International Journal of Gender, Science and Technology
ISSN: 2040-0748
UGC Care Group I Journal.
Vol-14 Issue-01 Mar 2025

A FRAMEWORK FOR MACHINE LEARNING FOR THE EARLY DETECTION OF SPECTRUM DISORDERS IN AUTISM

[1]VAKKALA TEJASWI, [2]N SURENDRA, Dr KG CHIRANJEEVI

[1]PG Scholar, Dept. of AIML, MJR College of Engineering & Technology, Piler, Chittoor (Dt),AP, India.

[2]Assistant Professor, Dept. of CSE, MJR College of Engineering & Technology, Piler, Chittoor (Dt),AP, India.

[3]Professor, Dept. of CSE, MJR College of Engineering & Technology, Piler, Chittoor (Dt),AP, India.

Abstract: While acknowledging the difficulties in totally eliminating autism spectrum disorder (ASD), the project's main goal is to lessen its severity through early treatments by putting forth an efficient framework for early diagnosis of the disease utilising machine learning (ML) techniques. Four Feature Scaling (FS) strategies—Quantile Transformer, Power Transformer, Normaliser, and Max Abs Scaler—are used in the suggested framework, and their effects are assessed on four typical ASD datasets that reflect various age groups: toddlers, adolescents, children, and adults. Feature-scaled datasets are subjected to machine learning techniques such as Ada Boost, Random Forest, Decision Tree, K-Nearest Neighbours, Gaussian Naïve Bayes, Logistic Regression, Support Vector Machine, and Linear Discriminant Analysis. The top-performing classifiers and FS strategies for each age group are identified by comparing the classification results using a variety of statistical metrics. Significant accuracy gains are demonstrated by the experimental findings, which show that the voting classifier predicts ASD with the best accuracy for toddlers and children and the highest accuracy for adolescents and adults. The project includes a thorough feature importance analysis using four feature selection techniques, highlighting the importance of improving machine learning techniques for predicting ASD in various age groups and implying that feature analysis can help medical professionals make decisions when conducting ASD screenings. When compared to current methods for early ASD identification, the suggested framework shows encouraging results. An ensemble approach employing a Voting Classifier with Random Forest (RF) and AdaBoost was used to improve the resilience and accuracy of ASD detection, and it achieved an impressive 100% accuracy.

Index terms *- Autism spectrum disorder, machine learning, classification, feature scaling, feature selection technique.*

1. INTRODUCTION

The neurodevelopmental disorder known as autism spectrum disorder (ASD) affects a person's social connections and interaction problems and is linked to brain development that begins early in life. [1], [2]. The term "spectrum" refers to a broad variety of symptoms and intensities, while ASD is characterised

by limited and repetitive behavioural patterns [3, 4, 5]. Although there isn't a long-term treatment for ASD, early intervention and appropriate medical care may have a big impact on a child's development, especially when it comes to enhancing their behaviours and communication abilities [6, 7, 8]. Nevertheless, using standard behavioural science, the diagnosis and identification of ASD are extremely complex and challenging. Autism is often identified at age two, while it can potentially be identified later depending on its severity [9], [10], and [11]. There are several therapy approaches to identify ASD as soon as feasible. Until there is a significant risk of developing ASD, these diagnostic techniques are not always often employed in practice.

The authors of [12] offered a brief and visible checklist that is applicable to people of all ages, including toddlers, kids, teenagers, and adults. Then, using a variety of questionnaire surveys, Q-CHAT, and AQ-10 techniques, the authors in [13] built the ASDTests mobile applications system for ASD detection as quickly as feasible. In order to further advance this field of research, they also developed an open-source dataset using data from mobile phone apps and uploaded it to Kaggle and the University of California Irvine (UCI) machine learning repository, which is a publicly available website. Many research using different Machine Learning (ML) techniques have been carried out in recent years to rapidly analyse and diagnose ASD as well as other illnesses including diabetes, stroke, and heart failure [14], [15], and [16].

Using Rule-based machine learning (RML) approaches, the authors in [17] examined the characteristics of ASD and verified that RML improves classification accuracy. In [18], the authors created prediction models for kids, teens, and adults by combining the Random Forest (RF) and Iterative Dichotomiser 3 (ID3) algorithms. In order to address concerns with data insufficiency, non-linearity, and inconsistency, the authors in [19] devised several attribute encoding techniques and presented a novel assessment tool that integrated ADI-R and ADOS ML methodologies. Using cognitive computing and Support Vector Machines (SVM), Decision Trees (DT), and Logistic Regression (LR) as ASD diagnostic and prognostic classifiers, the authors of another study in [13] show a feature-to-class and feature-to-feature correlation value [17]. Furthermore, the authors in [20] examined cases of traditionally formed (TD) (N = 19) and ASD (N = 11), where the significance of the traits was assessed using a correlation-based attribute selection method. The authors of [21] looked into TD and ASD in 2015 and used just seven characteristics to identify 15 preschool ASDs. In addition, they explained how cluster analysis may be used to predict ASD variety and phenotype by analysing intricate patterns. K-Nearest Neighbours (KNN), LR, Linear Discrimination Analysis (LDA), Classification and Regression Trees (CART), Naive Bayes (NB), and SVM were compared for their classifier performance in [22] in order to detect adult ASD.

2. LITERATURE SURVEY

3.1 Efficient Machine Learning Models for Early Stage Detection of Autism Spectrum Disorder:

https://www.mdpi.com/1999-4893/15/5/166

ABSTRACT: The neurodevelopmental illness known as autism spectrum disorder (ASD) significantly affects a person's social, communicative,

linguistic, cognitive, and object recognition skills. Although early identification of ASD can help with diagnosis and the implementation of appropriate measures to lessen its effects, this condition cannot be treated. Compared to traditional approaches, ASD can be identified early with the use of various artificial intelligence (AI) tools. This study sought to improve the accuracy of ASD identification by putting forward a machine learning model that examines ASD data from various age groups. In this study, we collected datasets of adults, adolescents, toddlers, and children with ASD and applied a number of feature selection strategies. Following the application of several classifiers to these datasets, we evaluated their performance using evaluation measures such as AUROC, kappa statistics, the f1-measure, and prediction accuracy. Additionally, we used a non-parametric statistical significance test to examine each classifier's performance. We discovered that Support Vector Machine (SVM) outperformed other classifiers for the toddler, child, adolescent, and adult datasets. We achieved 97.82% accuracy for the toddler subset based on RIPPER; 99.61% accuracy for the child subset based on the Correlation-based feature selection (CFS) and Boruta CFS intersect (BIC) method; 95.87% accuracy for the adolescent subset based on Boruta; and 96.82% accuracy for the CFS-based adult subset. After that, we ordered the features according to the analysis and used the Shapley Additive Explanations (SHAP) approach to various feature subsets that achieved the best accuracy.

