



# A Data-Driven Approach For Detecting Autism Spectrum Disorders

Manika Kapoor and David C. Anastasiu



## Introduction

- **Objective:** Predict Autism Spectrum Disorders (ASD) and characterize the type of stimuli needed for its detection.
- **Why?:**
  - No cure exists, but early diagnosis increases the chances of patients to function properly in society.
  - Current methods are either subjective or based on responses to single stimulus
- **How?:**
  - Creating machine learning based models using Electrocardiogram (ECG) and Skin Conductance (SC) data.



- **What is ASD?**
  - Neurodevelopmental disorder in which patients display diminished capacity of social interaction.
  - Instruments used for ASD assessment are lengthy to administer and are also not accurate.

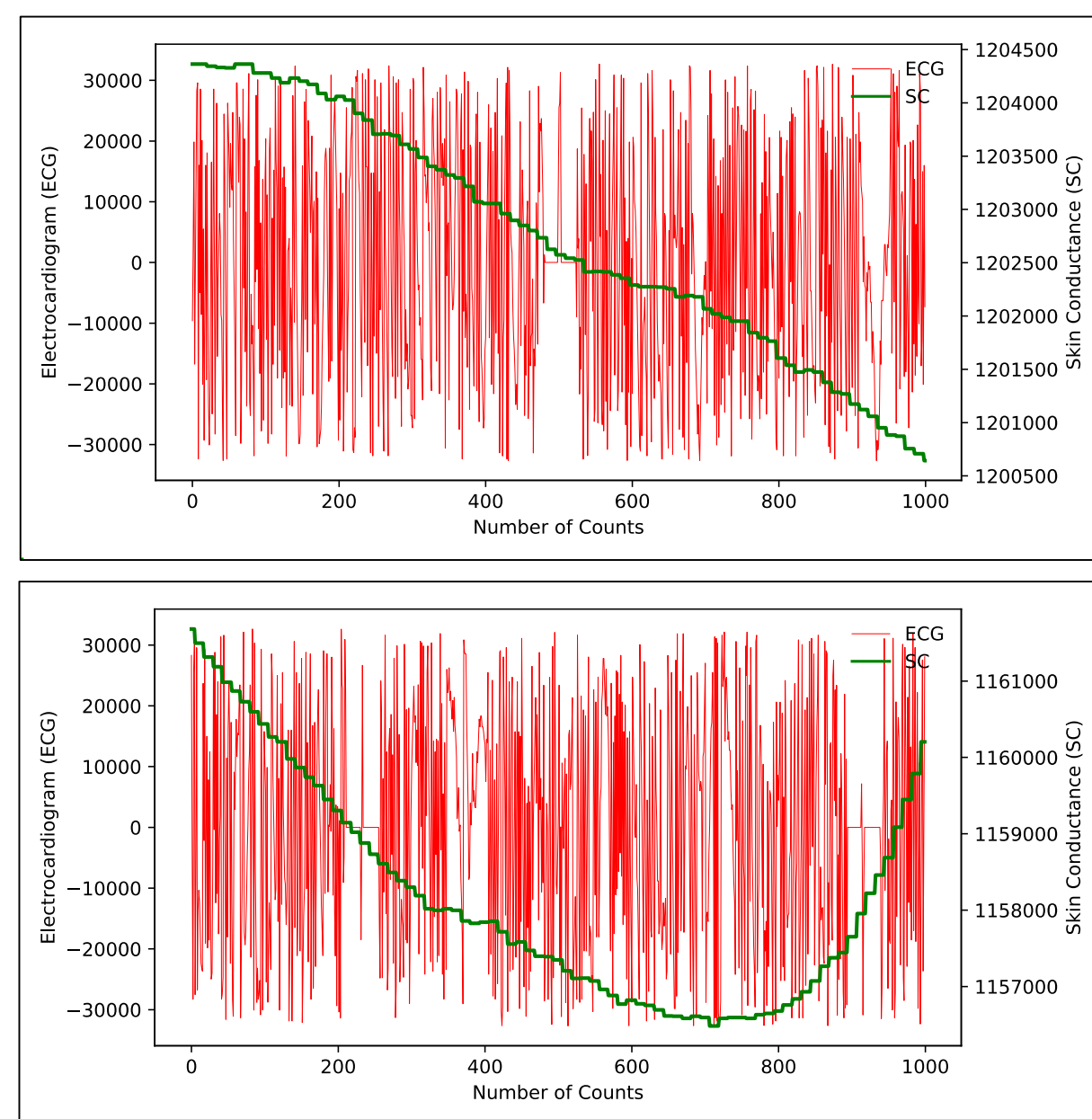


Fig1: Sample Dataset

- **Dataset:**
  - Collected during Sensory Challenge Protocol (SCP) [1].
  - Reactions to multiple stimuli were observed from 25 children with ASD and 25 Typically Developing (TD) children (5-12 years age).
