AUTISM CLASSIFICATION USING MACHINE LEARNING

Project Report submitted in partial fulfillment of
The requirements for the degree of

MASTER OF COMPUTER APPLICATION

Of

WEST BENGAL UNIVERSITY OF TECHNOLOGY

By

SUDIPTO PAUL, Roll no. - 29171020031
PHIROZE CHOWDHURY, Roll no. - 29171020046
SUPARNA DEBNATH, Roll no. - 29171020016

Under the guidance of
PROF. SUDIP DEY

DEPARTMENT OF COMPUTER APPLICATION

NETAJI SUBHASH ENGINEERING COLLEGE
TECHNO CITY, GARIA, KOLKATA – 700 152

2021-22

CERTIFICATE

This is to certify that this project report titled **AUTISM CLASSIFICATION USING MACHINE LEARNING** submitted in partial fulfillment of requirements for award of the degree **Master of Computer Application** of West **Bengal University of Technology** is a faithful record of the original work carried out by,

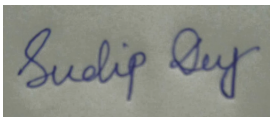
SUDIPTO PAUL, Roll no. 29171020031, **Regd. No.** 202910571010001 of 2020-21

PHIROZE CHOWDHURY, **Roll no**. 29171020046, **Regd. No.** 202910571010036 of 2020-21

SUPARNA DEBNATH, Roll no. 29171020016, **Regd. No.** 202910571010016 of 2020-21

Under my guidance and supervision

It is further certified that it contains no material, which to a substantial extent has been submitted for the award of any degree/diploma in any institute or has been published in any form, except the assistances drawn from other sources, for which due acknowledgement has been made.



Date: 13/01/2022

Guide's signature

PROF. SUDIP DEY
ASSISTANT PROFESSOR
DEPARTMENT OF COMPUTER APPLICATION,
NETAJI SUBHASH ENGINEERING COLLEGE,
TECHNO CITY, GARIA, KOLKATA – 700 152

_____

Head of the Department,
DEPARTMENT OF COMPUTER APPLICATION,
NETAJI SUBHASH ENGINEERING COLLEGE,
TECHNO CITY, GARIA, KOLKATA – 700 152

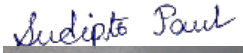
DECLARATION

We hereby declare that this project report titled

AUTISM CLASSIFICATION USING MACHINE LEARNING

is our own original work carried out as a under graduate student in **Netaji Subhash Engineering College** except to the extent that assistances from other sources are duly acknowledged.

All sources used for this project report have been fully and properly cited. It contains no material which to a substantial extent has been submitted for the award of any degree/diploma in any institute or has been published in any form, except where due acknowledgement is made.

Student's names:	Signatures:	Dates:
SUDIPTO PAUL		**13/01/2022**
PHIROZE CHOWDHURY		**13/01/2022**
SUPARNA DEBNATH		**13/01/2022**

CERTIFICATE OF APPROVAL

We hereby approve this dissertation titled

AUTISM CLASSIFICATION USING MACHINE LEARNING

Carried out by

SUDIPTO PAUL, Roll no. 29171020031, **Regd. No.** 202910571010001 of 2020-21

PHIROZE CHOWDHURY, **Roll no**. 29171020046, **Regd. No.** 202910571010036 of 2020-21

SUPARNA DEBNATH, Roll no. 29171020016, **Regd. No.** 202910571010016 of 2020-21

Under the guidance of
PROF. SUDIP DEY

Of Netaji Subhash Engineering College, Kolkata in partial fulfillment of requirements for award of the Master of Computer Application of West Bengal University of Technology

Date:

Examiners' signatures:

1. ……

2. ……

3. ……

Acknowledgement and Dedication

Firstly, we would like to thank our Project Guide **PROF. SUDIP DEY**, who guided us in doing these projects. He provided us with invaluable advice and helped us in difficult periods. His motivation and help contributed tremendously to the successful completion of the project.

Besides, we would like to thank all the teachers who helped us by giving us advice and providing the equipment which we needed.

At last but not in least, we would like to thank everyone who helped and motivated us to work on this project.

Dated: 13/01/2022

SUDIPTO PAUL

PHIROZE CHOWDHURY

SUPARNA DEBNATH

AUTISM CLASSIFICATION USING MACHINE LEARNING

INTRODUCTION TO MACHINE LEARNING

- **Machine learning** is a field of computer science that gives computers the ability to learn without being explicitly programmed.

- Machine Learning models consists algorithms that can learn and make predictions from data. Machine Learning has been evolved from prediction making and computational learning theory in artificial intelligence. It helps computers to learn and perform a certain task based on past experience. These models can be based on following:

  - ❖ Classification of predictions.
  - ❖ Classification of process where machine can recognize and categorize things from datasets.