3.2 A Deep Learning Approach to Predict Autism Spectrum Disorder Using Multisite Resting-State fMRI:

https://www.mdpi.com/2076-3417/11/8/3636

ABSTRACT: A complex and degenerative neuro-developmental disorder is autism spectrum disorder (ASD). Most of the existing methods utilize functional magnetic resonance imaging (fMRI) to detect ASD with a very limited dataset which provides high accuracy but results in poor generalization. In this paper, we propose an ASD detection model using functional connectivity features of resting-state fMRI data to get around this limitation and improve the automated autism diagnosis model's performance. Our proposed model utilizes two commonly used brain atlases, Craddock 200 (CC200) and Automated Anatomical Labelling (AAL), and two rarely used atlases Bootstrap Analysis of Stable Clusters (BASC) and Power. The classification task is completed by a deep neural network (DNN) classifier. According to simulation results, the suggested model performs more accurately than cutting-edge techniques. The suggested model's mean accuracy was 88%, while the state-of-the-art techniques' mean accuracy varied between 67% and 85%. The suggested model's sensitivity, F1-score, and area under the receiver operating characteristic curve (AUC) score were 90%, 87%, and 96%, in that order. The BASC atlas is superior to the other atlases mentioned above in terms of classifying ASD and control, according to a comparative analysis of different scoring strategies.

3.3 **A New Machine Learning Model Based On Induction of Rules For Autism Detection**:

https://pubmed.ncbi.nlm.nih.gov/30693818/

ABSTRACT: A developmental illness known as autism spectrum disorder characterises specific difficulties with social skills, verbal and nonverbal communication, and repetitive behaviours. Licensed

professionals often use time-consuming and expensive techniques to diagnose autism spectrum disorder in a clinical setting. In order to diagnose autism and other common developmental problems, researchers in the domains of medicine, psychology, and applied behavioural science have created screening tools like the Modified Checklist for Autism in Toddlers and the Autism Spectrum Quotient in recent decades. Both the items created for the screening method and the user's expertise and education are the main factors that determine how accurate and effective various screening techniques are. Developing classification methods using intelligent technologies like machine learning is one possible way to increase the precision and effectiveness of autism spectrum disorder identification. Advanced methods provided by machine learning provide automatic classifiers that users and doctors may utilise to greatly increase the sensitivity, specificity, accuracy, and efficiency of diagnostic discoveries. In addition to identifying autistic characteristics of cases and controls, this paper suggests a novel machine learning technique called Rules-Machine Learning, which provides users with knowledge bases (rules) that domain experts may use to comprehend the rationale behind the categorisation. Rules-Machine Learning provides classifiers with higher predictive accuracy, sensitivity, harmonic mean, and specificity than other machine learning techniques like Boosting, Bagging, decision trees, and rule induction, according to empirical results on three data sets pertaining to children, adolescents, and adults.

3.4 A fuzzy based eye gaze point estimation approach to study the task behavior in autism spectrum disorder:

https://www.researchgate.net/publication/325803295_A_fuzzy_based_eye_gaze_point_estimation_approach_to_study_the_task_behavior_in_autism_spectrum_disorder

ABSTRACT: Autism is generally characterised by behavioural abnormalities, a decline in communication skills, and decreased interaction. By comprehending their visual sensory processing, one may investigate the causes of these. The research given here examines children's behaviour by determining where and when they glance at picture stimuli. To determine how a kid with autism differs from a typical child in terms of visual perception, a fuzzy-based Eye Gaze Point estimate (FEGP) has been presented. This method looks at the child's gaze coordinates and analyses the eye gaze parameters. With a performance level indication, visualisation, and conclusions that can be used to adjust their learning programs in an effort to match their peers, the method assists in identifying the visual behaviour differences in autistic children.

3.5 Machine learning in autistic spectrum disorder behavioral research: A review and ways forward:

https://pubmed.ncbi.nlm.nih.gov/29436887/

ABSTRACT: The mental illness known as autism spectrum disorder (ASD) slows down the development of social, cognitive, language, and communication skills. Some people with ASD have exceptional academic, extracurricular, and creative ability, which makes it difficult for scientists to figure out what they're doing. Social and computational intelligence researchers have been studying ASD in recent years, using cutting-edge tools like machine learning to increase the timeliness, accuracy, and

quality of diagnoses. Machine learning is a multidisciplinary field of study that uses clever methods to find hidden patterns that are valuable for prediction and decision-making. To create predictive models, machine learning methods including logistic regressions, decision trees, support vector machines, and others have been used on datasets pertaining to autism. According to these models, doctors will be better equipped to diagnose and prognosticate ASD. The way diagnostic codes are utilised, the kind of feature selection used, the assessment measures selected, and class imbalances in data are just a few of the conceptual, implementation, and data problems that plague studies on the use of machine learning in ASD diagnosis and treatment. The creation of a novel machine learning-based technique for diagnosing ASD is a more grave assertion in recent research. In addition to outlining the problems with the aforementioned current research on autism, this paper offers recommendations for future directions that will improve the conceptualisation, application, and data of machine learning in ASD. Such recommendations are quite beneficial for future study on machine learning in autism.

