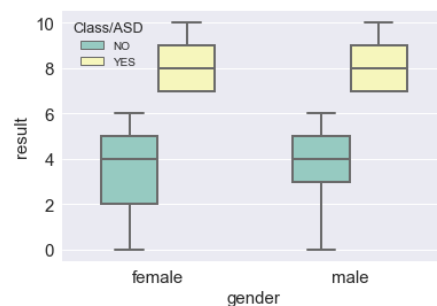
Fig. 7. The connection visualization between *gender* and *result*.

result, which indicates they have a linear relationship. However, other variables have a non-linear relationship (as shown in Figure 5). Therefore, a good machine learning model using these variables should consider both linear and non-linear relationships.

To obtain the first impression of the internal connections of some of the variables that are presented in the data set, we visualize the connection between several attributes and display how they are related with the target class in Figure 6 - 8. Figure 6 and 7 show a similar relationship between *gender* and *result* vs. *jaundice* and *result*. From both cases, the individual with a higher *result* score is more likely to have ASD, independent of the other variables. The variables *gender* and *jaundice* do not seem a major influence in deciding ASD. Nevertheless, from Figure 8, we can observe that when *jaundice* presents at birth, the individual with a higher *result* score will have ASD regardless of their gender. In Figure 9, we plot the data in terms of *age*, *gender*, *jaundice* and *relation* to visualize their

Fig. 8. The connection visualization among *jaundice*, *gender* and *result*.
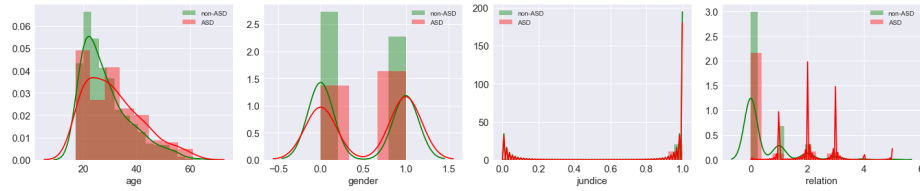
Fig. 9. Distribution plot for some of the input variables to machine learning models.

distribution on ASD class and non-ASD class. They have similar distribution on the data belongs to different class.

The purpose is to develop effective algorithms for ASD screening and diagnosis based on the data described above. However, from the above analysis, the variables in the dataset consist of two data types: continuous (e.g., age) and categorical (e.g., gender). Dealing with continuous numeric data is often easier than categorical data given that it can be fed into most of machine learning models after normalization. However, naively applying machine learning algorithms with integer representation for categorical variables does not work well. Since categorical variables are known to hide and mask lots of interesting information in a dataset and they might even be the most important variables in a model, we will present an advanced technique called deep embedding representation to deal with categorical variables in neural networks for ASD screening.

3   Prediction Methodology

We aim to use the dataset described in Section 2 to predict whether these patients actually have autism accurately. A main challenge to develop an algorithm is that the data contains both continuous numerical variables and categorical variables. Therefore, we first propose a deep categorical embedding method for feature engineering, and then present a neural network classifier using the embedded features for ASD screening. The model is denoted as DENN for short. The overall architecture of DENN is illustrated in Figure 10.

3.1   Deep Embedding Representation for Categorical Variables

For the continuous numerical data, we use identity mapping to map numerical values to feature vectors. Since neural network can only deal with numerical inputs by design, the categorical features can not be input to neural network directly. One-hot encoding is a commonly used method for converting a categorical variable into a continuous variable. By one-hot encoding, the new representation is a vector with one element being one and all others being zero, so it is also called dummy coding or the 1-of-c encoding where c is the number of total possible categories of a categorical feature. If we have P categorical features and the i-th feature can take c_i values, this encoding will result in dimensionality of d, such that,

