Islamic University of Technology (IUT)

Autism Detection using Visual and Behavioral Data

Authors

Nafisa Sadaf - 160041060

Karishma Shaer - 160041070

Farhan M Nafis Momin - 160041074

Supervisor

Hasan Mahmud

Assistant Professor

CSE Department, IUT

A thesis submitted in partial fulfillment of the requirements for the degree of B. Sc. Engineering in Computer Science and Engineering

Academic Year: 2019-2020

Department of Computer Science and Engineering (CSE)

Islamic University of Technology (IUT)

A Subsidiary Organ of the Organization of Islamic Cooperation (OIC)

Dhaka, Bangladesh

April 12, 2021

Declaration of Authorship

This is to certify that the work presented in this thesis is the outcome of the analysis and experiments carried out by Nafisa Sadaf, Karishma Shaer and Farhan M Nafis Momin under the supervision of Hasan Mahmud, Assistant Professor, Department of Computer Science and Engineering (CSE), Islamic University of Technology (IUT), Dhaka, Bangladesh. It is also declared that neither of this thesis nor any part of this thesis has been submitted anywhere else for any degree or diploma. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Authors:

Name
Student ID - 160041060

Karishma

Name
Student ID - 160041070

Name
Student ID - 160041074

Autism Detection using Visual and Behavioral Data

Approved By:

Supervisor:

Hasan Mahmud

Assistant Professor

Department of Computer Science and Engineering (CSE)

Islamic University of Technology (IUT), OIC

Acknowledgement

We would like to express our grateful appreciation to **Hasan Mahmud**, Assistant Professor, Department of Computer Science & Engineering, IUT for being our adviser and mentor. His motivation, suggestions and insights for this research have been invaluable. Without his support and proper guidance this research would never have been possible. His valuable opinion, time and input provided throughout the thesis work, from first phase of thesis topics introduction, subject selection, proposing algorithm, modification till the project implementation and finalization which helped us to do our thesis work in proper way. We are really grateful to him.

We are also grateful to **Md. Kamrul Hasan**, PhD., Department of Computer Science & Engineering, IUT for his valuable inspection and suggestions on our proposal of autism detection using visual and behavioral data.

Autism Detection using Visual and Behavioral Data

Abstract

Diagnosing Autism Spectrum Disorder (ASD) can be difficult as there is no existing medical test for detecting Autism. The only clinical method for diagnosing ASD are standardized tests which require prolonged diagnostic time and can be expensive. Autism diagnosis can be formulated as a typical machine learning classification problem between ASD patients and a control group, which requires large datasets with different modalities to be trained on, in order to yield accurate results. However, the unavailability of such robust datasets stands as a threat to this automated diagnosis. To resolve this, we propose a method of Autism Detection using Visual and Behavioral Data. The proposed technique first relates the two datasets by generating common attributes, distributing them into sub classes and then combining them. Finally, decision level integration is performed on data of two modalities (visual and behavioral). The main contribution of our work can be summarized as follows: we have achieved an accuracy of 97.57% in autism detection from visual and behavioral data, which is higher than detection from only visual data, we have shown that combining data within sub classes based on common attributes is more accurate than combining them arbitrarily, and finally, we have introduced a novel, integrated dataset in the ASD domain.

Keywords: *Autism Spectrum Disorder, Diagnosis, Machine learning, Data integration, Behavioral data, Visual data*

Contents

1 Introduction

1.1 Overview

Autism Spectrum Disorder (ASD) is a developmental disability [1] that can cause significant social, communication and behavioral challenges. There are many subtypes of ASD, most influenced by a combination of genetic and environmental factors. Because autism is a spectrum disorder, each person with autism has a distinct set of strengths and challenges. People with autism learn, think and problem-solve at varying range of capabilities. It can range from being highly intelligent to being severely challenged. Multiple factors influence the development of autism and is often based around sensory sensitivities and medical issues such as gastrointestinal (GI) disorders, seizures or sleep disorders. Other factors can be mental health challenges such as anxiety, depression and attention issues.

Diagnosing autism spectrum disorder can be difficult because there are no singularly definitive medical tests, like a blood test, MRI scan etc, to diagnose the disorder properly. Currently, the only clinical method for diagnosing ASD are standardized tests which require prolonged diagnostic time and increased medical costs. Medical professionals look at child's developmental history and behavior to determine diagnosis. This is a completely manual process that requires vast experience and extensive knowledge of psychology and neuroscience.

Past research has shown that ASD can often be detected at early ages - 18 months or younger. By the age of 2, diagnosis by a medical professional can be considered very reliable. However, many children do not receive a final diagnosis until they are much older. This kind of long delay means that children with ASD would be unlikely to get the proper help they need. Treatment can begin earlier if ASD can get diagnosed early. Applied Behaviour Analysis, or ABA, is the application of the science of behaviour analysis that helps those on the autism spectrum to lead fulfilled lives. Early intervention programs, particularly for children, have been shown to yield significant benefits in academic achievements, social behavior, educational progression and attainment, while also improved labor market success.

Autism diagnosis can be formulated as a typical machine learning classification problem between ASD patients and a control group. Autism diagnosis can be formulated as a typical classification problem (i.e., ASD vs non-ASD), where the classifier is able to evaluate whether an unknown subject has ASD or not based on the input features. Currently it is possible to diagnose ASD patients with a 61.8% accuracy [2] using machine learning models using personal characteristic data such as age, sex, handedness, verbal and performance IQ. They conclude, high possibility that future efforts could include combining PCD with neuroimaging data using machine learning models with fruitful output. It has also been concluded that addition of other features, such as medical tests or past or family history of disease, might boost the performance of their models to a clinically useful level. Addition of more features could potentially increase the performance of neural networks and allow for uses of much more complex architectures for greater benefits.

In [2], they developed and compared nine automated machine learning models by using a large Personal Characteristic Data (PCD) dataset. One of the main drawbacks was the presence of heterogeneity in the dataset. Another was the small size of the dataset which affected the generalization performances of the models. The researchers in [3] conducted an experiment to perform video gesture analysis for ASD Detection. They recorded a set of video clips of reach-to-grasp actions performed by children with ASD and IQ-matched normal children. Huiyu Duan, Guangtao Zhai, Xingkuo Min, Zhaohui Che, Yi Fang, Xiaokang Yang, Jesus Gutierrez and Patrick Le Callet [4] designed a dataset on eye movement of children with ASD Disorder. According to their studies, eye movements encode and hold information about attention retention and psychological features of individuals, which could be of help to characterize traits of ASD. Dr Fadi developed a dataset [5] related to autism screening for toddlers that contained influential features to be utilised for further analysis especially in determining autistic traits and improving the classification of ASD cases. Gerry [6] built an image dataset of ASD and non-ASD children; he extracted and used the facial features from the images of the dataset and fed them into classifier.

Our goal is to integrate multiple ASD datasets of different modalities into one efficiently large dataset with higher quality that can be used in further research works on diagnosis tools and higher accuracy ASD screening. Early detection of ASD is extremely important to a child's development. We believe the existence of such an enriched dataset can enhance the accuracy of ASD classifiers.

1.2 Problem Statement

The automated diagnosis of ASD at earlier ages aims to facilitate the treatment services available to children. Currently, the manual diagnosis requires professional/medical assistance in assessing whether a child has ASD or not. It requires direct interaction with medical professionals to check a child's developmental history, responsiveness, behaviour, attention span, and intelligence. By the age of three, children with ASD are more likely to have motor, communication, as well as sensory problems – and increasingly poorer emotional and social health. Advanced and robust classification methods play an important role in the early detection and diagnosis of ASD in children. The rapid growth in the number of ASD cases worldwide necessitates datasets with various traits. However, such datasets are rare, thus making it difficult to perform thorough analyses to improve the efficiency, sensitivity, specificity and predictive accuracy of the ASD screening process. Presently, very limited autism datasets are available and most of them are of one individual trait like either visual or behavioral or statistical. An efficient dataset can serve as a significant and useful tool in the generation of such models. An automated diagnostic tool that accurately classifies a child as ASD or non-ASD can replace the costly manual diagnosis. It can be used by parents to initially diagnose their child at home, before they can consult a medical professional or seek early childhood intervention programs.

1.3 Motivation and scope of research

The integration of multiple datasets however remains a key challenge in Autism Spectrum Disorder. In [7] they used an integrative approach to analyze multiple omics data and summarized their ability to address applications such as disease subtyping, biomarker prediction, and deriving insights into the data. Some of the problems that they faced were the underlying heterogeneity in individual omics data and the large size of data. In [8] a very large number of datasets were integrated at the same time by making use of six artificially constructed time series datasets. They concluded that it successfully captures the underlying structural similarity between the datasets. Multiple Dataset Integration (MDI) was performed, which is an unsupervised integrative modelling of multiple datasets. It is a novel Bayesian method. In [9] they developed a joint latent variable model for integrative clustering called iCluster. Flexible modeling of the associations between different data types is enabled by iCluster. The variance–covariance structure within data types in a single framework is also found, while the dimensionality of the datasets is reduced at the same time. But there is no existing work done for multiple ASD data integration. The main objective of our work is to focus on the lack of multivariate ASD datasets. The efficiency of any diagnostic classifier of ASD heavily depends on the size and diversity of the data available.