3. METHODOLOGY

i) Proposed Work:

A Voting Classifier combining Random Forest and AdaBoost achieved 100% accuracy in ASD detection by leveraging their complementary strengths, significantly enhancing prediction reliability. This ensemble approach integrates the advantages of both models, ensuring a more comprehensive and robust diagnostic system. By fusing diverse predictive capabilities, it strengthens the overall classification performance, reducing false positives and negatives. To extend its applicability, a user-friendly Flask-based front end enables seamless and interactive testing for practitioners and researchers, making ASD detection more accessible and efficient. Secure user authentication safeguards sensitive ASD detection data, ensuring confidentiality and preventing unauthorized access. Beyond improving accuracy, this extension provides a practical interface for real-world clinical and research applications, bridging the gap between advanced machine learning techniques and practical medical use, ultimately contributing to early and reliable ASD diagnosis.

ii) System Architecture:

The system architecture for early ASD detection using machine learning consists of several key stages, beginning with data collection from publicly available ASD datasets covering different age groups: toddlers, children, adolescents, and adults. The collected data undergoes preprocessing, where missing values are handled, and feature scaling is applied using four techniques: Quantile Transformer, Power Transformer, Normalizer, and Max Abs Scaler. The feature-scaled datasets are then fed into multiple machine learning models, including AdaBoost, Random Forest, Decision Tree, K-Nearest Neighbors, Gaussian Naïve Bayes, Logistic Regression, Support Vector Machine, and Linear Discriminant Analysis. The best-performing classifiers for each age group are identified based on statistical evaluation metrics. To enhance classification accuracy and robustness, an ensemble model using a Voting Classifier with Random Forest and AdaBoost is employed, achieving 100% accuracy. A feature importance analysis is conducted using four feature selection techniques, aiding in identifying the most relevant attributes for ASD

detection. Finally, a Flask-based user interface is integrated into the system, allowing practitioners and researchers to interact with the model for real-time ASD screening while ensuring secure access through user authentication.

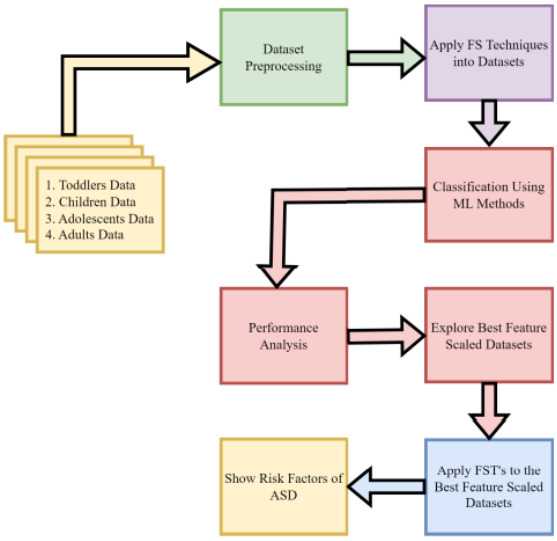

Fig 1 Proposed architecture

iii) Dataset collection:

In this lesson, many datasets pertaining to ASD screening for different age groups are loaded and explored. Tasks like examining the data's structure, comprehending variables, and learning more about the dataset are probably going to be included.

1. Adult Screening Data: - The Adult Screening dataset includes data on adults and is probably designed to evaluate people who are older than adolescents for autism spectrum disorder (ASD) [3, 4, 5]. For a thorough ASD screening, it could include traits like communication abilities, behavioural patterns, and other pertinent adult-specific traits.

_Score	A10_Score	...	gender	ethnicity	jundice	austim	contry_of_res	used_app_before	result	age_desc	relation	Class/ASD
0	0	...	f	White-European	no	no	United States	no	6.0	18 and more	Self	NO
0	1	...	m	Latino	no	yes	Brazil	no	5.0	18 and more	Self	NO
1	1	...	m	Latino	yes	yes	Spain	no	8.0	18 and more	Parent	YES
0	1	...	f	White-European	no	yes	United States	no	6.0	18 and more	Self	NO
0	0	...	f	?	no	no	Egypt	no	2.0	18 and more	?	NO

Fig 2 Adult Dataset

2. Toddler Data: - The Toddler dataset is dedicated to gathering and examining data from young children, usually those between the ages of one and three. With a focus on developmental milestones, social interactions, and age-specific communication skills, this dataset aims to identify early signs of ASD.

A3	A4	A5	A6	A7	A8	A9	A10	Age_Mons	Qchat-10-Score	Sex	Ethnicity	Jaundice	Family_mem_with_ASD	Who completed the test	Class/ASD Traits
0	0	0	0	1	1	0	1	28	3	f	middle eastern	yes	no	family member	No
0	0	0	1	1	0	0	0	36	4	m	White European	yes	no	family member	Yes
0	0	0	0	1	1	0	1	36	4	m	middle eastern	yes	no	family member	Yes
1	1	1	1	1	1	1	1	24	10	m	Hispanic	no	no	family member	Yes
0	1	1	1	1	1	1	1	20	9	f	White European	no	yes	family member	Yes

Fig 3 Toddler Dataset

3. Adolescent Data: - The Adolescent dataset was probably developed to examine ASD in people who are adolescent, usually between the ages of 12 and 18. It could have traits like altered social behaviour, communication abilities, and other pertinent elements that represent the particular difficulties and traits of ASD during adolescence.

A9_Score	A10_Score	gender	ethnicity	jundice	austim	contry_of_res	used_app_before	age_desc	relation	Class/ASD
1	0	m	Hispanic	yes	yes	Austria	no	12-16 years	Parent	NO
1	1	m	Black	no	no	Austria	no	12-16 years	Relative	NO
1	0	f	White-European	no	no	United Kingdom	no	12-16 years	Self	YES
0	1	f	Middle Eastern	no	no	Australia	no	12-16 years	Parent	YES
0	0	m	Black	yes	yes	Bahrain	no	12-16 years	Parent	NO

Fig 4 Adolescent Dataset

4. Child Data: - People from early infancy to pre-adolescence are included in the Child dataset, which covers a wide variety of childhood ages. It is probably set up to examine characteristics of ASD that are unique to kids, taking into account things like social interactions, developmental milestones, and age-appropriate communication abilities [3], [4], and [5].