  - Time taken for each protocol: 45–90 minutes.
  - Included three phases: baseline, sensory challenge, and recovery.
  - **Baseline and Recovery periods:** 3 minutes with no stimulation.
  - **Sensory Challenge:** 6 stimuli, each administered for 3 seconds and was presented at least 8 times.
  - **Six Stimuli:**
    - Auditory tones (at 84 dB)
    - Visual cues (20W strobe light at 10Hz)
    - Auditory siren sound (at 78 dB)
    - Olfactory (wintergreen oil passed under the nose)
    - Tactile (touch along the jaw bone a feather)
    - Vestibular (chair tilted back to a 30 degree angle)

## Hypothesis and Supporting Evidence

- **Hypothesis:**
  - We hypothesize that autistic children are greatly affected by certain sensory stimulation and thus may take longer to return to normal state.
  - In contrast, TD children can quickly recover to a normal state after the sensory trial.

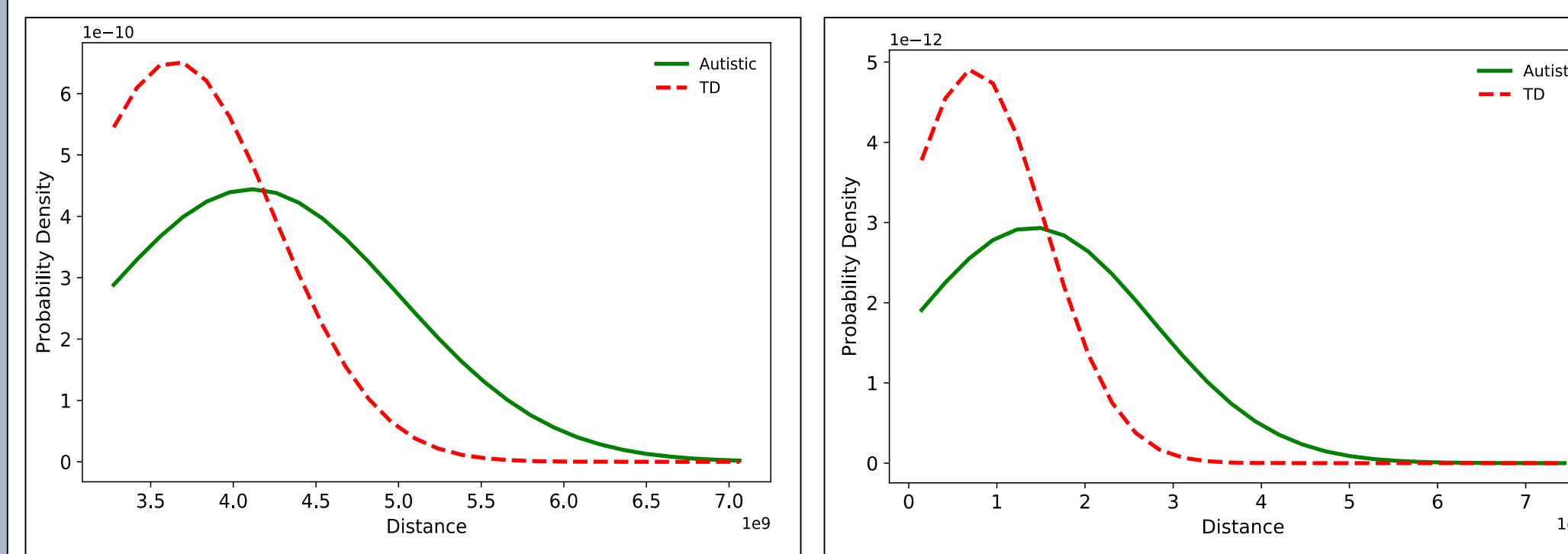


Fig2: Normal Distribution of Baseline and Recovery Distance using EKG (left) and SC (right) data