This publication [10] has very promising preliminary results with an accuracy of 94.6%. But the image dataset they used had duplicate images, improper age ranges, and lack of validation about the conditions of the individuals in each photo. According to them, improving the data set could yield better results, which we attempted to fix in our experiment. Sufficiently large datasets on ASD are not easily available, and even if they are large enough, they mostly contain data of singular modality, like only images, or only Personal Characteristic Data (PCD), or video, or hand gestures etc. The two datasets that we found most promising for our experiment have been collected from subjects that are similar in age group, gender and ethnic background. From the work done in [2], integrated features can improve Autism detection. In [11], PCD and rs-fMRI taken from the same subjects have been successfully integrated. Lastly, an integrated model using both Visual and Behavioral Data for diagnosis has not been done before in the domain of autism.

1.4 Research Challenges

Detection of ASD via automated computerized methods encounters a range of problems, while Data Integration stands to its own complications. ASD patients often exhibit much of external traits. Among ASD patients, if we look at language skills for example, these skills span the entire continuum- ranging from no speech or functional speech, to high levels of comprehension, production, and literacy, while autism severity ranges from very severe to very mild. The greatest variability, however, occurs in their intellectual abilities (IQ) and cognitive profiles (verbal vs. nonverbal IQ). Although IQ appears as a strong predictor of social and adaptive functioning in young children, adaptive behavior was determined to be most dysfunctional in older children, despite IQ, indicating the wisdom in employing ongoing surveillance with these children.

Despite being classified as a heritable biological condition, to the best of our knowledge, there has not been a medical or genetic test to verify the existence of ASD as for fragile X syndrome (FXS), Williams syndrome, and Down syndrome. The syndromes mentioned are marked with intellectual and language difficulties and they may occur with ASD, particularly in individuals with FXS. The ASD diagnosis on the other hand, is based on observations of the behavior and reports based on it. Depending on cognitive function and assessment age, children with ASD demonstrate widely varied phenotypes, and wide-ranging symptom severity and developmental paths, making accurate and stable ASD identification a challenging process, especially in early development. Charman [12] notes that although desirable, reliable early identification is hampered due to regression or loss of skills; by age 2 they may exhibit reduced orienting to name, poorer joint attention, some early motor abnormalities, and reduced emotional expression. These are pointers that are observed by medical professionals.

While ASD has no known treatment, regular therapy sessions and check-ups are advised, that can be very costly overtime. So for an automated or computerized model to be developed related to ASD, there needs to be accurate and large dataset collection. Unfortunately these kinds of datasets are quite scare and most require advanced permissions that are inaccessible for most individuals, as it can be concern of privacy breach for most families. Such factors make development quite challenging.

We have observed the existence of individual feature datasets that focus on a wide range of attributes ranging from eye movement, genotype and phenotype data, digital image datasets, video gestures, personal characteristics data, statistical data of toddlers, emotional and behavioral information, attention and psychological factors. These visual or behavioral, statistical or other types of datasets can provide more effective information if two or more of them can be successfully integrated. The vast majority of classification studies in ASD have trained and tested predictive models within relatively small research samples using cross validation methods. Finding a method of integration that works for such a huge range of data types is our biggest challenge. As we have seen in the study using pattern classification, heterogeneity in samples can be taken into account through normative approaches that allow for complexity. Such normative approaches could be integrated into our work too.

After merging the datasets, a suitable validation methodology is also crucial. It will provide the evidence of whether the integrated dataset is better at building a predictive model than the individual datasets. In other words, metrics such as 'overall accuracy' can be highly study specific and not generalizable to the wider population or across the autism spectrum and thus should be interpreted with a certain degree of caution. In this publication [13], real world data was used as a predictive quantifier along with comparison of the merged data against the original. Simulation was an important validation method here too, since it was reproducible and scalable for integration of any number of datasets. Validation approaches such as these can make our results more credible.

Since image dataset has been promising in ASD research, there is emerging interest in combining different imaging modalities together in single classification models. Integrating different modules of ASD datasets into one for ASD screening will pose issues in storage of the large amount of data and creation of the dataset without losing valuable information from the constituent sets. 3D facial images of autistic and non-autistic children have created objective accurate measurements, but such methods are not generally available in most image datasets. The sex and age of the individual also has an influence on diagnosis using image data, so sexually dimorphic facial features and proper division according to age should be taken into account. Organic dataset derived from real-life environments increases the accuracy and robustness of image classification algorithms.

1.5 Thesis Outline

In Chapter 1 we have discussed our study in a precise and concise manner. Chapter 2 deals with the necessary literature review for our study and the developments so far. In Chapter 3 we have stated the skeleton of our proposed method and provided a flowchart to give a detailed insight to the working procedure of our proposed method of Autism Detection using Visual and Behavioral Data. Chapter 4 shows the experimental results and comparative analysis of successful implementation of our proposed method. The final segment of this study contains all the references and credits used.

2 Background Study and Literature Review

2.1 Autism Spectrum Disorder (ASD)

Recent advances in the study of ASD may pave the way for groundbreaking future treatment. We know that ASD is a neurodevelopmental condition characterized by deficits in social interaction and communication, as well as repetitive and restrictive behaviours. Autistic patients struggle with social skills, repetitive behaviors, and communication, both verbal and nonverbal. With these symptoms that are often utilized for the diagnosis of ASD by both the experts and data driven algorithms, we look to the recent studies that have shed light on identifying some of the most robust features for detection of Autism at an early age.

According to [14], investigations into the facial structure of individuals of ASD have the potential to reveal greater insights into the biological pathways leading to autism. A dataset containing the facial images of children were used as the adult facial morphology may be influenced by pubertal hormone actions. Facial phenotypes are widely used in the clinical diagnosis of numerous neurodevelopmental and other disorders, and provide an important research tool for identifying rare diseases and links between genotype and phenotype. Facial images were acquired using a 3dMDface system. This study used landmark data from 3D images of autistic and non-autistic children to create objective measurements of sexually dimorphic facial features. For each sex, increased facial masculinity was observed in the ASD group relative to control group. Furthermore, increased facial masculinity in the ASD group correlated with more social-communication difficulties. Sample sizes in the study are small, this limits the ability to generalise the current findings to the wider population. Nevertheless, sexual dimorphism in the three linear features found in this study have been previously reported in other studies too, so the findings are in line with the generally accepted hypothesis.

The study in [10] reveals that patients have a common pattern of distinct facial deformities, allowing researchers to analyze only an image of the child to determine if the child has the disease. The deep learning model used in this work uses MobileNet and two dense layers in order to perform feature extraction and image classification. The model is trained and tested using both images of children with autism and children without it. The first dense layer is dedicated to distribution, and allows customisation of weights to input into the second dense layer. Thus, the second dense layer allows for classification of the images. We have used a similar model in our experiment with the image data. One of the challenges faced by earlier work on image classification was the fact that the age cutoffs for the subjects did not typically have hard divisions, but for the sake of this study, it is necessary. This creates issues when individuals are right at the boundary of two age groups. To prevent similar issues with this experiment, the images are simply classified as "Autistic" and "Non-Autistic" rather than additionally classifying the levels of autism. Another issue was that many of the proven studies in autism diagnosis used ideal images, most of which were frontal, occlusion-free, with a clean background, consistent lighting, and limited facial expressions. In this study, contrary to what was done before, the facial images are derived from real-life environments and the dataset was constructed organically.

Using eye movement tracking to create a diagnostic tool for ASD patients has already been mentioned. On this topic, work had been done in [15]. A theoretical framework with eye tracking as a linking method to relate aberrant social cognition and empathy with neurocognitive processes and potential endophenotypes had been proposed. In this initial work, the multidimensionality of social cognition was not replicated and sufficient fits of empathy for proposed structure equations were not retrieved. Small sample sizes also limited the representativeness of the findings. These limitations had been addressed in later work on the creation of a dataset of eye movements for the children with autism spectrum disorder [4]. Here, the attention and psychological factors of an individual, which could help to characterize the traits of ASD, were taken into account, diminishing the limitations of the previous publication. Predicting autism diagnosis using image with fixations and synthetic saccade patterns has been attempted in [16], though the dataset was small, so some data argumentation methods have been applied to alleviate the overfitting problem. The color for the input image is jittered and random noise is added to the location of data points.

Eye gaze has also been shown to be a useful attribute in detection of ASD in [17]. It proposes the formation of a dual eye-tracking setup capable of recording gaze with high resolution was used to investigate how gaze behavior in interaction is related to traits of Autism Spectrum Disorder (ASD), and Social Anxiety Disorder (SAD). Eye gaze tracking is a sub-system of facial information processing. One big disadvantage is that the generalizability of these findings to social settings may be limited. The main argument is that in social interaction, the content of faces is more dynamic and dependent on the interplay between interaction partners, than the content of a non-responsive face (e.g. pictures or videos) as portrayed in a typical experiment. As clinical ASD and SAD groups have exhibited deficiencies in reciprocal social behavior (social behaviour and actions that are done as a response to another person's action), traits of these two conditions were assessed in a general population. The study reports that gaze behavior in interaction of individuals who score high on ASD and SAD traits corroborates hypotheses posed in typical face-processing research using non-responsive stimuli. Moreover, the findings on the relation between paired gaze states (when and how often pairs look at each other's eyes simultaneously or alternately) and ASD and SAD traits bear resemblance to prevailing models in the ASD literature (the 'gaze aversion' model) and SAD literature (the 'vigilant-avoidance' model). Pair-based analyses of gaze may reveal behavioral patterns crucial to our understanding of ASD and SAD, and more general to our understanding of eye movements as social signals in interaction.