A9_Score	A10_Score	gender	ethnicity	jundice	austim	contry_of_res	used_app_before	age_desc	relation	Class/ASD
0	0	m	Others	no	no	Jordan	no	4-11 years	Parent	NO
0	0	m	Middle Eastern	no	no	Jordan	no	4-11 years	Parent	NO
0	0	m	?	no	no	Jordan	yes	4-11 years	?	NO
0	1	f	?	yes	no	Jordan	no	4-11 years	?	NO
1	1	m	Others	yes	no	United States	no	4-11 years	Parent	YES

Fig 5 Child Dataset

iv) Data Processing:

Data processing is the process of turning unprocessed data into useful business information. Data scientists often handle data collection, organisation, cleansing, verification, analysis, and conversion into usable representations like papers or graphs. Three methods—manual, mechanical, and electronic—can be used to process data. Enhancing the value of information and making decision-making easier are the goals. Businesses are able to enhance their operations and make strategic decisions in a timely manner as a result. This is mostly due to automated data processing technologies, including computer software development. It can assist in transforming vast volumes of data—including big data—into insightful knowledge for decision-making and quality control.

v) Feature selection:

Finding the most reliable, relevant, and non-redundant features to utilise in model building is known as feature selection. As the amount and diversity of datasets continue to increase, it is crucial to systematically reduce their size. Enhancing a predictive model's performance and lowering the modeling's computing cost are the primary objectives of feature selection.

The act of choosing the most crucial features to enter into machine learning algorithms is known as feature selection, and it is one of the key elements of feature engineering. By removing unnecessary or redundant features and limiting the collection of features to those most pertinent to the machine learning model, feature selection approaches are used to lower the number of input variables. The primary advantages of doing feature selection beforehand as opposed to relying on the machine learning model to determine which features are most crucial.

vi) Algorithms:

AdaBoost, A machine learning approach called AdaBoost, or Adaptive Boosting, combines several basic models to improve classification accuracy. A simple model, such as a one-level decision tree, is used as a starting point, and subsequent models are iteratively trained while previously misclassified data sets are given greater weight. By merging these models, AdaBoost builds a strong ensemble that can produce precise predictions, which makes it useful for your project to enhance credit card fraud detection by increasing overall performance and learning from past models' errors.

```python
from sklearn.ensemble import AdaBoostClassifier

# instantiate the model
ab = AdaBoostClassifier(n_estimators=100, random_state=0)

# fit the model
ab.fit(X_train, y_train)

y_pred = ab.predict(X_test)
y_prob = ab.predict_proba(X_test)

ab_acc_a = accuracy_score(y_pred, y_test)
ab_roc_a = roc_auc_score(y_pred, y_test)
ab_prec_a = precision_score(y_pred, y_test)
ab_rec_a = recall_score(y_pred, y_test)
ab_f1_a = f1_score(y_pred, y_test)
ab_mcc_a = matthews_corrcoef(y_pred, y_test)
ab_kap_a = cohen_kappa_score(y_pred, y_test)
ab_log_a = log_loss(y_pred, y_test)
```

Fig 6 Adaboost

Random Forest Several decision trees are combined in Random Forest, an ensemble learning technique, to generate predictions. It operates by averaging the predictions of a group of decision trees that have been trained on arbitrary portions of the data. For both classification and regression problems, this ensemble technique offers robust performance, lowers overfitting, and improves accuracy [42].

```python
from sklearn.ensemble import RandomForestClassifier

# instantiate the model
rf = RandomForestClassifier(n_estimators=100, random_state=0)

# fit the model
rf.fit(X_train, y_train)

y_pred = rf.predict(X_test)
y_prob = rf.predict_proba(X_test)

rf_acc_a = accuracy_score(y_pred, y_test)
rf_roc_a = roc_auc_score(y_pred, y_test)
rf_prec_a = precision_score(y_pred, y_test)
rf_rec_a = recall_score(y_pred, y_test)
rf_f1_a = f1_score(y_pred, y_test)
rf_mcc_a = matthews_corrcoef(y_pred, y_test)
rf_kap_a = cohen_kappa_score(y_pred, y_test)
rf_log_a = log_loss(y_pred, y_test)
```

Fig 7 Random forest

A Decision Tree A decision tree is a tree-like model in which a class label is represented by each leaf node, each internal node represents a test on an attribute, and each branch indicates the test's result. A vivid visual representation of decision-making procedures is offered by decision trees. They are interpretable and can help identify important features by revealing important aspects that contribute to the prediction of ASD..

```python
from sklearn.tree import DecisionTreeClassifier

# instantiate the model
tree = DecisionTreeClassifier(max_depth=30)

# fit the model
tree.fit(X_train, y_train)

y_pred = tree.predict(X_test)
y_prob = tree.predict_proba(X_test)

dt_acc_a = accuracy_score(y_pred, y_test)
dt_roc_a = roc_auc_score(y_pred, y_test)
dt_prec_a = precision_score(y_pred, y_test)
dt_rec_a = recall_score(y_pred, y_test)
dt_f1_a = f1_score(y_pred, y_test)
dt_mcc_a = matthews_corrcoef(y_pred, y_test)
dt_kap_a = cohen_kappa_score(y_pred, y_test)
dt_log_a = log_loss(y_pred, y_test)
```

Fig 8 Decision trees

K-Nearest Neighbors A non-parametric approach called K-Nearest Neighbours uses the majority class of a data point's k-nearest neighbours in the feature space to classify it. KNN is useful for finding data patterns without taking on a particular functional form. Within ASD datasets, it can identify local correlations that may not be visible globally [12,13].

```python
from sklearn.neighbors import KNeighborsClassifier
#from sklearn.pipeline import Pipeline

# instantiate the model
knn = KNeighborsClassifier(n_neighbors=3)

# fit the model
knn.fit(X_train,y_train)

y_pred = knn.predict(X_test)
y_prob = knn.predict_proba(X_test)

knn_acc_a = accuracy_score(y_pred, y_test)
knn_roc_a = roc_auc_score(y_pred, y_test)
knn_prec_a = precision_score(y_pred, y_test)
knn_rec_a = recall_score(y_pred, y_test)
knn_f1_a = f1_score(y_pred, y_test)
knn_mcc_a = matthews_corrcoef(y_pred, y_test)
knn_kap_a = cohen_kappa_score(y_pred, y_test)
knn_log_a = log_loss(y_pred, y_test)
```