INTRODUCTION TO AUTISM

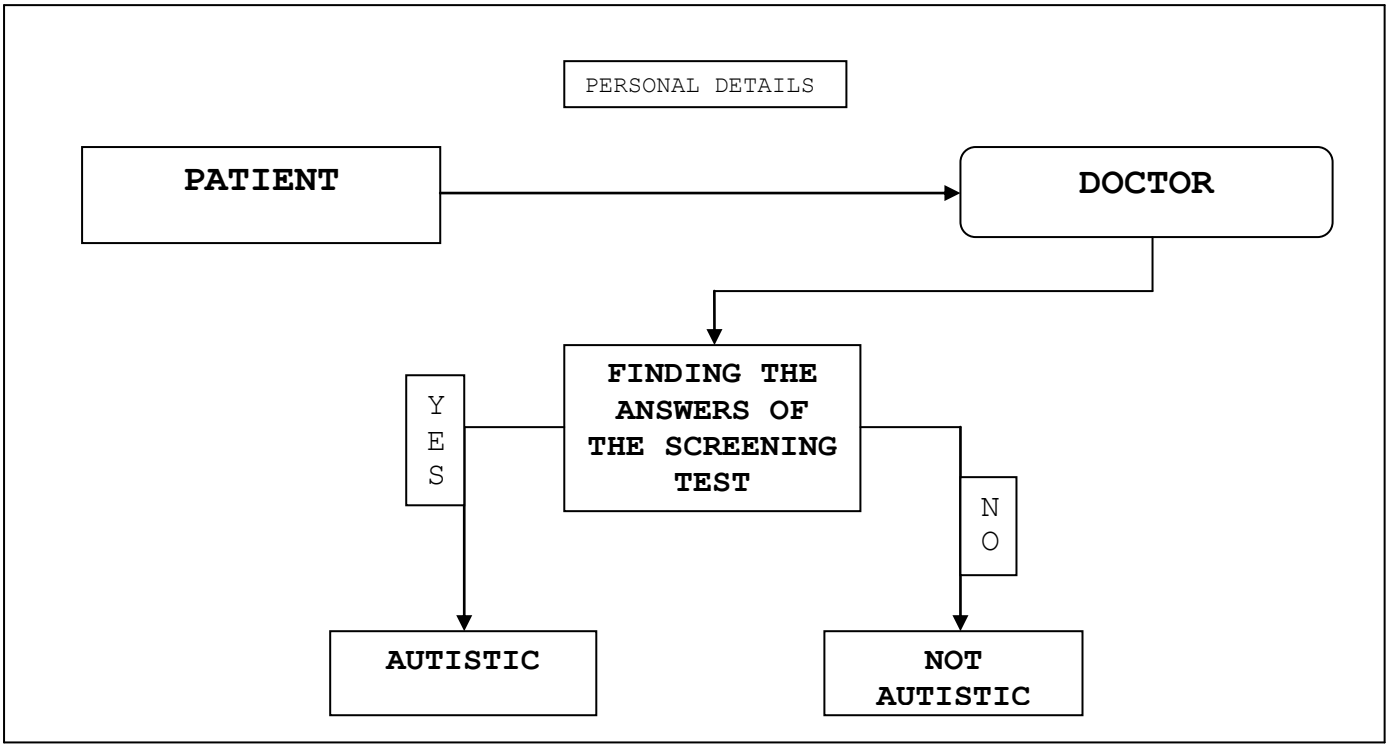
Autism Spectrum Disorder (ASD) is mainly reflected in the communication and language barriers, difficulties in social communication, and it is a kind of neurological developmental disorder. Most researches have used the machine learning method to classify patients and normal controls, among which Support Vector Machines (SVM) are widely employed. But the classification accuracy of SVM is usually low, due to the usage of a single SVM as classifier. Thus, we used multiple SVMs to classify ASD patients and Typical Controls (TC). Abnormal brain regions could also be found, such as Inferior Frontal Gyrus (IFG) (orbital and opercula part), hippocampus, and precuneus. It is indicated that the method of random SVM cluster may apply to the auxiliary diagnosis of ASD.

Key Words

- ASD
- SVM
- ML
- SkLEARN
- DECISION TREE
- LOGISTIC REGRESSION
- AUTISTIC/NON-AUTISTIC
- BINARY CLASSIFIER
- SCREENING
- SUPERVISED

MODULES USED

- Data flow implemented in typical deep learning projects. Boxes represent software infrastructure to be developed and arrows represent the data flow.
- This is what is going to be shown from this DFD diagram .Patients with autism will take their problems to the doctor .Then doctor will look at each of the symptoms and detect if there is any or not.