2.2 Machine Learning in ASD Detection

In addition to the above mentioned approaches, there are multiple further methods of detecting ASD patients using machine learning. Emotional and behavioral information has been used in identifying the likelihood of autism in children in [18].

This publication has investigated and compared emotional and behavioural profiles in children with ASD aged 6–15 years and explored to what extent language problems contributed to the emotional and behavioural needs (EBN). Specialized questionnaires (SDQ) and checklists (CCC-2) were filled out by parents. Statistical methods like Personal Characteristic Data (PCD) dataset and video gesture analysis have already been talked about earlier. Pattern Classification has been used in [19] to identify brain imaging markers (neurological markers) in ASD. A variety of these techniques have been increasingly applied in classification models of ASD across a wide range of neuroimaging features. Heterogeneity in clinical phenotypes is missing in most samples pertaining to ASD. Several overlapping symptoms may hinder model performance in real world scenarios. In order to tackle these issues, a normative approach is taken that is not limited to traditional diagnostic categories but can accommodate the complexities of an ASD population like heterogenous phenotypes, misdiagnosed individuals and related comorbidities.

Now that we have looked into the wide range of methods and datasets available for research on Autism detection, it is clear that there are multiple datasets which have a wide range of data types present to create predictive models on. Our core goal for research is to create a method for creating an efficient integrated dataset that can serve as a tool that significantly increases the accuracy in the generation of classification models. Large amounts of data afford simple models much more power; outliers are easier to classify and the underlying distribution of that data is clearer. With more data we can use X-fold cross validation to build models which may have a better predictive power, which may not really be possible with very limited data. Let us take an overview of the methods we can consider for our solution approach to integrate datasets.

2.3 ASD Datasets

Researchers of [3] conducted an experiment to perform video gesture analysis for ASD detection. They recorded a set of video clips of reach-to-grasp actions performed by children with ASD and IQ-matched typically developing (TD) children and made a video gesture dataset. Their dataset consists of 40 subjects (ASD: 20, non-ASD: 20). Another dataset, which we used in our experiments, contains facial images of ASD and non-ASD children [6]. It has 2936 instances in total (ASD: 1418). Each image is in 224 x 224 x 3 jpg format. In [4], they designed a dataset of eye movements of children with ASD disorder. It consists of 300 natural scene images and the corresponding eye movement data collected from 14 children with ASD and 14 healthy controls. In particular, fxation maps and scanpaths are available in the dataset. Autism Brain Imaging Data Exchange (ABIDE) [20] has aggregated functional and structural brain imaging data collected from laboratories around the world to accelerate the understanding of the neural bases of autism. ABIDE I contains resting state functional magnetic resonance imaging (R-fMRI), anatomical and phenotypic datasets. ABIDE II consists of over 1000 additional datasets with greater phenotypic characterization, particularly in regard to measures of core ASD and associated symptoms. ALong with these there is another Personal Characteristic Data (PCD) dataset containing 1114 instances from 521 individuals with ASD and 593 controls (age range: 5-64 years). The PCD features include age, sex, handedness, performance IQ, verbal IQ, full IQ etc. [5] contains a dataset related to autism screening of toddlers which we have also used in our experiments. This dataset was also used in [21]. It consists of behavioral data with 1054 instances in total and has 18 attributes including the class label. The attributes include Q-Chat-10-Toddler feature scores (A1-A10) collected through the Q-Chat questionnaire application, age, sex, ethnicity, family member with ASD history etc. Another dataset called Autism Adult Screening Data contains 704 instances (189 - ASD) which was used in [22].

2.4 Gender and Race Prediction

Gender and race predictions of unfiltered faces classify unconstrained real-world facial images into predefined age and gender. Significant improvements have been made in this research area due to its usefulness in intelligent real-world applications. More recently, Convolutional Neural Networks (CNNs) based methods have been extensively used for classification due to their excellent performance in facial analysis. Four widely used face detection tools, which are Face++, IBM Bluemix Visual Recognition, AWS Rekognition, and Microsoft Azure Face API were evaluated using multiple datasets to determine their accuracy in inferring user attributes, including gender, race, and age [23]. Results show that the tools are generally proficient at determining gender, with accuracy rates greater than 90%, except for IBM Bluemix. Concerning race, only one of the four tools provides this capability, Face++, with an accuracy rate of greater than 90%, although the evaluation was performed on a high-quality dataset. All tools infer the attributes of age and gender, and Face++ can detect ethnicity as well. However, its ethnicity classification is limited to white, black, or Asian; thus, it is actually attempting to detect race, not ethnicity, as stated. Within a race, there can be many ethnicities, like "Asian" can include "Korean" and "Japanese." Noisy data used for such classification refers to images of low quality or where the frontal of the face is not fully revealed in the image. In the recent work by A.K. Vani, faces and the facial features which include eyes, mouth, and nose are detected using Haar Cascade based on Viola-Jones face detection algorithm [24]. Before gender detection, the noise is reduced by applying adaptive filters, thereby increasing the accuracy. The obtained facial features are given as the input or test data to the neural network. The neural network is designed to obtain the features and acts as a classifier to detect the genders. In [25] a novel end-to-end CNN approach is proposed to achieve robust age group and gender classification of unfiltered real-world faces. The two-level CNN architecture includes feature extraction and classification itself. The feature extraction extracts features corresponding to age and gender, while the classification classifies the face images to the correct age group and gender. The EGA database [26] is an integrated database consisting of images from CASIA-Face V5, FEI, FERET, FRGC, JAFFE and Indian Face Dataset . The EGA database has five ethnicities: a) African-American, b) Asian c) Caucasian, d) Indian, e) Latinos. There are two genders for each ethnicity: a) male; b) female and three age ranges: a) young, b) adult and c) middle-aged. For finding the relationship with the gander and age of toddlers, the "Young EGA" database would be most useful. There are 153 pictures in the "young" category of EGA. EGA dataset is balanced with respect to subject gender with 52.4% male and 47.6% female. It is slightly less balanced with respect to age with 32.6% young, 48,5% adult and 18,9% middle-aged. To obtain the EGA dataset the interested user has to obtain the individual permissions from the data set sources first (which are all freely available for research purposes). Then a set of scripts are provided, which allow, from the images in their original format, to rebuild the entire structure of the EGA dataset, by making an appropriate re-allocation of names to the files concerned. One such dataset that is present in the EGA database is the FEI dataset [27] which includes 2,800 images of 200 subjects, 100 male e 100 female, each with 14 images. These are all colour images with a resolution of 640×480. The faces have been acquired on a white background in FEI Laboratory in São Bernardo do Campo, São Paulo, Brazil, and belong to subjects from 19 to 40 years old, mostly of Latin ethnicity. The features that this dataset mainly focuses on are age, gender and expression. The main drawback of the FEI dataset with respect to our work is that it does not contain any data of toddlers. Another promising dataset is FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age dataset [28] which contains 108,501 images, with an emphasis of balanced race composition in the dataset. It consists of 7 race groups: White, Black, Indian, East Asian, Southeast Asian, Middle East, and Latino. All the data are labeled with race, gender, and age groups with an age range of 0 to 70+ years. The ratio of toddlers to adult in this dataset is quite small with only 5730 out of 108,501 images being of toddlers. The dataset that we found to be most promising for our work is the UTK Dataset, a large-scale face dataset that was used in [29]. It is composed of over 20,000 images with their respective annotations of age, gender and race. The age ranges from 0 to 116 years. Compared to the FairFace dataset of [28], UTK Dataset has a higher ratio of images of toddlers to adults. The ethnicities are as follows: White (42.5%), Black (19.1%), Asian (14.5%), Indian (16.8%) and

Others (7.14%). As is evident, similar to our visual and behavioral datasets, the UTK Dataset is also white dominant, thus making it consistent with our work. It has an almost equal distribution of gender with males comprising 52.3% and females 47.7% of the entire data. In a study by Caifeng Shan [30], researchers used Local Binary Patterns (LBP) to describe faces. Through the application of support vector machines (SVM), they were able to achieve a 94.81% success rate in determining the gender of the subject. The main breakthrough of this study was its ability to use only real-life images in their classification.