$$d = \sum_{i=1}^{P} c_i \tag{1}$$

Although one-hot encoding is simple and common, it often yields a very sparse and high dimensional representation of the data, and it ignores the informative relations between them because it treats different values of categorical variables completely independent of each other. In this section, we present an advanced method that aims to capture the different categories with a much smaller dimension and represents the categorical features more efficiently.
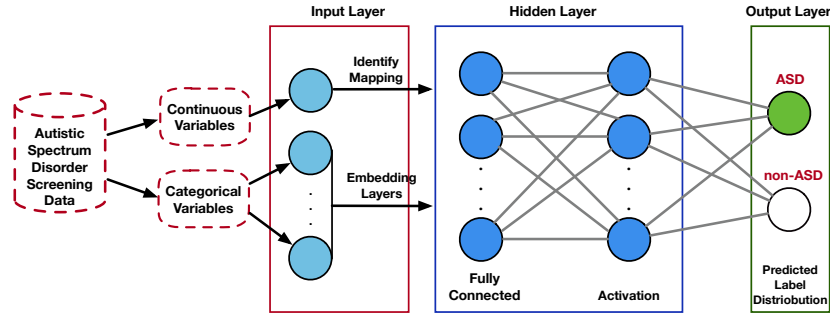

Fig. 10. Overall architecture of DENN.

We map categorical variables in a function approximation problem into Euclidean spaces, which is to build a vector embedding to every category type, such that $e_i : x_i \mapsto \mathbf{x}_i$. For each categorical variable, we initialise a $m \times D$ embedding matrix $\boldsymbol{E}$ as

where m is the number of total possible categories of a categorical variable, hyperparameter D is the desired dimension for embedding representation, which is usually less than m.

$$\boldsymbol{E} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \ddots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mD} \end{pmatrix} = (x_{ij}) \in \mathbb{R}^{m \times D} \tag{2}$$

Then we add an embedding layer in a deep neural network to do a lookup for a given value from the embedding matrix $\boldsymbol{E}$, which returns a vector $\mathbf{x}_k = (x_{kj}) \in \mathbb{R}^{1 \times D}$. The new representation $\mathbf{x}_k$ along with the numerical variables would then be fed into the next layer in the neural networks, and all the embeddings are updated and learned through backpropagation.

3.2   Neural Network Architecture for ASD Screening

As shown in Figure 10, the embedded categorical variables are concatenated with numerical features as new feature vectors that can be fed into a dense layer. There are several dense layers in our neural network architecture, and each dense layer followed by an activation layer. We used ReLu as the activation function to introduce non-linearity. After the ReLu activation, we have $f(z) = max(0, z)$ that gives an output z if z is positive and 0 otherwise.

Then, an output layer is placed after the last hidden layer. From hidden layer to output layer is a sigmoid function to output the predicted probability of a feature vector $\mathbf{x}$ belonging to class 1:

$$Pr(y = 1 | \mathbf{x}, \mathbf{w}_o) = \frac{1}{1 + exp(-\mathbf{w}_o^T \mathbf{x})} \tag{3}$$

where $\mathbf{w}_o$ is the weights of the output layer (the subscript o represents the parameters in the output layer).

Since this is a binary classification problem, we encoded the class label as one-hot vector and set categorical cross entropy as the loss function, as follows:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{K} (y_{i,t} * log(\hat{y}_{i,t})) \tag{4}$$

where N is the total number of training samples, $\hat{y}$ is the predicted label, and K is the total number of classes (K equals 2 for ASD screening). To learn and optimize the parameters of the model, we minimize the loss function $\mathcal{L}$. The loss minimization and parameter optimization can be performed through the backpropagation using mini-batch stochastic gradient descent.

4   Experiments and Evaluation

4.1   Datasets and Setup

We trained the proposed approach on the Autistic Spectrum Disorder Screening Data from the University of Irvine data collection [5]. After preprocessing, this dataset contains 702 samples, including 513 patients or children that have been screened for autism (i.e., 189 cases and 513 controls). We randomly selected 80% as training data and the remaining 20% as testing data.