- **Evidence:**
  - Compare the sensory data recorded during **baseline** stage and during the **recovery** stage.
  - No stimulus was administered during either rest stages.
  - Compute the Euclidean DTW distance of the ECG and SC time series recorded during the rest periods.
  - Euclidean DTW is the distance between two time-dependent sequences which may have different speeds and length.
  - Mean distance for the autistic children is **approximately 60% more** than that for the TD children.

## Methods

- **Feature Extraction:**
  - Transform data for each stimulus into a form that is representative of the data but enables efficient analysis.
  - Three methods:
    - **Equal Width Partition (EWP):** SCP has stimuli administered in specific number of contiguous trials. Divide each stimulus data into 'n' equal parts and then take
      - mean and standard deviation of each split.
      - slope and intercepts of peaks and slope and intercepts of valleys in ECG and slope and intercept in SC data.

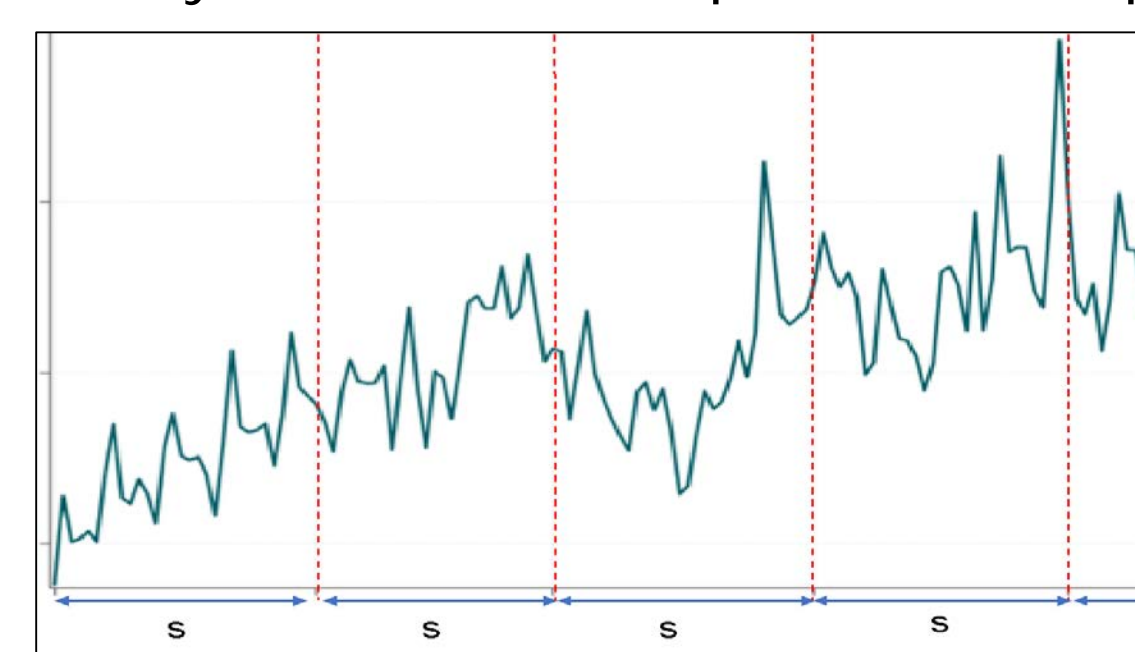


Fig3: Equal Width Partition

- **Dynamic Time Warping (DTW):**
  - Identify similar patterns in two time series even if one of them is stretched out.
  - Calculate the DTW Euclidean distance between ECG and SC data of every subject with every other subject.
  - Working with huge time-series is computationally very expensive, so we divide the data for each stimulus into 8 equal parts with 10% data points from the neighboring splits.