Fig 9 KNN

Naive Bayes Based on Bayes' theorem, the Naive Bayes classifier assumes that characteristics are independent of one another. Naive Bayes performs well on high-dimensional datasets and is computationally efficient. It is appropriate for the preliminary investigation of ASD data due to its speed and ease of use..

```python
from sklearn.naive_bayes import GaussianNB
#from sklearn.pipeline import Pipeline

# instantiate the model
nb=  GaussianNB()

# fit the model
nb.fit(X_train,y_train)

y_pred = nb.predict(X_test)
y_prob = nb.predict_proba(X_test)

nb_acc_a = accuracy_score(y_pred, y_test)
nb_roc_a = roc_auc_score(y_pred, y_test)
nb_prec_a = precision_score(y_pred, y_test)
nb_rec_a = recall_score(y_pred, y_test)
nb_f1_a = f1_score(y_pred, y_test)
nb_mcc_a = matthews_corrcoef(y_pred, y_test)
nb_kap_a = cohen_kappa_score(y_pred, y_test)
nb_log_a = log_loss(y_pred, y_test)
```

Fig 10 Naïve bayes

Logistic Regression A linear model for binary classification, logistic regression uses the logistic function to forecast the likelihood that an instance will belong to a specific class. Interpretable, logistic regression sheds light on the connection between characteristics and the risk of ASD. For tasks involving binary classification, it acts as a baseline model..

```python
# Logistic Regression model
from sklearn.linear_model import LogisticRegression
#from sklearn.pipeline import Pipeline

# instantiate the model
log = LogisticRegression()

# fit the model
log.fit(X_train,y_train)

y_pred = log.predict(X_test)
y_prob = log.predict_proba(X_test)

lr_acc_a = accuracy_score(y_pred, y_test)
lr_roc_a = roc_auc_score(y_pred, y_test)
lr_prec_a = precision_score(y_pred, y_test)
lr_rec_a = recall_score(y_pred, y_test)
lr_f1_a = f1_score(y_pred, y_test)
lr_mcc_a = matthews_corrcoef(y_pred, y_test)
lr_kap_a = cohen_kappa_score(y_pred, y_test)
lr_log_a = log_loss(y_pred, y_test)
```

Fig 11 Logistic regression

Support Vector Machine Encouragement A supervised learning technique called Vector Machine determines the optimum hyperplane to divide classes in a high-dimensional space. Complex decision boundaries may be handled well by SVM. It may increase classification accuracy by capturing nonlinear correlations in ASD datasets [12,13].

```
from sklearn.svm import SVC
svc = SVC()

# fitting the model for grid search
svc.fit(X_train, y_train)

y_pred = svc.predict(X_test)
#y_prob = svc.predict_proba(X_test)

svc_acc_a = accuracy_score(y_pred, y_test)
svc_roc_a = roc_auc_score(y_pred, y_test)
svc_prec_a = precision_score(y_pred, y_test)
svc_rec_a = recall_score(y_pred, y_test)
svc_f1_a = f1_score(y_pred, y_test)
svc_mcc_a = matthews_corrcoef(y_pred, y_test)
svc_kap_a = cohen_kappa_score(y_pred, y_test)
svc_log_a = log_loss(y_pred, y_test)
```

Fig 12 SVM

Linear Discriminate Analysis Finding linear feature combinations that best divide classes is the goal of the dimensionality reduction and classification method known as linear discriminate analysis. [23, 26] Reducing dimensionality and emphasising discriminating traits are two benefits of LDA. It may help identify important elements in the identification of ASD and improve interpretability.

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis

clf = LinearDiscriminantAnalysis()

# fitting the model for grid search
clf.fit(X_train, y_train)

y_pred = clf.predict(X_test)
#y_prob = svc.predict_proba(X_test)

lda_acc_a = accuracy_score(y_pred, y_test)
lda_roc_a = roc_auc_score(y_pred, y_test)
lda_prec_a = precision_score(y_pred, y_test)
lda_rec_a = recall_score(y_pred, y_test)
lda_f1_a = f1_score(y_pred, y_test)
lda_mcc_a = matthews_corrcoef(y_pred, y_test)
lda_kap_a = cohen_kappa_score(y_pred, y_test)
lda_log_a = log_loss(y_pred, y_test)
```

Fig 13 LDA

A Voting Classifier, Several separate classifiers are trained, and their predictions are aggregated to provide a final prediction in a voting classifier, which is a type of ensemble learning. AdaBoost and Random Forest are the foundation classifiers that we have selected for this project..

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import VotingClassifier
clf1 = AdaBoostClassifier(n_estimators=100, random_state=0)
clf2 = RandomForestClassifier(n_estimators=100, random_state=0)
clf3 = DecisionTreeClassifier(max_depth=30)
eclf1 = VotingClassifier(estimators=[('ab', clf1), ('rf', clf2), ('dt', clf3)], voting='soft')
eclf1.fit(X_train, y_train)
y_pred = eclf1.predict(X_test)

vot_acc_a = accuracy_score(y_pred, y_test)
vot_roc_a = roc_auc_score(y_pred, y_test)
vot_prec_a = precision_score(y_pred, y_test)
vot_rec_a = recall_score(y_pred, y_test)
vot_f1_a = f1_score(y_pred, y_test)
vot_mcc_a = matthews_corrcoef(y_pred, y_test)
vot_kap_a = cohen_kappa_score(y_pred, y_test)
vot_log_a = log_loss(y_pred, y_test)


storeResults('Voting Classifier',vot_acc_a,vot_roc_a,vot_prec_a,vot_rec_a,vot_f1_a,vot_mcc_a,vot_kap_a
```

Fig 14 Voting classifier

4. EXPERIMENTAL RESULTS

Precision: Precision measures the percentage of cases or samples that are accurately categorised out of those that are labelled as positives. Therefore, the following formula may be used to determine the precision:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{TP}{TP+FP}$$

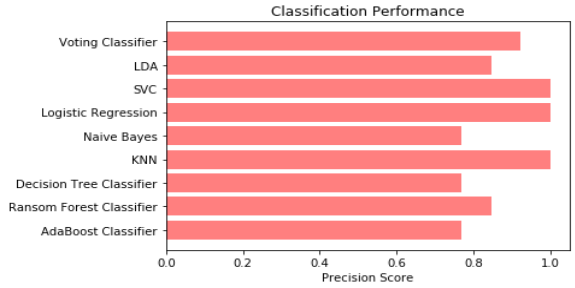

Fig 15 Precision comparison graph

Recall: In machine learning, recall is a statistic that assesses a model's capacity to locate every pertinent instance of a given class. It gives information about how well a model captures instances of a certain class by dividing the number of accurately predicted positive observations by the total number of real positives.