Since our model is based on supervised learning, the input of the models utilizes a training dataset of cases (the number is 155) and controls (the number is 406) that have already been diagnosed. Usually, the cases and controls have been generated using a screening tool such as ADOS-R, ADI-R, etc., in a clinic by a behaviorist, clinical psychologist, or a licensed clinician specialized in that tool. In our experiment settings, we use two fully connected layers (1000 and 500 neurons respectively) on top of the embedding layer. The neural network is trained for 50 epochs.

---

[5] https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++

Table 2. Confusion Matrix for ASD screening.

		Predicted Class	
		ASD	non-ASD
Actual Class	ASD	True Positive (TP)	False Negative (FN)
	non-ASD	False Positive (FP)	True Negative (TN)

4.2 Evaluation Criteria

The experiments are designed to validate whether the ASD screening model, which combines deep embedding representation and neural networks, can achieve better performance than using one-hot encoding and shallow machine learning models. All the experiments were implemented in Python and Keras framework. The source codes used for this experiments and results can be found in this github repository [6].

Following the most common procedure for evaluating models for ASD screening, we use Sensitivity (Recall), Specificity (Precision), F-measure (F1-Score) Receives Operating Characteristic (ROC) Curve, and Area Under ROC Curve (AUC) to evaluate the ASD screening approaches. To compute the measurements, we can use a confusion matrix to summarize the performance *w.r.t.* various models, as shown in Table 2. For classification problem, if the sample is positive and it is classified as positive, it is counted as a true positive (Eq. 5); If the sample is negative while it is classified as positive, it is counted as a false positive (Eq. 6).

$$TP = \frac{Correctly\ classified\ positives\ samples}{Total\ no.\ of\ positives} \tag{5}$$

$$FP = \frac{Incorrectly\ classified\ negative\ samples}{Total\ no.\ of\ negative\ samples} \tag{6}$$

Following formulas are used to measure above mentioned performance measures are shown below.

$$Specificity = \frac{\sum_{n=1}^{N} TN_n}{\sum_{n=1}^{N} TN_n + FP_n}, \ Sensitivity = \frac{\sum_{n=1}^{N} TP_n}{\sum_{n=1}^{N} TP_n + FN_n} \tag{7}$$

$$F1 = 2 \times \frac{Specificity \times Sensitivity}{Specificity + Sensitivity} \tag{8}$$

Baselines. We also conducted comparison experiments with different methods, including Random Forest, Support Vector Machine (SVM), Gradient Boosting, and Neural Networks without embedding layer. These machine learning-

---

[6] https://github.com/BlindReview/ASD_Prediction

based baselines have been applied and reported [23, 9, 1, 19, 2, 11] for ASD screening based on the same dataset. The inputs for all of the baselines are one-hot-encoded features. The baselines are implemented using the scikit-learn library of python and default parameters.