2.5 Data Integration

According to [31], modern data generated in many fields are in a strong need of integrative machine learning models in order to better make use of heterogeneous information in decision making and knowledge discovery. Data from multiple sources can be incorporated in a learning system and that is a key step for a successful analysis. On the one hand, this provides us an opportunity to understand a complex object (or system or patient) from multiple angles and make precise data-driven decisions. On the other hand, it poses a challenge to wisely optimize the use of these data. Intelligent learning systems play crucial roles while dealing with big data. Data fusion techniques can be classified into three categories: early, intermediate, and late integration. In early integration methods, all features are concatenated into a vector and then fitted onto an unsupervised or a supervised model. In late integration methods, separate models are first learned using their corresponding feature subsets, and then their outputs are further combined to make the final determination. An intermediate strategy involves data integration in a learning process. Bayesian methods are promising because from a Bayesian perspective, we can consider treating some features as prior knowledge. Decision trees for classification or regression are representatives of rule-based learning. Decision trees can be applied to select features by ranking the features with respect to their summed improvements in class purity. In a decision tree with T internal nodes, the importance score of the i-th feature can be defined. The values of continuous variables are partitioned into intervals of different lengths, thus decision rules can be created for continuous variables of a variety of distributions. There is no need to standardize the input data. In fact, decision trees are invariant under feature scaling and transformation. However, decision trees are notorious with their high risk of overfitting, resulting into pruning being a necessary remedy. Moreover, building a decision tree for high-dimensional data is very much time-consuming.

Ensemble learning has been one of the influential ideas behind data integration. As a successful example of collective intelligence, ensemble learning, builds a population of weak learners for a much better performance. Bagging and Boosting are popular ensemble learning models where decision trees are often used as weak learners. While bagging simply combining the decisions of multiple weak learners, boosting tweaks the weak learners to focus on hard examples. There are three ways to integrate data through ensemble learning. The first way involves using the concatenated features as input to a random forest. The second way requires to build multiple trees for each data source, and then use all trees of all data sources to reach the final decision via voting. Ensemble-learning based data integration is considered to be advantegeous mostly due its good manipulability and interpretability. Class imbalance problems can be addressed by random forest in its bootstrapping. Also, granularity of features can be carefully considered while sampling features. However, since it is a late-integration principle, the interactions of features from separate sources cannot be detected. In the third way, meta-features learned from different sources can be used to grow trees. This idea can be inferred from West's group, where they incorporated both clinical factors and genomic data in predictive survival assessments. A meta-feature (named meta-gene) is defined as the first principal component of a cluster of genes grouped by k mean clustering. Then, the model grows a forest of statistical classification and prediction trees. In each tree, features used in the nodes are decided by the significance of Bayesian factor tests on the features (meta-genes and clinical factors). Multiple significant features can be distributed in multiple trees so that the correlations between trees are reduced. The final decision is determined by a weighted combination of the decisions of all trees, where the probabilities of trees are used as weights. One advantage of the meta-feature based ensemble model is that the information from different sources can be incorporated in the model learning. Since the meta-features are used instead of the original features, complexity of trees can thus be reduced.

A comparative study [32] considered only decision level fusion techniques and observed that the performance of the highest confidence (HC) algorithm is the best among all the algorithms taken into account. It had an accuracy of 95.08 per cent rate, thus proving the usefulness of dynamic reliability estimation by the relative distance between classes. Arithmetic rules perform better than both abstract- and rank-level schemes, as SUM and PRODUCT rules attained 95.47 and 95.34 per cent average identification rates, respectively.

The individual similarity (kernel) matrix of a data source can be computed by an off-the shelf kernel function semantically sensible for the specific data source or by sophisticated metric learning. Metric learning aims to learn a metric function from data such that the distances between within-class samples are closer, and the distances between inter-class samples are farther. The main strength of kernel methods is that their optimizations are independent of the number of features. This is known as being dimension-free. However, optimization corresponding to a large sample size still remains the a major bottleneck. For example, an optimal MKL learning may be essentially a semi definite programming problem.

There are several benefits of using feature extraction in data integration. First, the natures of heterogeneous data from multiple sources can be separately well considered. In spite of the original data types, the new features in the corresponding feature spaces are usually numeric. Second, the high-dimensionality is dramatically reduced so that the downstream analysis will be more efficient. Third, extracting features separately for each data source implements the principle of divide-and conquer, thus computational complexity can be significantly reduced. Fourth, relational data can be well incorporated with the help of kernel feature extraction methods. However, one drawback of the feature-extraction based integrative principle is that the interactions (correlation) between features from different sources cannot be taken into consideration in the feature extraction procedures.

Deep neural networks can be used for multi-modal learning and it can be extended to data integration too. Different deep learning models can be applied to the individual data sources. The integrative network can combine the information from the sub-networks. The model can be either directed or undirected; either supervised or unsupervised. The deep neural network based multi-modal structure can merge the output of individual sub-networks in higher layers. The sub-networks provide the flexibility of choosing appropriate deep learning models respectively for individual data sources, such as deep belief net for binary data, convolutional network for image data, recurrent neural network for speech signal, and deep feature selection for choosing discriminative features.

Bayesian data integration transcriptional modules have been briefly discussed earlier. These modules have been able to integrate gene expression and transcription factor binding (ChIP-chip) data in [8]. Algorithms for Learning Kernels Based on Centered Alignment [33] is an integration method developed at Google Research. In the past, the uniform combination solution has proven to be difficult to improve upon. The newly proposed algorithms are based on the notion of centered alignment which is used as a similarity measure between kernels or kernel matrices. A number of novel algorithmic, theoretical, and empirical results for learning kernels based on centered alignment are presented. The problem is reduced to a simple QP and a one-stage algorithm for learning both a kernel and a hypothesis based on that kernel using an alignment-based regularization.

Two key concepts that underlie most integrated models are spatial point processes and joint likelihood methods, as illustrated in [34]. A spatial point process describes the distribution of event locations across some spatial domain (e.g. locations of individuals of the same species). Points arise from a random process, described by the local intensity which measures the expected density of points at a given location in space. It is possible to use the toolbox developed for modelling spatial point processes to estimating distributions for many data types. The result suggests that if it is possible to link multiple data types to a common point process, it is possible to combine the methods in an integrated estimator. It is also possible to develop shared parameterizations across a wide set of data types and spatial resolutions. This may involve different measured response variables (e.g. probability of occurrence vs. local density) and responses measured at different spatial resolutions.

As we look into a very different field, the field of satellite data, methodologies on data integration that are used to merge multiple types of data in that field can provide insight into the field of Autism detection at an early age. A new method for estimation of spatially distributed rainfall through merging satellite observations, rain gauge records, and terrain digital elevation model data has been formed by in [13]. The merged rainfall data was tested and confirmed using the information on the 21 July 2012 Beijing flood. In this method, pairwise comparisons of the rain gauge record against various satellite observation datasets are implemented to determine the most suitable satellite observation. For testing the validity of merged data, there needs to be a robust method for validation and comparison. Validation of the data merging method is conducted through comparing the merged rainfall data against the two original rainfall datasets (i.e., the rain gauge record and the satellite observation) and another merged rainfall dataset without consideration of the third dataset (elevation influence). To further validate the method, the merged rainfall data and the original datasets are used to drive a hydrological model and to simulate the streamflow. This simulation method for validation can be useful for further work on data integration since it is reproducible and scalable for integration of any number of datasets.

A network embedding-based multiple information integration (NEMII) method for the MiRNA-disease association prediction has been developed in [35]. Firstly, miRNA-family associations are used to represent miRNAs; disease-disease similarity is calculated and then diseases are represented using this similarity. Then diseases and their relationships can be transformed into a directed acyclic graph (DAG). The DAGs can then be used to calculate disease semantic similarity. Secondly, the known miRNA-disease associations are formulated as a bipartite network, and node embeddings in the bipartite network are learned by using SDNE which in turn are used to represent miRNAs and diseases. Thirdly, all representations of miRNAs and diseases are combined to represent miRNA-disease pairs. Finally, a prediction model is constructed based on the miRNA-disease pairs by using random forest. Random forest is well-known for its ability to deal with unbalanced datasets. Experimental results reveal that NEMII performs better than the models using biological features alone and models using embedding representations alone, and SDNE produces better results than using other network embedding methods. NEMII can be a useful technique for creating a predictive dataset for ASD too, allowing similar methodology to be adopted.

In [36], they proposed a method of data integration by combining big data and survey sample data. By stratifying the population into a big data stratum and a missing data stratum, they can estimate the missing data stratum by using a fully responding probability sample, and hence the population as a whole by using a data integration estimator. By expressing the data integration estimator as a regression estimator, they can handle measurement errors in the variables in big data and also in the probability sample. They also propose a fully nonparametric classification method for identifying the overlapping units and develop a bias-corrected data integration estimator under misclassification errors. Finally, they develop a two-step regression data integration estimator to deal with measurement errors in the probability sample. An advantage of the approach is that there are no unrealistic missing-at-random assumptions for the methods to work. This is a novel method of data integration for handling big data by incorporating survey sample data. The proposed method is applied to the real data example and validated.

3 Proposed Approach

3.1 System Architecture

Initially common data clusters are identified between the two classes by filtering the records that have the same race and gender in both datasets. The records corresponding to each cluster are merged to obtain the integrated dataset. The dataset is then trained through two binary classifiers corresponding to each modality of the data to generate class scores. These class scores are then combined using three different methods and the resulting binary classification scores are evaluated.