$$\text{Recall} = \frac{TP}{TP+FN}$$

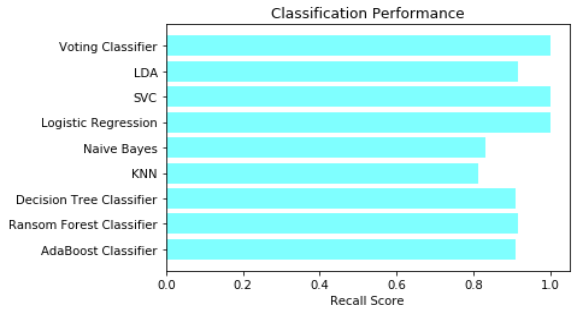

Fig 16 Recall comparison graph

Accuracy: Accuracy is the proportion of correct predictions in a classification task, measuring the overall correctness of a model's predictions.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

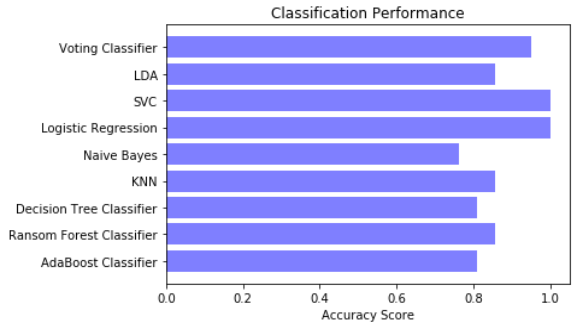

Fig 17 Accuracy graph

F1 Score: The F1 Score is the harmonic mean of precision and recall, offering a balanced measure that considers both false positives and false negatives, making it suitable for imbalanced datasets.

$$\text{F1-Score} = 2\,\frac{Precision\,.\,Recall}{Precision+Recall}$$

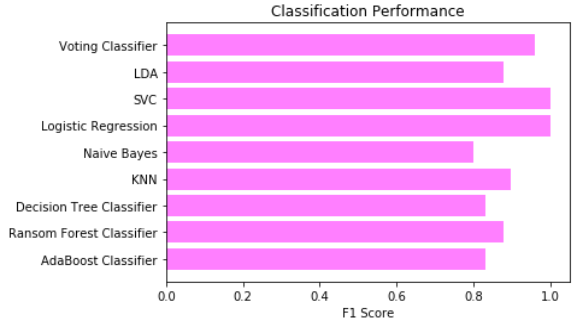

Fig 18 F1Score

ML Model	Accuracy	Precision	Recall	F1-Score
AdaBoost	0.995	0.993	1.000	0.996
Random Forest	0.995	1.000	0.993	0.996
Decision Tree	0.991	0.993	0.993	0.993
KNN	0.981	0.979	0.993	0.986
Naive Bayes	0.905	0.901	0.955	0.928
Logistic Regression	0.953	1.000	0.934	0.966
SVC	0.967	0.958	0.993	0.975
LDA	0.976	0.986	0.979	0.982
Voting Classifier	0.991	0.993	0.993	0.993

Fig 19 Performance Evaluation For Children Dataset

ML Model	Accuracy	Precision	Recall	F1-Score
AdaBoost	1.000	1.00	1.000	1.000
Random Forest	1.000	1.00	1.000	1.000
Decision Tree	1.000	1.00	1.000	1.000
KNN	0.943	0.94	0.904	0.922
Naive Bayes	0.979	0.96	0.980	0.970
Logistic Regression	0.993	0.98	1.000	0.990
SVC	0.993	1.00	0.980	0.990
LDA	0.936	0.88	0.936	0.907
Voting Classifier	1.000	1.00	1.000	1.000

Fig 19 Performance Evaluation For Adult Dataset

ML Model	Accuracy	Precision	Recall	F1-Score
AdaBoost	0.810	0.769	0.909	0.833
Random Forest	0.857	0.846	0.917	0.880
Decision Tree	0.810	0.769	0.909	0.833
KNN	0.857	1.000	0.812	0.897
Naive Bayes	0.762	0.769	0.833	0.800
Logistic Regression	1.000	1.000	1.000	1.000
SVC	1.000	1.000	1.000	1.000
LDA	0.857	0.846	0.917	0.880
Voting Classifier	0.952	0.923	1.000	0.960