## Methods

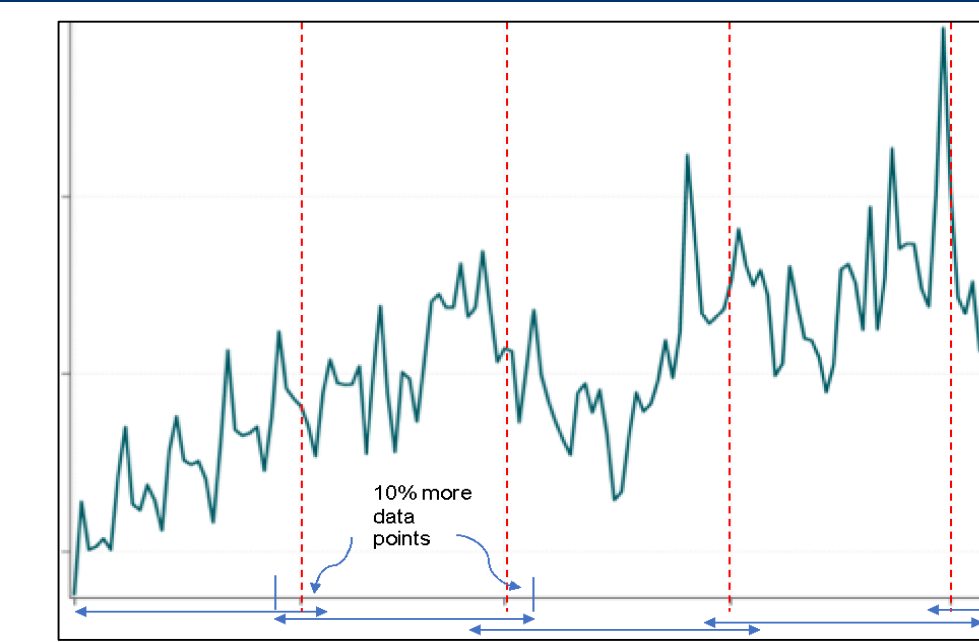


Fig4: Initial Splitting for DTW Experiments

- **Motif Based Segmentation (MBS):**
  - Represent each stimulus using motifs.
  - Calculate Euclidean distance between ECG and SC data of every subject with every other subject.

## Developing Predictive Models for Autism Detection:

- Binary classification problem.
- Built eight different types of models, namely:
  - K-Nearest Neighbors (KNN)
  - Support Vector Machine (SVM)
  - Random Forest (RF)
  - DTW-based KNN model (DTW-KNN)
  - Motif-based KNN model (M-KNN)
  - Decision Tree (DT)
  - Naive Bayes (NB)
  - XGBoost (XGB)

## Degree with which each stimulus affects Autistic children:

- Contribution of each stimulus towards predicting autism in children.
- Use Stochastic Gradient Descent (SGD) to find appropriate weights for each stimulus for predicting autism in children.