4.3   Experimental Results

Table 3 and Figures 11-13 present the performance of ASD screening based on different machine learning models on the test dataset. In comparison to the baselines on the test dataset, we observed that the deep embedding representation model performs better than baselines in terms of specificity, sensitivity and f1 score. The baseline using neural network without embedding layer (denoted as NN for short) performs better than other baselines but worse than the DENN model. This is because the deep embedding representation is more efficient than one-hot encoding for categorical feature representation, and neural network fits better than shallow machine learning models on the autism dataset. As we can observe from the ROC curves in figure 11, under the same false positive rate, we are able to identify ASD with high true positive rate, which is better than that of in the baselines.
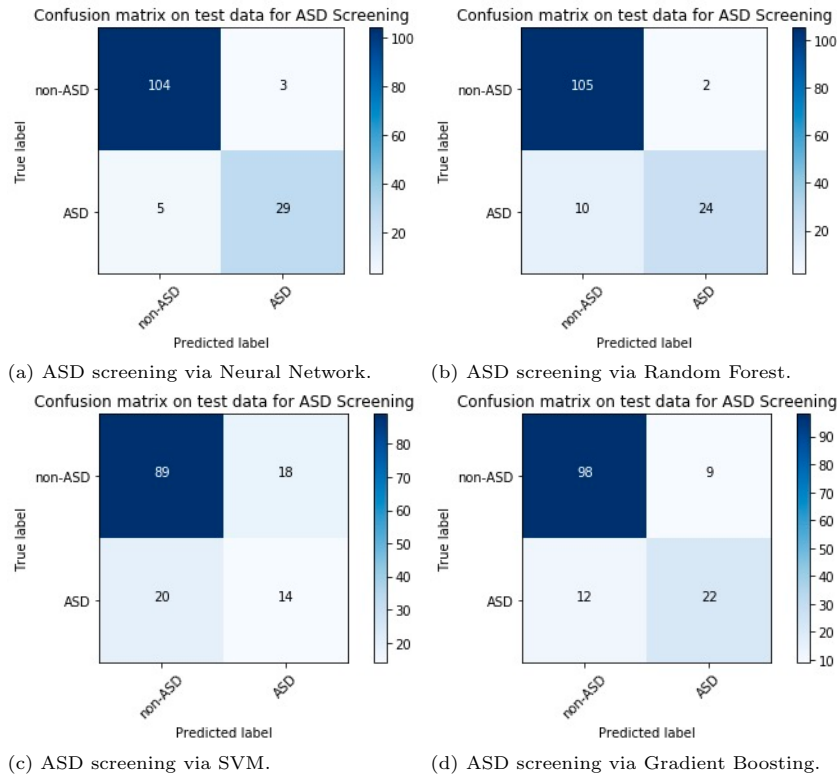
Table 3. Comparison of different methods *w.r.t.* specificity, sensitivity and F1-score.

Method	Specificity	Sensitivity	F1-Score
DENN	**0.99**	**0.99**	**0.99**
NN	0.94	0.94	0.94
Random Forest	0.92	0.91	0.91
SVM	0.73	0.73	0.73
Gradient Boosting	0.85	0.85	0.85

We visualize the confusion matrix of the DENN model on the test data in Figure 12. From Figure 12, we can observe that there is only one sample is misclassified on the test data. The deep embedding representation with neural networks is able to achieve almost 100% accuracy for ASD screening on the dataset. However, other baselines have higher misclassification rate for ASD group because there are fewer ASD samples in the dataset. The baselines are not able to learn enough information due to their limited learning capability. From the experimental results, we can conclude that effective categorical feature embedding method can help to improve the learning ability and model performance. The data has both linearity and nonlinearity nature, resulting in DENN achieves the best performance, followed by neural network based approach.
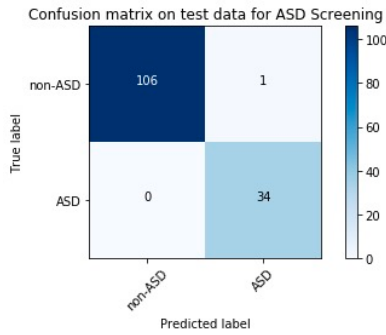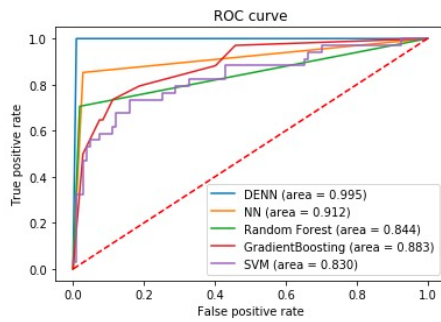
Fig. 11. ROC curve *w.r.t.* various ASD screening approaches.

Fig. 12. Confusion matrix of DENN model on the test data for ASD screening.

(a) ASD screening via Neural Network.

(b) ASD screening via Random Forest.

(c) ASD screening via SVM.

(d) ASD screening via Gradient Boosting.

Fig. 13. Confusion matrix on the test data *w.r.t.* various baseline methods.