Fig 19 Performance Evaluation For Adolescent Dataset

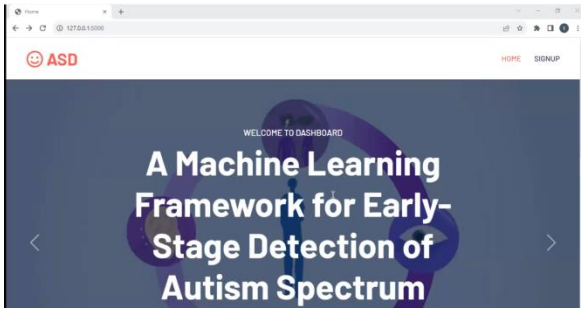

Fig 20 Home page

Fig 21 Signin page

Fig 22 Login page

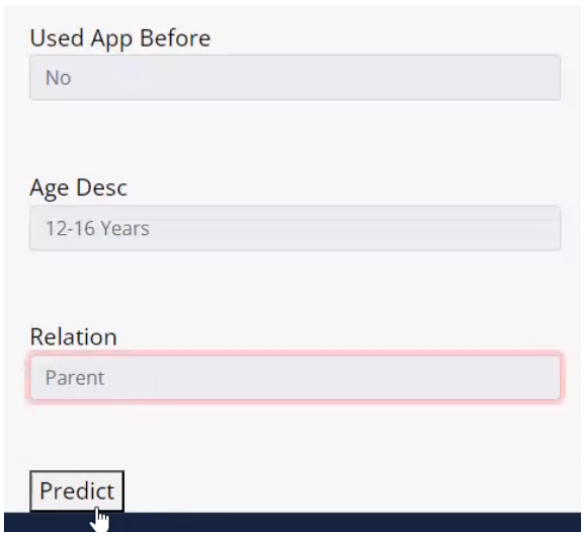

Fig 23 User input

Fig 24 Predict result for given input

5. CONCLUSION

By combining cutting-edge algorithms with feature scaling techniques, the research has effectively launched a novel machine learning framework for the early identification of autism spectrum disorder (ASD). The framework's strong performance across a range of age groups, as demonstrated by thorough testing on typical ASD datasets that include toddlers, adolescents, children, and adults, highlights its adaptability and possible clinical relevance [12,13]. The framework's best classification and feature scaling strategies are identified, providing a sophisticated and successful strategy for early ASD identification with possible ramifications for prompt therapies. In terms of ASD identification, the ensemble algorithm—which combines Random Forest and AdaBoost—has shown remarkable performance, attaining increased accuracy. Its practicality and efficacy in real-world applications are further demonstrated by its smooth integration into an intuitive front end where feature values may be entered and evaluated with ease. By utilising feature selection methodologies, the project offers informative attribute rankings that highlight important risk factors and significant characteristics that are essential for comprehending the intricacies of ASD and facilitating precise diagnosis.

6. FUTURE SCOPE

In order to enhance the identification of ASD and other neuro-developmental diseases, the project aims to gather more information about ASD and build a more universal prediction model for individuals of all ages [18]. This suggests that future research could include enlarging the study's dataset to encompass a more extensive and varied sample of people with ASD. In order to increase the precision and dependability of ASD identification, the project also recommends creating a more generalised prediction model, which may entail adding new machine learning methods or improving the current framework. The project's future scope may potentially include examining more neuro-developmental problems and looking at how the suggested framework may be used to identify and forecast them. The project's overall future scope encompasses more data collecting, model improvement, and possible extension to additional neuro-developmental problems.

REFERENCES

[1] M. Bala, M. H. Ali, M. S. Satu, K. F. Hasan, and M. A. Moni, ‘‘Efficient machine learning models for

early stage detection of autism spectrum disorder,'' Algorithms, vol. 15, no. 5, p. 166, May 2022.

[2] D. Pietrucci, A. Teofani, M. Milanesi, B. Fosso, L. Putignani, F. Messina, G. Pesole, A. Desideri, and G. Chillemi, ''Machine learning data analysis highlights the role of parasutterella and alloprevotella in autism spectrum disorders,'' Biomedicines, vol. 10, no. 8, p. 2028, Aug. 2022.

[3] R. Sreedasyam, A. Rao, N. Sachidanandan, N. Sampath, and S. K. Vasudevan, ''Aarya—A kinesthetic companion for children with autism spectrum disorder,'' J. Intell. Fuzzy Syst., vol. 32, no. 4, pp. 2971–2976, Mar. 2017.

[4] J. Amudha and H. Nandakumar, ''A fuzzy based eye gaze point estimation approach to study the task behavior in autism spectrum disorder,'' J. Intell. Fuzzy Syst., vol. 35, no. 2, pp. 1459–1469, Aug. 2018.

[5] H. Chahkandi Nejad, O. Khayat, and J. Razjouyan, ''Software development of an intelligent spirography test system for neurological disorder detection and quantification,'' J. Intell. Fuzzy Syst., vol. 28, no. 5, pp. 2149–2157, Jun. 2015.

[6] F. Z. Subah, K. Deb, P. K. Dhar, and T. Koshiba, ''A deep learning approach to predict autism spectrum disorder using multisite resting-state fMRI,'' Appl. Sci., vol. 11, no. 8, p. 3636, Apr. 2021.

[7] K.-F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis, ''The contribution of machine learning and eye-tracking technology in autism spectrum disorder research: A systematic review,'' Electronics, vol. 10, no. 23, p. 2982, Nov. 2021.

[8] I. A. Ahmed, E. M. Senan, T. H. Rassem, M. A. H. Ali, H. S. A. Shatnawi, S. M. Alwazer, and M. Alshahrani, ''Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques,'' Electronics, vol. 11, no. 4, p. 530, Feb. 2022.

[9] P. Sukumaran and K. Govardhanan, ''Towards voice based prediction and analysis of emotions in ASD children,'' J. Intell. Fuzzy Syst., vol. 41, no. 5, pp. 5317–5326, 2021.

[10] S. P. Abirami, G. Kousalya, and R. Karthick, ''Identification and exploration of facial expression in children with ASD in a contact less environment,'' J. Intell. Fuzzy Syst., vol. 36, no. 3, pp. 2033–2042, Mar. 2019.

[11] M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, ''Detecting autism spectrum disorder using machine learning techniques,'' Health Inf. Sci. Syst., vol. 9, no. 1, pp. 1–13, Dec. 2021.

[12] C. Allison, B. Auyeung, and S. Baron-Cohen, ''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,'' J. Amer. Acad. Child Adolescent Psychiatry, vol. 51, no. 2, pp. 202–212, 2012.

[13] F. Thabtah, F. Kamalov, and K. Rajab, ''A new computational intelligence approach to detect autistic features for autism screening,'' Int. J. Med. Inform., vol. 117, pp. 112–124, Sep. 2018.

[14] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, ''Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison,''

Comput. Biol. Med., vol. 136, Sep. 2021, Art. no. 104672.

[15] E. Dritsas and M. Trigka, ‘‘Stroke risk prediction with machine learning techniques,’’ Sensors, vol. 22, no. 13, p. 4670, Jun. 2022

[16] V. Chang, J. Bailey, Q. A. Xu, and Z. Sun, ‘‘Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms,’’ Neural Comput. Appl., early access, pp. 1–17, Mar. 2022.

[17] F. Thabtah, ‘‘Machine learning in autistic spectrum disorder behavioral research: A review and ways forward,’’ Inform. Health Social Care, vol. 44, no. 3, pp. 278–297, 2018.

[18] K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi, and M. N. Islam, ‘‘A machine learning approach to predict autism spectrum disorder,’’ in Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE), Feb. 2019, pp. 1–6.

[19] H. Abbas, F. Garberson, E. Glover, and D. P. Wall, ‘‘Machine learning approach for early detection of autism by combining questionnaire and home video screening,’’ J. Amer. Med. Informat. Assoc., vol. 25, no. 8, pp. 1000–1007, 2018.