SOFTWARES USED

- VISUAL STUDIO CODE
- MS EXCEL
- MS POWER POINT
- ADOBE ACROBAT DC PDF READER

HARDWARES REQUIREMENT

- Windows 10 OS
- 64-bit, four-core, 2.5 GHz minimum per core Processor(CPU)
- 8GB RAM
- 512GB HDD

ALGORITHMS UESD

- **DECISION TREE:** Decision trees are employed to visually represent decisions and show or inform decision making. When working with machine learning and data mining, decision trees are used as a predictive model. These models map observations about data to conclusions about the data's target value.

```
from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier(random_state=0)
clf.fit(X_train,y_train)
y_prd = clf.predict(X_test)

mse = mean_squared_error(list(y_test), list(y_prd))
rmse = np.sqrt(mse)
print(" Decision Tree RMSE= ", rmse)
```

- **SUPPORT VECTOR CLASSIFIER(SVC):** Support Vector Classifier(SVC) is a supervised machine learning algorithm which can be used for classification or regression problems. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible output.

```
from sklearn.svm import SVC
clfsvc = SVC(gamma='auto')
clfsvc.fit(X_train,y_train)
y_prd1 = clfsvc.predict(X_test)

svcmse = mean_squared_error(list(y_test), list(y_prd))
svcrmse = np.sqrt(mse)
print("SVM RMSE = ", svcrmse)
```

- **LOGISTIC REGRESSION:** Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

```
from sklearn.linear_model import LogisticRegression
logcls = LogisticRegression()
```

```
    logcls.fit(X_train,y_train)
    y_prd = logcls.predict(X_test)

    from sklearn.metrics import mean_squared_error
    mse = mean_squared_error(list(y_test), list(y_prd))
    rmse = np.sqrt(mse)
    print("Logistic Regression RMSE = ", rmse)
```

CASE STUDIES

1. Arun was brought for consultation with Dr. A.M. Reddy by his parents. He was about 4 years old, the second child to the parents. Even while he was being brought into the room, we could hear his loud wailing. It took some time for the child to calm down and later we could observe that the child was very restless. He was running around the room, pulling down cushions and generally creating chaos in the room and mother was quite harried in trying to control him. He was diagnosed with ASD (Autism Spectrum Disorder).
2. Tomeika is a three-year-old girl. She was recently diagnosed with autistic disorder. Tomeika is able to make many vocalizations and is able to say one recognizable word. Tomeika will say “juice”, which she pronounces as “oos.” Throughout the day, Tomeika cries and falls to the floor to gain access to food, obtain a favorite toy, or when she wants to be picked up. Her parents, Mr. and Mrs. Williams, would like for Tomeika to communicate her desires with words, but do not know how to help her. Tomeika recently began attending an early childhood special education classroom for learners with ASD in the Hampton County Public School System for six hours a day, four days a week. On Wednesdays, Tomeika and her peers do not go to school. Instead, on this day, Tomeika’s interventionist, Mrs. Dell, has parent conferences in her students’ homes. During the conferences, Mrs. Dell discusses educational programming, learner progress, areas of concern, and also provides training to parents.
3. Mikey is an eight-year old male who was diagnosed with ASD with a severity requiring substantial support at age four by his pediatrician. Mikey is in third grade and has seen an occupational therapist through his public school system for the past two years. Mikey has difficulty maintaining social interactions and engaging in age appropriate play with his peers and sibling. Mikey’s parents have expressed that he is having difficulty with his morning routine, causing behavioral issues. When Mikey has these issues his entire day is thrown off, resulting in trouble concentrating on schoolwork and increasing difficulties interacting with his peers throughout the day. Mikey’s parents would like a home evaluation and recommendations from an occupational therapist in order to allow him to better function, especially during the mornings. Mikey’s family would be willing to purchase accommodations to better help Mikey function, and would commit to occupational therapy treatment outside of the school.