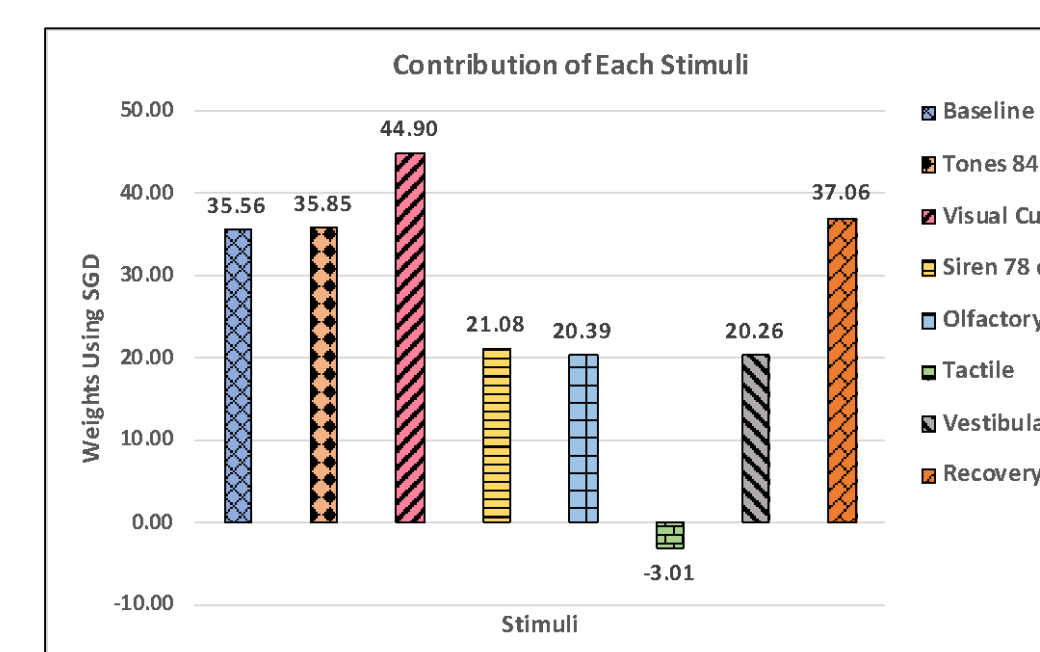


Fig5: SGD Result

## Results and Conclusions

- Using EWP, the best model accuracy achieved is **93.33%** by XGBoost model created using only SC data and using slope and intercept of represent each partition.