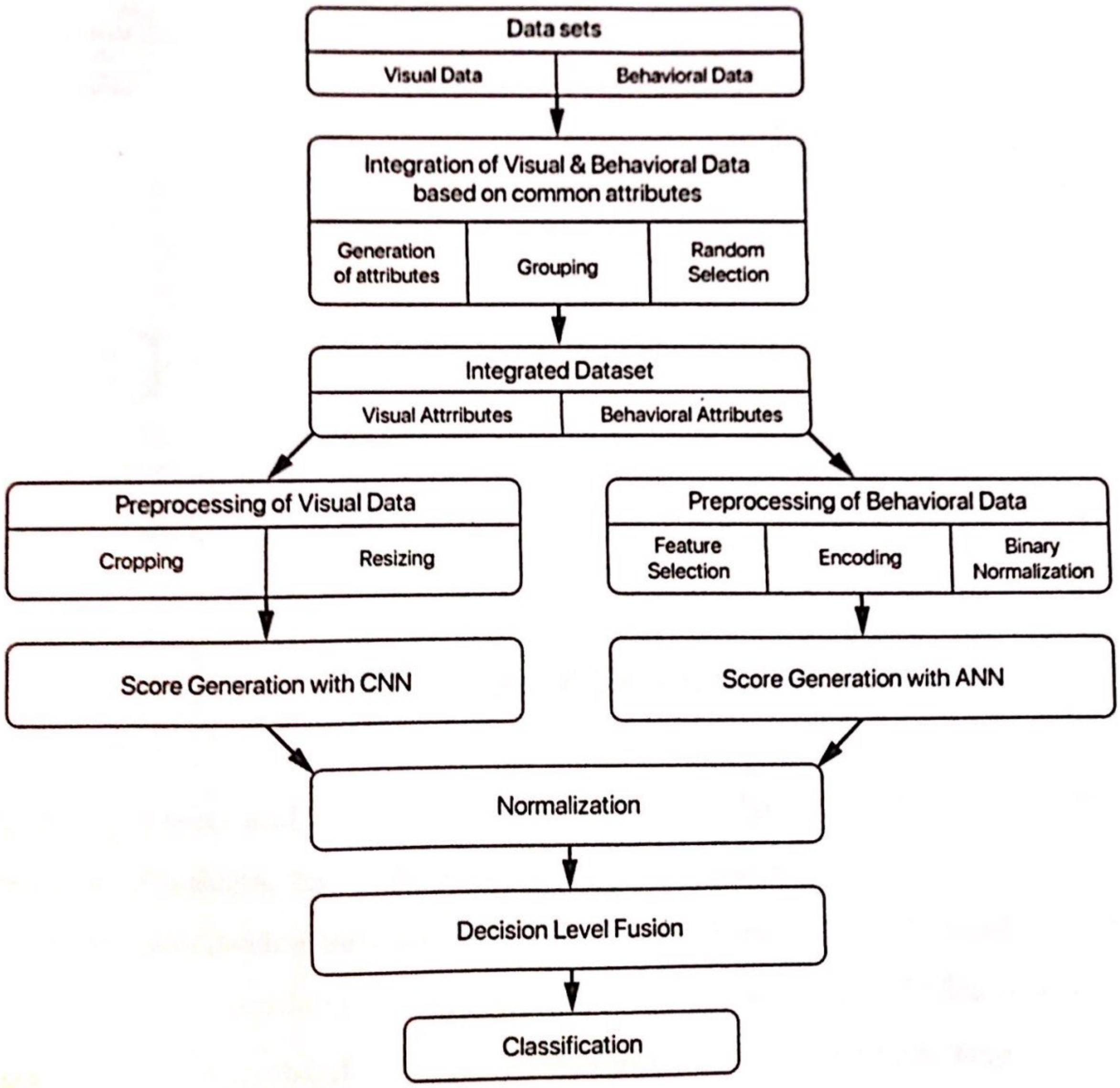

Figure 1: Proposed Approach

3.2 Integrated Record Generation

The image dataset being used in the proposed approach was obtained from Kaggle, which consists of 3014 images of both autistic and non-autistic children, equally divided in each class. The behavioral dataset is related to autism screening of toddlers and contains influential features to be utilised for further analysis especially in determining autistic traits and improving the classification of ASD cases. The dataset has 1054 records, ten behavioural features (Q-Chat-10) and other individual characteristics that have proven to be effective in detecting the ASD cases in behaviour science.

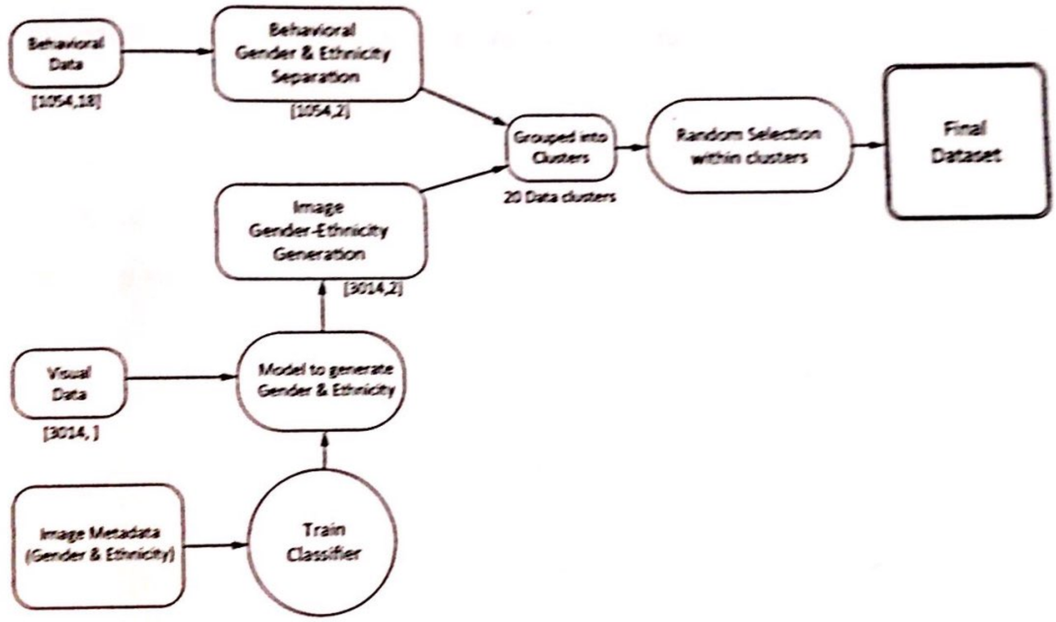

Figure 2: Integrated Dataset Generation

Since the Visual and Behavioral Data being used have no common records and different modalities, two different methods of integration are required that approximate the relationship between the two datasets. Firstly, the relationship between the individual records of the two datasets are approximated. To find the association between two sets of data, predictive common attributes are generated using the visual data. The attributes that are generated are the Gender and the Race of the subjects in each of the facial images. Gender is classified into two classes:

a) Male and b) Female. Race is classified into five classes: a) White, b) Black c) Asian, d) Indian, e) Others. In this process, a large-scale face dataset called the UTK Dataset [37] is used as our image metadat that consists of over 20,000 images, to train a multi-output CNN model that will output the predicted gender and race. The multi-output CNN model consists of three branches: the age, gender and race branch respectively. Each branch follows a similar set of hidden layers: Conv2D with ReLu activation followed by BatchNormalization followed by Pooling followed by Dropout. This is then connected to two fully connected layers: the first with 128 nodes and the second has number of nodes equal to the number of classes (for instance, for gender branch there are 2 classes and for race branch there are 5 branches). The entire architecture of the multi-output CNN model is given below. After training, our visual data is fed into this model to obtain the predicted gender and race of each image. Apart from these two predictions, another attribute that is used to associate the records in the datasets is the binary class label: a) Autistic and b) Non Autistic. This procedure of visual metadata generation is depicted in the diagram below.

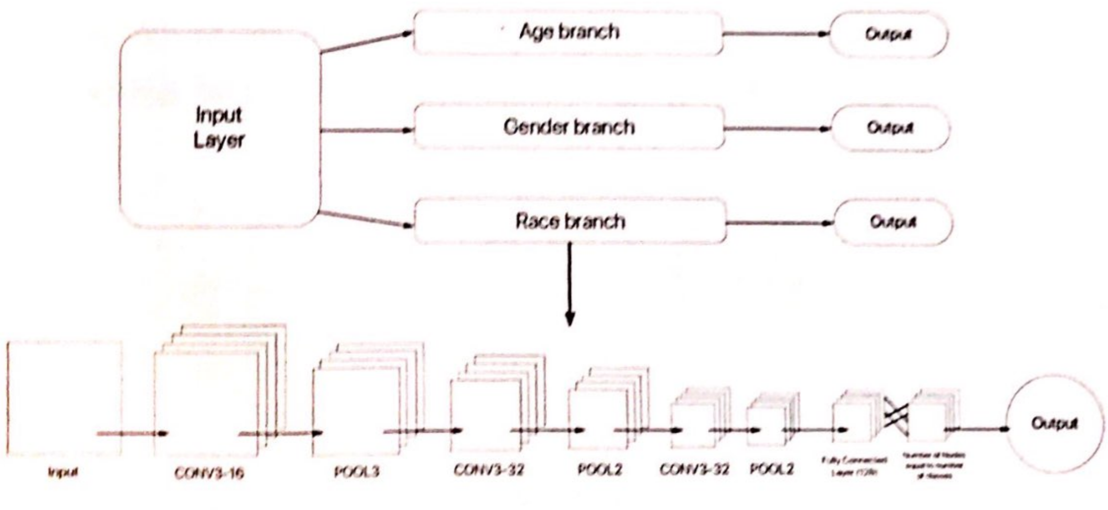

Figure 3: Multi-output CNN model Architecture

Taking all of these generated attributes into account, the whole image dataset is divided into 20 data clusters corresponding to each combination of gender, race and class label. Similarly, the tabular data of the behavioral dataset is also divided