PROJECT CODE

```
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt

# Loading the dataset
data = pd.read_csv("Autism-Adult-Data.csv", na_values=["?"])
data_rv = data.drop("age_desc", axis=1)

print("the data attributes are: ")
print(data_rv.columns)

#filling empty data
data_rv["ethnicity"].fillna("missing", inplace=True)
data_rv["relation"].fillna("missing", inplace=True)

#Encoding categorical data
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
data_encode = data_rv.apply(le.fit_transform)

#print("##########After Label Encoding###########")
#print(data_encode)

X = data_encode.iloc[:, :19].values
y = data_encode.iloc[:, 19].values

#print("#########Features#######################")
#print(X)
#print("####################lebel#######################")
#print(y)
#print("###############################################")

#print("data_encode.head()/n", data_encode.head())
print("data_encode.info()/n",data_encode.info())
print("data_encode.describe()/n",data_encode.describe())

# # For plotting histogram
```

```
data_encode.hist(bins=50, figsize=(20, 15))
plt.show()

#corr_matrix1 = data_encode.corrwith("ASD")
print(data_encode[data_encode.columns[1:]].corr()['ASD'].sort_values(ascending=False))
#corr_matrix['austim'].sort_values(ascending=False)
from pandas.plotting import scatter_matrix
#attributes = X.columns
#scatter_matrix(data_encode)

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test =train_test_split(X, y, test_size = .2, random_state = 42)



from sklearn.linear_model import LogisticRegression
logcls = LogisticRegression()
logcls.fit(X_train,y_train)
y_prd = logcls.predict(X_test)

from sklearn.metrics import mean_squared_error
mse = mean_squared_error(list(y_test), list(y_prd))
rmse = np.sqrt(mse)
print("Logistic Regression RMSE = ", rmse)

from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier(random_state=0)
clf.fit(X_train,y_train)
y_prd = clf.predict(X_test)

mse = mean_squared_error(list(y_test), list(y_prd))
rmse = np.sqrt(mse)
print(" Decision Tree RMSE= ", rmse)

from sklearn.svm import SVC
clfsvc = SVC(gamma='auto')
clfsvc.fit(X_train,y_train)
y_prd1 = clfsvc.predict(X_test)

svcmse = mean_squared_error(list(y_test), list(y_prd))
svcrmse = np.sqrt(mse)
print("SVM RMSE = ", svcrmse)
```

ADVANTAGES OF CLASSIFICATION USING MACHINE LEARNING

- This algorithm is simple to implement, robust to noisy training data, and effective if training data is large.

- Data classification helps you prioritize your data protection efforts to improve data security and regulatory compliance.

- A common job of machine learning algorithms is to recognize objects and being able to separate them into categories. This process is called classification, and it helps us segregate vast quantities of data into discrete values, i.e.: distinct, like 0/1, True/False, or a pre-defined output label class.

DISADVANTAGES OF CLASSIFICATION USING MACHINE LEARNING

- A disadvantage to classification is that many of the classifications themselves are based on subjective judgments, which may or may not be shared by everyone participating. This would lead to differences in perceived value.
- In ML, we can choose the algorithms based on accurate results. For that, we have to run the results on every algorithm. The main problem occurs in the training and testing of data. The data is huge, so sometimes removing errors becomes nearly impossible. These errors can cause a headache to users. Since the data is huge, the errors take a lot of time to resolve.
- In ML, we constantly work on data. We take a huge amount of data for training and testing. This process can sometimes cause data inconsistency.

FUTURE SCOPE OF THE PROJECT

- If more data can be collected then there could be a better result with the machine learning algorithms

- In future this project can be implemented as an application, from where it would be easier for the doctors for the patients to dectect the disease

FUTURE TREATMENTS OF AUTISM

- Behavioural management therapy
- Cognitive behaviour therapy
- Early intervention
- Educational and school-based therapies.

CONCLUSION

The rapid growth in the number of ASD cases worldwide necessitates datasets related to behaviour traits. However, such datasets are rare making it difficult to perform thorough analyses to improve the performance of the screening. Presently, limited autism datasets associated with clinical diagnosis or screening are available and most of them are genetic in nature. Hence, we propose new machine learning framework with datasets related to the autism screening of adults and adolescents that have influential features and perform predictive analysis using Logistic Regression. In these datasets, we record ten behavioural features based on AQ (adults, adolescents) screening methods plus an individual's characteristics; these have proved to be effective in detecting the ASD cases from controls in behavioural science. We also perform an in-depth feature analysis on the two datasets using feature selection to determine the effective features that can be utilized in screening for autism. In conclusion, this research reported that machine learning technology specially function based ones (logistic regression) showed promising results in ASD screening at least for the adults and adolescents. In the near future, we intend to implement a new screening method using machine learning technology for toddlers and children.

BIBLIOGRAPHY

- DATASET:
  https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult

- CASE STUDY:
  1. https://positiveautism.com/case-study-of-autism/
  2. https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=3638&context=honors_theses
  3. https://autismpdc.fpg.unc.edu/sites/autismpdc.fpg.unc.edu/files/PII-casestudy-tomeika.pdf
- OTHER LINKS:
  1. https://www.frontiersin.org/articles/10.3389/fninf.2020.575999/full
  2. https://www.autism-society.org/what-is/facts-and-statistics/
  3. https://machinelearningmastery.com/naive-bayes-for-machine-learning/
  4. https://www.knowledgehut.com/blog/data-science/logistic-regression-for-machine-learning
  5. Duda M, Ma R, Haber N, Wall DP. Use of machine learning for behavioral distinction of autism and ADHD. Transl Psychiatry. 2016;6: e732.
  6. Google Scholar, Bosl, W. J., Tager-Flusberg, H., and Nelson, C. A. (2018). Eeg analytics for early detection of autism spectrum disorder: a data-driven approach. Sci. Rep. 8:6828. doi: 10.1038/s41598-018-24318-x