5   Related Work

Autism spectrum disorder (ASD) is a developmental disorder, affecting about 1% of the global population [19]. ASD screening is crucial in helping a child

with autism live a more normal life in society. Diagnosing ASD typically takes two steps: Developmental Screening and Comprehensive Diagnostic Evaluation. Developmental Screening determines if children are learning basic skills at the right age or later. Doctors may recommend developmental tests to determine if cognitive, language, and social skills acquisition is delayed. Comprehensive evaluation is a thorough review that includes looking at the child's behavior and development and interviewing the parents. It may also include hearing and vision screening, genetic testing, neurological testing, and other medical testing.

To date, behavior-based tests are the standard clinical approach for ASD diagnosis [19, 7], hence, traditional tools for autism screening are based on score-sheets with questions for the parent or the medical practitioner [2], and the summation score is compared with predetermined thresholds to produce results. For example, the Modified Checklist for Autism in Toddlers (M-CHAT) [5] and the Child Behavior Checklist (CBCL) [4]. A number of other existing clinical diagnosis methods have also been used for ASD identification, such as Autism Diagnostic Interview-Revised (ADI-R) [18] and Autism Diagnostic Observation Schedule-Revised (ADOS-R) [17], have shown superior performance. Yet those methods rely on handcrafted rules that employ mathematical summation formulas of scores to come up with the appropriate diagnosis [23]. Moreover, the majority of existing ASD screening tools require substantial time to produce a complete diagnosis. Therefore, they are time-consuming and costly.

Several studies have recently employed machine learning and deep learning to improve the diagnosis process for ASD. Researchers have adopted several supervised machine learning techniques (such as Neural Network, Decision Trees, Logistic Regression, Random Forest, Support Vector Machines(SVM), k-Nearest Neighbors(kNN), Naive Bayes) to solve the classification problem of predicting whether an individual with certain characteristics has ASD [23, 9, 1, 19, 2]. The study [11] applied a machine learning algorithm to identify ASD from attention deficit hyperactivity disorder based on a 65-item Social Responsiveness Scale. Another study [10] combines the Social Responsiveness Scale with the ADI-R score to train their models to distinguish ASD from controls. More recently, studies [16, 15, 3] have developed machine learning models using the Autism Brain Imaging Data Exchange (ABIDE) towards the automated diagnosis of ASD based on brain neuroimaging data.

The main purposes of the existing machine learning-based ASD diagnostic tools were to improve diagnosis accuracy, and speed up diagnosis time to provide timely access to healthcare services. However, the existing machine learning or deep learning based approaches for ASD either rely on neuroimaging data or simply applying traditional learning algorithms without considering the latent feature characteristics. Therefore, new effective feature representation methods for machine learning based ASD diagnosis is significant to improve the diagnosis performance.

6   Conclusion

Autism spectrum disorder (ASD) is a developmental disability that can cause significant social, communication and behavioral challenges. ASD screening is significant because it enables early intervention. Early treatment is more effective than later treatment for ASD. In this paper, we are able to predict autism in patients with about 99% accuracy based on the UCI ASD screening data. Since there are plenty of health applications are going to have categorical data, we also learned how to deal with categorical data using deep embedding representation along with a neural network as a classifier for ASD screening. Compared with other models for ASD screening, the DENN model is more efficient and accurate for ASD screening.