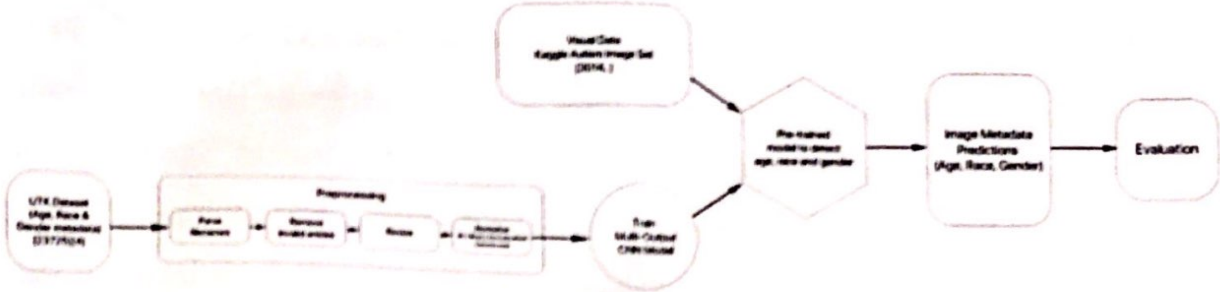

Figure 4: Generation of Visual metadata

into 20 clusters. For each image in the visual dataset, a tabular record from within the data cluster is chosen arbitrarily and concatenated with the visual record to form a record in the integrated dataset. Each of the integrated records consists of an image and a corresponding tabular record of the behavioral data.

3.3 Score Generation

The images and the behavioral data from the integrated records are trained on two different classifiers to generate score values. Decision level fusion of the datasets are done due to the multimodal nature of the data, which results in different preprocessing, feature generation and training to be best suited to each data type.

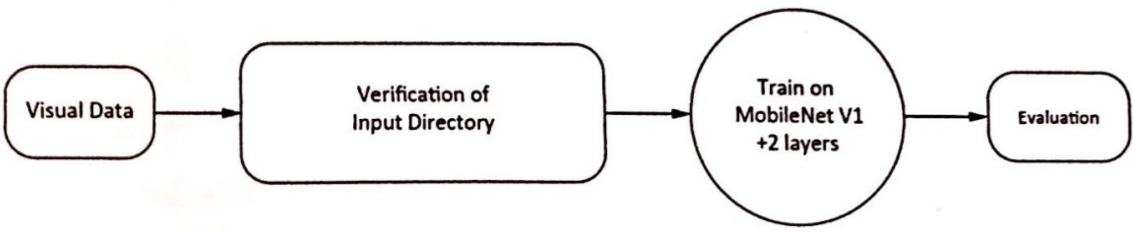

Figure 5: Processing of visual data

Since the visual data is already cropped and cleaned to have uniform sized images, there is minimal preprocessing required on it, as shown in the figure below. The input directories of images are verified and after that the images from the integrated dataset are trained on a MobileNet classifier with two augmented dense layers. MobileNet has been shown to be just as accurate as the larger CNN architectures while significantly reducing the computing. To perform deep learning on the dataset, MobileNet is utilized followed by two dense layers. The first layer (with L2 regularization and ReLu activation is dedicated to distribution, and allows customisation of weights to input into the second dense layer. A dropout of 0.4 is applied to the first layer to prevent overfitting. Then the second dense layer allows for generation of the logit scores. The second layer consists of two nodes, with no activation function applied to it. These scores are the prediction weights of the of the classes given by the classifier. The architecture model of the CNN is given below.

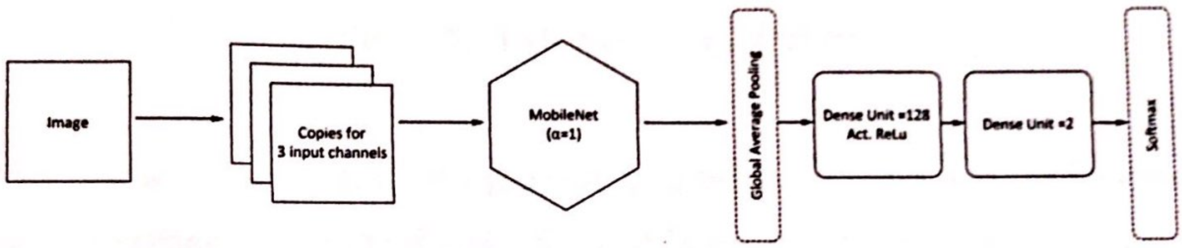

Figure 6: CNN model Architecture

The behavioral data records are taken from the integrated dataset and trained separately. The data is preprocessed before being trained with a neural network to generate logit scores. Feature selection step is used to select relevant records from the dataset. The 10 attributes pertaining to the AQ-10 questionnaire were selected, along with the gender, race, age and whether any family member has ASD. After that, the data is cleaned to remove all the records with incomplete or missing information. Then the categoriacal data is encoded into numerical values to be used in the neural network. The whole numerical dataset is normalised in order to prevent any bias to be coming from the varying range of the attribute values. The behavioral data is then passed through an Artificial Neural Network (ANN) consisting of 3 hidden layers: with 64, 128 and 64 layers. A dropout rate of 0.8 is applied to prevent overfitting. The last layer of the ANN consists of two nodes without any activation function applied to them. These two nodes generate two score values that are used to calculate the final prediction of the combined model. The architecture model of the ANN is given below.

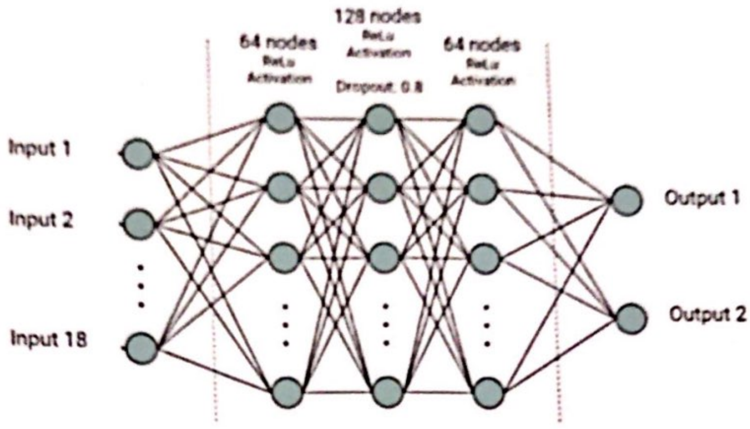

Figure 7: ANN model Architecture

The figures below illustrate the processing on behavioral data and the hyperparameter setting of both ANN and CNN models, respectively. For the ANN model, we chose the best set of hyperparameters (number of nodes = 128, dropout = 0.4, classifier = adam) which gave an accuracy of 100%. For the CNN model, we chose the set of hyperparameters as used in [10] (dropout = 0.4, regularization strength = 0.015) which gave an accuracy of 93.57%.

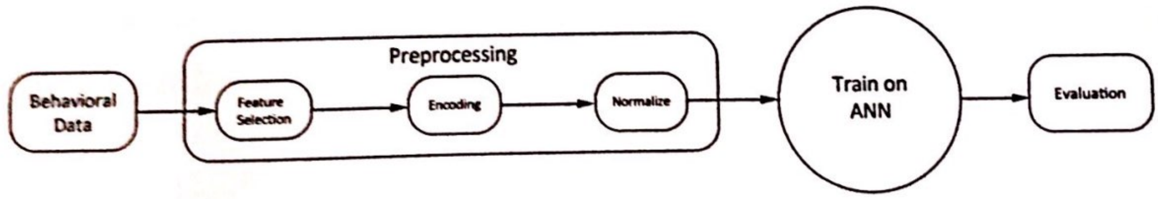

Figure 8: Processing on behavioral data

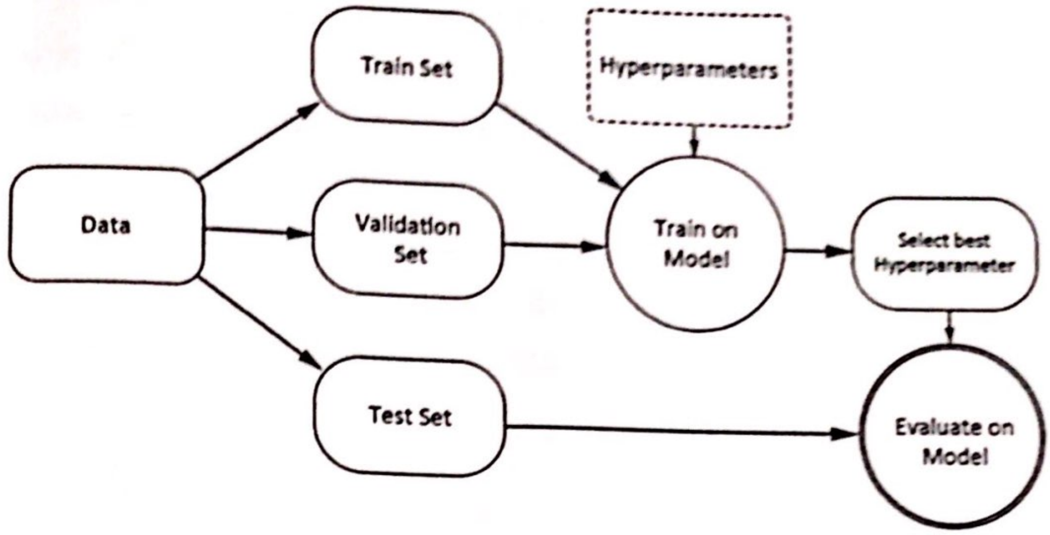

Figure 9: Hyperparameter setting

3.4 Decision Level Fusion

The scores obtained from the ANN and the MobileNet architecture are used to give final prediction of the class labels for the test records. There are two scores obtained from the ANN and two from the MobileNet architecture. This is to ensure that there is no bias of any dataset due to unequal number of scores from any dataset. The range of values can also cause bias of particular attributes, so all the scores are normalised. After that, to ensure that the best decision level integration method is used, three different methods are used to integrate the scores and the final predictions are evaluated.