[20] K. L. Goh, S. Morris, S. Rosalie, C. Foster, T. Falkmer, and T. Tan, ‘‘Typically developed adults and adults with autism spectrum disorder classification using centre of pressure measurements,’’ in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Mar. 2016, pp. 844–848.

[21] A. Crippa, C. Salvatore, P. Perego, S. Forti, M. Nobile, M. Molteni, and I. Castiglioni, ‘‘Use of machine learning to identify children with autism and their motor abnormalities,’’ J. Autism Develop. Disorders, vol. 45, no. 7, pp. 2146–2156, 2015.

[22] B. Tyagi, R. Mishra, and N. Bajpai, ‘‘Machine learning techniques to predict autism spectrum disorder,’’ in Proc. IEEE Punecon, Jun. 2019, pp. 1–5.

[23] F. Thabtah and D. Peebles, ‘‘A new machine learning model based on induction of rules for autism detection,’’ Health Informat. J., vol. 26, no. 1, pp. 264–286, Mar. 2020.

[24] M. Duda, R. Ma, N. Haber, and D. P. Wall, ‘‘Use of machine learning for behavioral distinction of autism and ADHD,’’ Transl. Psychiatry, vol. 6, no. 2, pp. e732–e732, Feb. 2016.

[25] S. B. Shuvo, J. Ghosh, and A. S. Oyshi, ‘‘A data mining based approach to predict autism spectrum disorder considering behavioral attributes,’’ in Proc. 10th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Jul. 2019, pp. 1–5.

[26] O. Altay and M. Ulas, ‘‘Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children,’’ in Proc. 6th Int. Symp. Digit. Forensic Secur. (ISDFS), Mar. 2018, pp. 1–4.

[27] F. N. Buyukoflaz and A. Ozturk, ‘‘Early autism diagnosis of children with machine learning algorithms,’’ in Proc. 26th Signal Process. Commun. Appl. Conf. (SIU), May 2018, pp. 1–4.

[28] M. F. Misman, A. A. Samah, F. A. Ezudin, H. A. Majid, Z. A. Shah, H. Hashim, and M. F. Harun, ‘‘Classification of adults with autism spectrum

disorder using deep neural network,'' in Proc. 1st Int. Conf. Artif. Intell. Data Sci. (AiDAS), Sep. 2019, pp. 29–34.

[29] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, ''Applications of support vector machine (SVM) learning in cancer genomics,'' Cancer Genomics Proteomics, vol. 15, no. 1, pp. 41–51, Jan./Feb. 2018.

[30] A. S. Haroon and T. Padma, ''An ensemble classification and binomial cumulative based PCA for diagnosis of Parkinson's disease and autism spectrum disorder,'' Int. J. Syst. Assurance Eng. Manage., early access, pp. 1–16, Jul. 2022.

[31] R. Abitha, S. M. Vennila, and I. M. Zaheer, ''Evolutionary multiobjective optimization of artificial neural network for classification of autism spectrum disorder screening,'' J. Supercomput., vol. 78, no. 9, pp. 11640–11656, Jun. 2022.

[32] M. Alsuliman and H. H. Al-Baity, ''Efficient diagnosis of autism with optimized machine learning models: An experimental analysis on genetic and personal characteristic datasets,'' Appl. Sci., vol. 12, no. 8, p. 3812, Apr. 2022.

[33] S. P. Kamma, S. Bano, G. L. Niharika, G. S. Chilukuri, and D. Ghanta, ''Cost-effective and efficient detection of autism from screening test data using light gradient boosting machine,'' in Intelligent Sustainable Systems. Singapore: Springer, pp. 777–789, 2022.

[34] U. Gupta, D. Gupta, and U. Agarwal, ''Analysis of randomization-based approaches for autism spectrum disorder,'' in Pattern Recognition and Data Analysis with Applications. Singapore: Springer, pp. 701–713, 2022.

[35] T. Akter, M. Shahriare Satu, M. I. Khan, M. H. Ali, S. Uddin, P. Lio, J. M. W. Quinn, and M. A. Moni, ''Machine learning-based models for early stage detection of autism spectrum disorders,'' IEEE Access, vol. 7, pp. 166509–166527, 2019.

[36] Kaggle. (2022). Autism Spectrum Disorder Detection Dataset for Toddlers. [Online]. Available: https://www.kaggle.com/fabdelja/autism-screeningfor-toddlers

[37] UCI. (2022). UCI Machine Learning Repository: Autistic Spectrum Disorder Screening Data for Adolescent Data Set. [Online]. Available: https://shorturl.at/fhxCZ

[38] UCI. (2022). UCI Machine Learning Repository: Autism Screening Adult Data Set. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult

[39] UCI. (2022). UCI Machine Learning Repository: Autistic Spectrum Disorder Screening Data for Children Data Set. [Online]. Available: https://shorturl.at/fiwLU

[40] D. Singh and B. Singh, ''Investigating the impact of data normalization on classification performance,'' Appl. Soft Comput., vol. 97, Dec. 2020, Art. no. 105524.

[41] D. Mease, A. J. Wyner, and A. Buja, ''Boosted classification trees and class probability/quantile estimation,'' J. Mach. Learn. Res., vol. 8, no. 3, pp. 409–439, 2007.

[42] Q. Wang, W. Cao, J. Guo, J. Ren, Y. Cheng, and D. N. Davis, ‘‘DMP_MI: An effective diabetes mellitus classification algorithm on imbalanced data with missing values,’’ IEEE Access, vol. 7, pp. 102232–102238, 2019.

[43] S. M. M. Hasan, M. A. Mamun, M. P. Uddin, and M. A. Hossain, ‘‘Comparative analysis of classification approaches for heart disease prediction,’’ in Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng. (ICME), Feb. 2018, pp. 1–4.

[44] D. Ramesh and Y. S. Katheria, ‘‘Ensemble method based predictive model for analyzing disease datasets: A predictive analysis approach,’’ Health Technol., vol. 9, no. 4, pp. 533–545, Aug. 2019.

[45] A. Arabameri and H. R. Pourghasemi, ‘‘Spatial modeling of gully erosion using linear and quadratic discriminant analyses in GIS and R,’’ in Spatial Modeling in GIS and R for Earth and Environmental Sciences. Amsterdam, The Netherlands: Elsevier, pp. 299–321, 2019.