References

1. Abbas, H., Garberson, F., Glover, E., Wall, D.P.: Machine learning for early detection of autism (and other conditions) using a parental questionnaire and home video screening. In: 2017 IEEE International Conference on Big Data (Big Data). pp. 3558–3561. IEEE (2017)
2. Abbas, H., Garberson, F., Glover, E., Wall, D.P.: 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 (2018)
3. Abraham, A., Milham, M.P., Di Martino, A., Craddock, R.C., Samaras, D., Thirion, B., Varoquaux, G.: Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example. NeuroImage **147**, 736–745 (2017)
4. Achenbach, T.M., Rescorla, L.A.: Manual for the ASEBA preschool forms and profiles, vol. 30. Burlington, VT: University of Vermont, Research center for children, youth . . . (2000)
5. Allely, C.S., Wilson, P.: Diagnosing autism spectrum disorders in primary care. The Practitioner **255**(1745), 27–31 (2011)
6. Altay, O., Ulas, M.: Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and k-nearest neighbor in children. In: 2018 6th International Symposium on Digital Forensic and Security (ISDFS). pp. 1–4. IEEE (2018)
7. Association, A.P., et al.: Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub (2013)
8. Asuncion, A., Newman, D.: Uci machine learning repository (2007)
9. van den Bekerom, B.: Using machine learning for detection of autism spectrum disorder (2017)
10. Bone, D., Goodwin, M.S., Black, M.P., Lee, C.C., Audhkhasi, K., Narayanan, S.: Applying machine learning to facilitate autism diagnostics: pitfalls and promises. Journal of autism and developmental disorders **45**(5), 1121–1136 (2015)
11. Duda, M., Ma, R., Haber, N., Wall, D.: Use of machine learning for behavioral distinction of autism and adhd. Translational psychiatry **6**(2), e732 (2016)
12. Galliver, M., Gowling, E., Farr, W., Gain, A., Male, I.: Cost of assessing a child for possible autism spectrum disorder? an observational study of current practice in child development centres in the uk. BMJ paediatrics open **1**(1) (2017)

13. Ghiassian, S., Greiner, R., Jin, P., Brown, M.R.: Using functional or structural magnetic resonance images and personal characteristic data to identify adhd and autism. PloS one **11**(12), e0166934 (2016)
14. Guo, X., Dominick, K.C., Minai, A.A., Li, H., Erickson, C.A., Lu, L.J.: Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method. Frontiers in neuroscience **11**, 460 (2017)
15. Heinsfeld, A.S., Franco, A.R., Craddock, R.C., Buchweitz, A., Meneguzzi, F.: Identification of autism spectrum disorder using deep learning and the abide dataset. NeuroImage: Clinical **17**, 16–23 (2018)
16. Li, H., Parikh, N.A., He, L.: A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes. Frontiers in neuroscience **12**, 491 (2018)
17. Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C., Pickles, A., Rutter, M.: 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 **30**(3), 205–223 (2000)
18. Lord, C., Rutter, M., Le Couteur, A.: Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of autism and developmental disorders **24**(5), 659–685 (1994)
19. Parikh, M.N., Li, H., He, L.: Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data. Frontiers in computational neuroscience **13** (2019)
20. Rad, N.M., Kia, S.M., Zarbo, C., van Laarhoven, T., Jurman, G., Venuti, P., Marchiori, E., Furlanello, C.: Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. Signal Processing **144**, 180–191 (2018)
21. Thabtah, F.: Asdtests. a mobile app for asd screening (2017)
22. Thabtah, F.: Autism spectrum disorder screening: machine learning adaptation and dsm-5 fulfillment. In: Proceedings of the 1st International Conference on Medical and Health Informatics 2017. pp. 1–6. ACM (2017)
23. Thabtah, F.: Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. Informatics for Health and Social Care pp. 1–20 (2018)
24. de la Torre-Ubieta, L., Won, H., Stein, J.L., Geschwind, D.H.: Advancing the understanding of autism disease mechanisms through genetics. Nature medicine **22**(4), 345 (2016)
25. Wang, H., Cui, Z., Chen, Y., Avidan, M., Abdallah, A.B., Kronzer, A.: Predicting hospital readmission via cost-sensitive deep learning. IEEE/ACM Transactions on Computational Biology and Bioinformatics (2018)
26. Yahata, N., Morimoto, J., Hashimoto, R., Lisi, G., Shibata, K., Kawakubo, Y., Kuwabara, H., Kuroda, M., Yamada, T., Megumi, F., et al.: A small number of abnormal brain connections predicts adult autism spectrum disorder. Nature communications **7**, 11254 (2016)