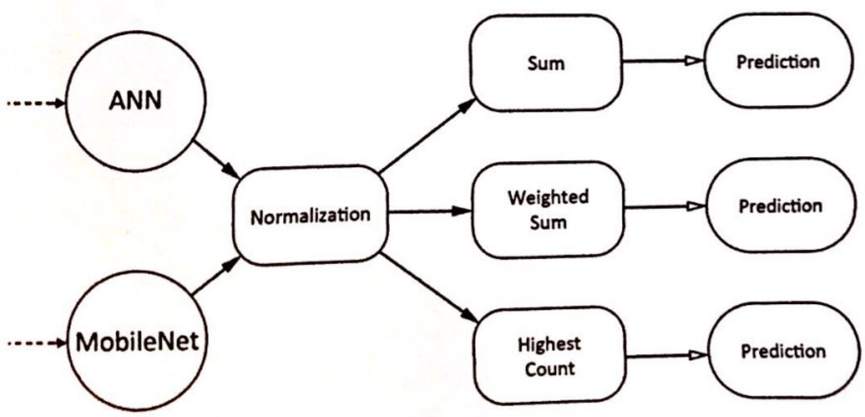

Figure 10: Decision Level Fusion

The three methods are: a) Sum- The average of the scores are taken for each class by adding corresponding scores from each dataset and dividing by 2 (number of datasets). b) Weighted Sum- This is similar to sum, but before addition, a weight is assigned the scores from each dataset. c) Highest Count (Max)- This method compares the scores from each dataset and takes the maximum value. After applying these methods to integrate the scores, each score is passed through a softmax classifier to generate the probability of the integrated record belonging to each class.

4 Experimental Evaluation and Result Analysis

4.1 Overview

The proposed approach to detect ASD from two distinct datasets will be tested on the Integrated records of the dataset and on each dataset individually. Large datasets will require exponentially more time to compute the Cartesian product of the records, so instead the image data rows are randomly concatenated with the behavioral data within each sub class. To test the hypothesis that Integrated Dataset would indeed perform better than initial datasets, three experiments are performed and the performance is evaluated in terms of sensitivity, specificity, accuracy, f-score, precision and recall each time. The experiments are: a) ASD Detection using the Visual Dataset, b) ASD detection using the Behavioral Dataset c) ASD detection on the Integrated Dataset. Three methods of decision level fusion are used, namely: sum, weighted sum and maximum. All three methods of fusion are compared with each other by evaluation based on the sensitivity, specificity, accuracy, f-score, precision and recall. Thereby we can experimentally discern if the Integration method is useful and if it is, which type of integration gives the best performance.

4.2 Experimental Setup

Our experiment utilised three datasets, Dataset 1 and Dataset 2 contained the behavioral and the visual data respectively to be integrated. Dataset 3 contained the gender and ethnicity labeled facial images that have been used to train the classifier that generated the image metadata before integration.

4.2.1 Dataset Preparation

Each dataset underwent the necessary preprocessing steps that made them suitable to be used with the classifier models. The preprocessing required and the overview of each dataset have been outlined below.

4.2.1.1 Dataset 1 Description Our data set 1 "Autistic Spectrum Disorder Screening Data for Toddlers ", which was also used in [21], was recorded via Q-Chat-10: Quantitative Checklist for Autism in Toddlers. It consists of 1,054 instances with each instance having 18 attributes (including the class labels). The attributes are of categorical, continuous and binary types. Q-Chat-10 gives scores based on ten behavioral attributes which it calculates from a questionnaire. For preprocessing, the features relevant to autism detection are selected, then the categorical features were encoded, and finally 0-1 normalization was performed on the values to ensure they are standardized within a fixed range.

4.2.1.2 Dataset 2 Description Our data set 2 "Detect Autism from a facial image", which was also used in [10], was obtained from Kaggle, and consists of 3,014 children's facial images in total. The images are evenly split between two classes: autistic (1507) and non-autistic (1507). Each image is in 224 x 224 x 3 jpg format. The images were obtained from online, both through Facebook groups and through Google Image searches. Once all the images were gathered, they were subsequently cropped so that the faces occupied the majority of the image. Prior to training, the images are split into three categories: train, validation, and test. This dataset has several versions and the ones we used in our experiments are version 1 and version 9 and compared the performances with both. The version 1 contains few duplicate images which could account for the slight increase in accuracy as mentioned in [10]. Since the images of this dataset had already been cropped and resized, the size of the images was compatible the model and no further preprocessing was necessary.

4.2.1.3 Dataset 3 Description Our data set 3 "UTK Dataset" [37], which was also used in [29], is a large-scale face dataset. It consists of over 20,000 images in total. The images cover large variation in pose, facial expression, illumination, occlusion, resolution, etc. The images are evenly split between two genders. Each image has its respective annotations of age, gender and race. Each record is stored in the following format: age_gender_race_datetime.jpg, where age is an integer from

0 to 116, gender is an integer: 0 - male, 1 - female, race is an integer: 0 - white, 1 - black, 2 - asian, 3 - indian, 4 - others, date and time, denoting when the picture was taken. The annotation was obtained from the filenames by parsing them. After that the invalid entires, such as unlabelled or partially labelled rows were removed from the dataset. Finally the images were resized and column wise 0-1 normalisation was performed on them before being used for training the gender and ethnicity classifier model.

4.2.2 Software and Hardware Configuration

The experiments have been performed on an Intel Core i5-6500 3.2 Ghz processor with 4 cores. The chipset used was of the Intel Skylake Series and a RAM of 8 GB @ 3000 Mhz was used. All the experiments have been performed in Python language on the platform of Google Colaboratory.

4.3 Result Analysis

The results obtained verify that the integrated dataset performs better at ASD detection compared to the individual datasets, namely behavioural data, visual data version 1 and visual data version 9.

4.3.1 Experimental Results

This result is evident across all three decision level fusion techniques, with the highest set of classification accuracies being obtained for the integrated dataset that was composed of the visual data version 1 and the behavioral data. The accuracy of the randomly integrated data where the records were combined arbitrarily was significantly less than the accuracy of the integrated dataset. The sensitivity, specificity, precision, recall and F1 scores are calculated for each of the experiments performed.

Table 1: Comparison of Individual and Integrated Datasets

			Accuracy - single fold validation (%)	Accuracy - 10 fold cross validation (%)
Individual Datasets	(1) Visual data - Version 1		93.57	81.30
	(2) Visual data - Version 9		90.33	-
	(3) Behavioral data		99.05	99.40
Integrated Dataset	(1) & (3)	Sum	97.40	96.20
		Weighted sum	96.43	95.96
		Max	97.57	96.44
	(2) & (3)	Sum	96.00	-
		Weighted sum	94.00	-
		Max	96.00	-

4.3.2 Analysis

The improved accuracy results for the experiments done with visual data version 1 are more likely to be caused due to the the presence of duplicates. For further analysis and detection of any possible overfitting, a 10 fold cross validation was performed on the visual data version 1, behavioral data and the integrated dataset formed after combining them. As can be seen from the table above, the accuracy values after 10 fold cross validation does not deviate significantly from our reference model, thus validating our model.

The accuracy of the integrated dataset created from data clusters grouped by gender and race features that are common in both datasets is significantly higher. This demonstrates the effectiveness of combining records based on common attributes. As before, the integrated dataset containing visual data version 1 has higher accuracy than the dataset containing visual data version 9 due to duplicates.

Table 2: Comparison of Integrated Dataset with Randomly Integrated Data

		Accuracy (%)	
		Visual data (Version 9)	Visual data (Version 1)
Randomly Integrated Data	Sum	52.00	84.65
	Weighted Sum	51.00	76.75
	Max	55.55	69.64
Integrated Dataset (Grouped by common features)	Sum	96.00	97.40
	Weighted Sum	94.00	96.43
	Max	96.00	97.57

The results reflected from the evalutaion metrics are aligned with the assumption that integration of datasets improves the accuracy of detection of ASD. This is due to the accuracy of the integrated dataset being higher than the accuracy of visual dataset individually. The integrated dataset contains same samples present in the individual datasets, so discrepancies in the accuracy measures are more likely to be caused due to the integration.

Table 3: Comparison of Evaluation metrics

	Accuracy (%)	Sensitivity	Specificity	Precision	Recall	F1 score
Behavioral data	99.05	1.00	0.99	0.99	1.00	1.00
Visual data (Version 1)	93.57	0.93	0.94	0.94	0.93	0.93
Integrated data (sum)	97.40	0.97	0.98	0.98	0.97	0.97
Integrated data (weighted sum)	96.43	0.96	0.96	0.96	0.96	0.96
Integrated data (max)	97.57	0.97	0.98	0.98	0.97	0.97

5 Conclusion and Future Work

We have achieved an accuracy of 97.57% in autism detection from visual and behavioral data. Our work proves that integration of multi modal data (visual and behavioral) can be better for detecting ASD than any individual visual dataset(93.57%). We have also shown that combining data within sub classes based on common attributes like gender and race, is more accurate than combining them arbitrarily. The datasets used in ASD studies usually consist of small number of subjects and only one mode of data. Through our work we have also developed a novel, integrated dataset which consists of multi modal data (visual and behavioral).

However, we believe the evaluation would be more accurate if ground truth of ASD labels for the data had been obtained via case studies on actual ASD patients. The image data had duplicates present and there was lack of ground truth labels for the predicted image metadata. Both the datasets used for integration had skewed distributions, which may have reduced the accuracy. The dataset used for behavioral data was not robust, and using alternative datasets should give greater significant improvement to results. Finally, our prediction algorithms can only predict the race, not the ethnicity of a subject. Race and ethnicity have variations among them, race being related to the physical characteristics of the person and ethnicity being the culture that their ancestors identify with. Thus considering ethnicity and race to be the same may have introduced some inaccuracies.

For future improvements to the experiments, we can incorporate handcrafted visual features like Masculinity, Eye contact, Patterns of expression that have been shown in previous studies to be useful for ASD detection. Case studies can be performed to ascertain ground truth of the labels. So far, three simple types of Decision Level Fusion techniques have been attempted on the dataset. If the performance of the dataset is not adequate, a neural network could be used that takes the score values as input and computes the class prediction.

References

[1] `https://www.cdc.gov/ncbddd/developmentaldisabilities/facts.html`, last accessed on 07/12/20.

[2] M. N. Parikh, H. Li, and L. He, "Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data," *Frontiers in computational neuroscience*, vol. 13, p. 9, 2019.

[3] A. Zunino, P. Morerio, A. Cavallo, C. Ansuini, J. Podda, F. Battaglia, E. Veneselli, C. Becchio, and V. Murino, "Video gesture analysis for autism spectrum disorder detection," in *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 3421–3426, IEEE, 2018.

[4] H. Duan, G. Zhai, X. Min, Z. Che, Y. Fang, X. Yang, J. Gutiérrez, and P. L. Callet, "A dataset of eye movements for the children with autism spectrum disorder," in *Proceedings of the 10th ACM Multimedia Systems Conference*, pp. 255–260, 2019.

[5] `https://www.kaggle.com/fabdelja/autism-screening-for-toddlers/version/1?select=Toddler+Autism+dataset+July+2018.cs`, last accessed on 07/12/20.

[6] `https://www.kaggle.com/gpiosenka/autistic-children-data-set-traintestvalidate`, last accessed on 07/12/20.

[7] I. Subramanian, S. Verma, S. Kumar, A. Jere, and K. Anamika, "Multi-omics data integration, interpretation, and its application," *Bioinformatics and biology insights*, vol. 14, p. 1177932219899051, 2020.

[8] P. Kirk, J. E. Griffin, R. S. Savage, Z. Ghahramani, and D. L. Wild, "Bayesian correlated clustering to integrate multiple datasets," *Bioinformatics*, vol. 28, no. 24, pp. 3290–3297, 2012.

[9] R. Shen, A. B. Olshen, and M. Ladanyi, "Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis," *Bioinformatics*, vol. 25, no. 22, pp. 2906–2912, 2009.

[10] M. Beary, A. Hadsell, R. Messersmith, and M.-P. Hosseini, "Diagnosis of autism in children using facial analysis and deep learning," *arXiv preprint arXiv:2008.02890*, 2020.

[11] K. Niu, J. Guo, Y. Pan, X. Gao, X. Peng, N. Li, and H. Li, "Multichannel deep attention neural networks for the classification of autism spectrum disorder using neuroimaging and personal characteristic data," *Complexity*, vol. 2020, 2020.

[12] T. Charman, "Developmental approaches to understanding and treating autism," *Folia Phoniatrica et Logopaedica*, vol. 62, no. 4, pp. 166–177, 2010.

[13] H. Shi, J. Chen, T. Li, and G. Wang, "A new method for estimation of spatially distributed rainfall through merging satellite observations, rain-gauge records, and terrain digital elevation model data," *Journal of Hydro-environment Research*, vol. 28, pp. 1–14, 2020.

[14] D. W. Tan, S. Z. Gilani, M. T. Maybery, A. Mian, A. Hunt, M. Walters, and A. J. Whitehouse, "Hypermasculinised facial morphology in boys and girls with autism spectrum disorder and its association with symptomatology," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.

[15] N. Müller, "Eye tracking as a diagnostic tool in autism spectrum disorder," 2017.

[16] C. Wu, S. Liaqat, S.-c. Cheung, C.-N. Chuah, and S. Ozonoff, "Predicting autism diagnosis using image with fixations and synthetic saccade patterns," in *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 647–650, IEEE, 2019.

[17] R. S. Hessels, G. A. Holleman, T. H. Cornelissen, I. T. Hooge, and C. Kemner, "Eye contact takes two-autistic and social anxiety traits predict gaze behavior in dyadic interaction," *Journal of Experimental Psychopathology*, vol. 9, no. 2, pp. jep–062917, 2018.

[18] W. A. Helland and T. Helland, "Emotional and behavioural needs in children with specific language impairment and in children with autism spectrum disorder: The importance of pragmatic language impairment," *Research in Developmental Disabilities*, vol. 70, pp. 33–39, 2017.

[19] D. S. Andrews, A. Marquand, C. Ecker, and G. McAlonan, "Using pattern classification to identify brain imaging markers in autism spectrum disorder," in *Biomarkers in Psychiatry*, pp. 413–436, Springer, 2018.

[20] `http://fcon_1000.projects.nitrc.org/indi/abide/`, last accessed on 07/12/20.

[21] T. Akter, M. S. Satu, M. I. Khan, M. H. Ali, S. Uddin, P. Lio, J. M. 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.

[22] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for autism detection," *Health informatics journal*, vol. 26, no. 1, pp. 264–286, 2020.

[23] S.-G. JUNG, J. An, H. Kwak, J. Salminen, and B. J. Jansen, "Assessing the accuracy of four popular face recognition tools for inferring gender, age, and race," 2018.

[24] A. Vani, R. Raajan, D. H. Winmalar, and R. Sudharsan, "Using the keras model for accurate and rapid gender identification through detection of facial features," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 572–574, IEEE, 2020.

[25] O. Agbo-Ajala and S. Viriri, "Deeply learned classifiers for age and gender predictions of unfiltered faces," *The Scientific World Journal*, vol. 2020, 2020.

[26] D. Riccio, G. Tortora, M. De Marsico, and H. Wechsler, "Ega—ethnicity, gender and age, a pre-annotated face database," in *2012 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings*, pp. 1–8, IEEE, 2012.

[27] `https://fei.edu.br/~cet/facedatabase.html`, last accessed on 02/03/20.

[28] K. Kärkkäinen and J. Joo, "Fairface: Face attribute dataset for balanced race, gender, and age," *arXiv preprint arXiv:1908.04913*, 2019.

[29] P. Chandaliya, V. Kumar, M. Harjani, and N. Nain, *SCDAE: Ethnicity and Gender Alteration on CLF and UTKFace Dataset*, pp. 294–306. 03 2020.

[30] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern recognition letters*, vol. 33, no. 4, pp. 431–437, 2012.

[31] Y. Li and A. Ngom, "Data integration in machine learning," in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1665–1671, IEEE, 2015.

[32] B. Gokberk and L. Akarun, "Comparative analysis of decision-level fusion algorithms for 3d face recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3, pp. 1018–1021, IEEE, 2006.

[33] C. Cortes, M. Mohri, and A. Rostamizadeh, "Algorithms for learning kernels based on centered alignment," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 795–828, 2012.

[34] D. A. Miller, K. Pacifici, J. S. Sanderlin, and B. J. Reich, "The recent past and promising future for data integration methods to estimate species' distributions," *Methods in Ecology and Evolution*, vol. 10, no. 1, pp. 22–37, 2019.

[35] Y. Gong, Y. Niu, W. Zhang, and X. Li, "A network embedding-based multiple information integration method for the mirna-disease association prediction," *BMC bioinformatics*, vol. 20, no. 1, p. 468, 2019.

[36] J.-k. Kim and S.-M. Tam, "Data integration by combining big data and survey sample data for finite population inference," *arXiv preprint arXiv:2003.12156*, 2020.

[37] `http://aicip.eecs.utk.edu/wiki/UTKFace`, last accessed on 02/03/20.