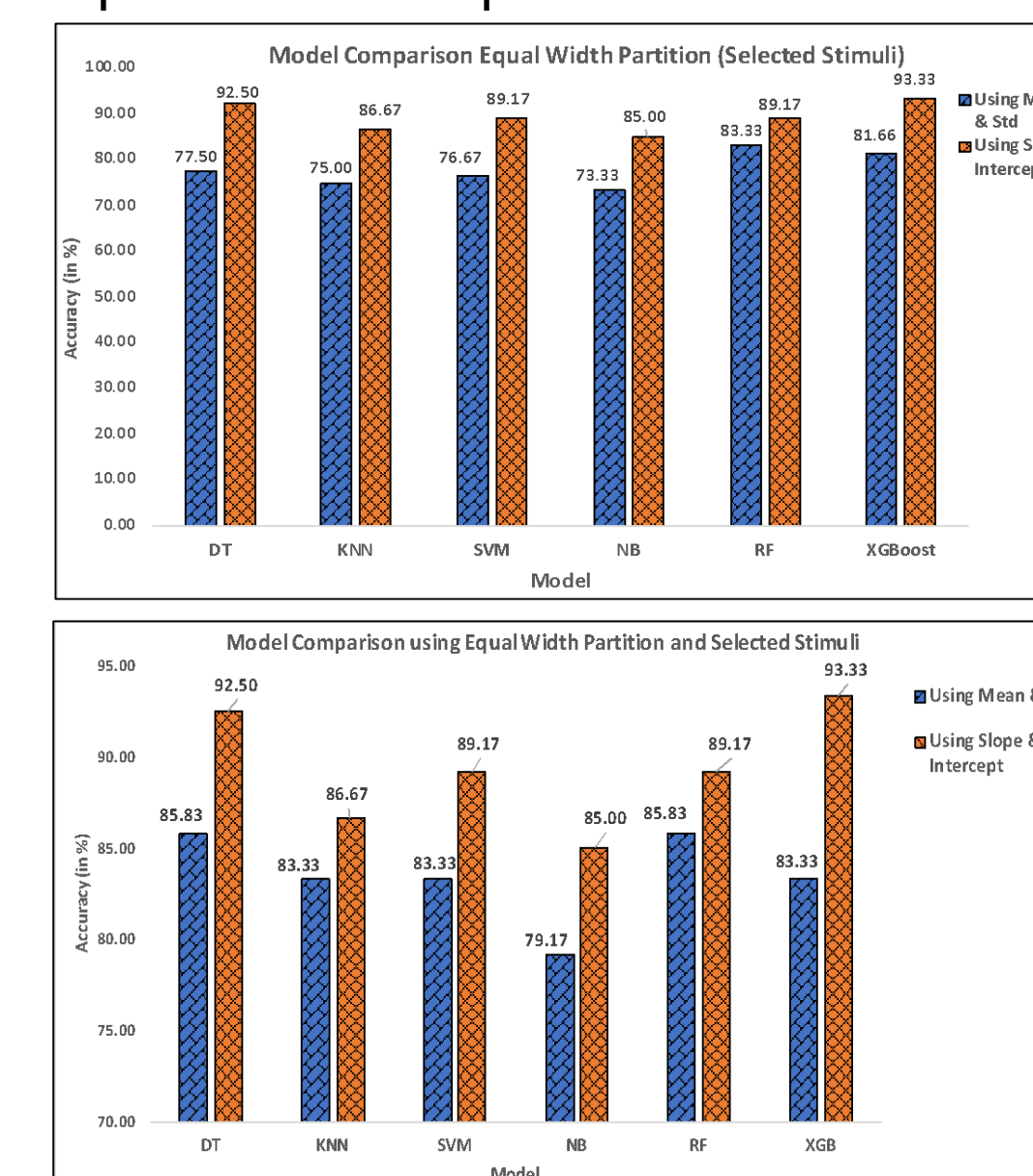


Fig6: Model Accuracy Comparison using Equal Width Partition

- Using **MBS**, the best accuracy achieved is **95.83%** by KNN model built using both SC and ECG data.
- Using **DTW**, the best accuracy achieved is **77.50%** by the KNN model built using both SC and ECG data.

## Results and Conclusions

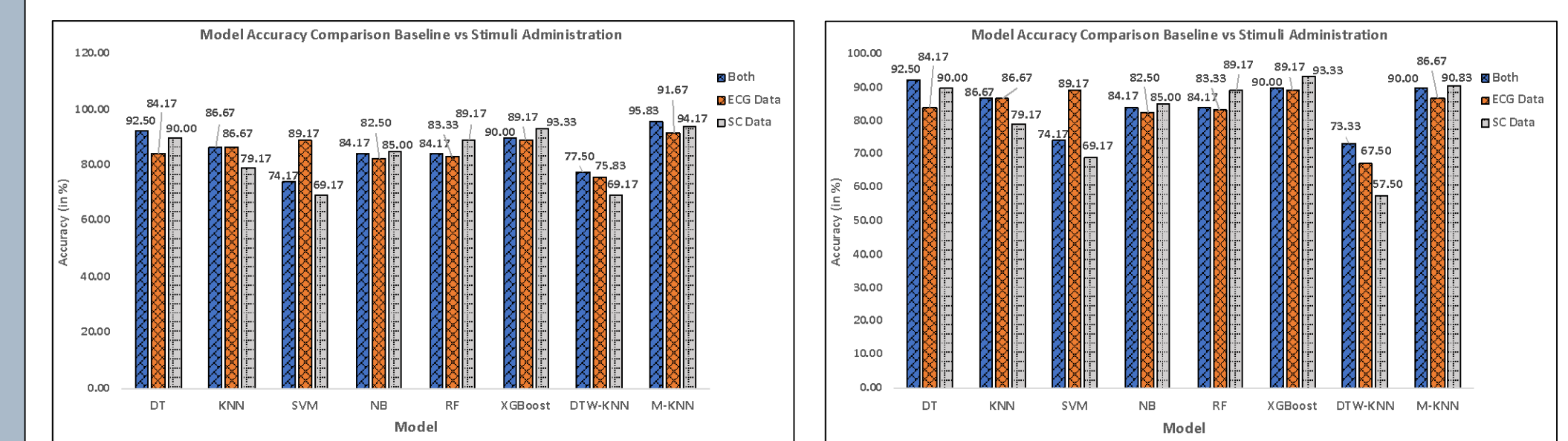


Fig7: Overall Model Accuracy Comparison

- Stimuli application helps in predicting ASD.
- Without stimuli application, best accuracy achieved is **82.50%** which increases to **95.83%** after stimuli application.

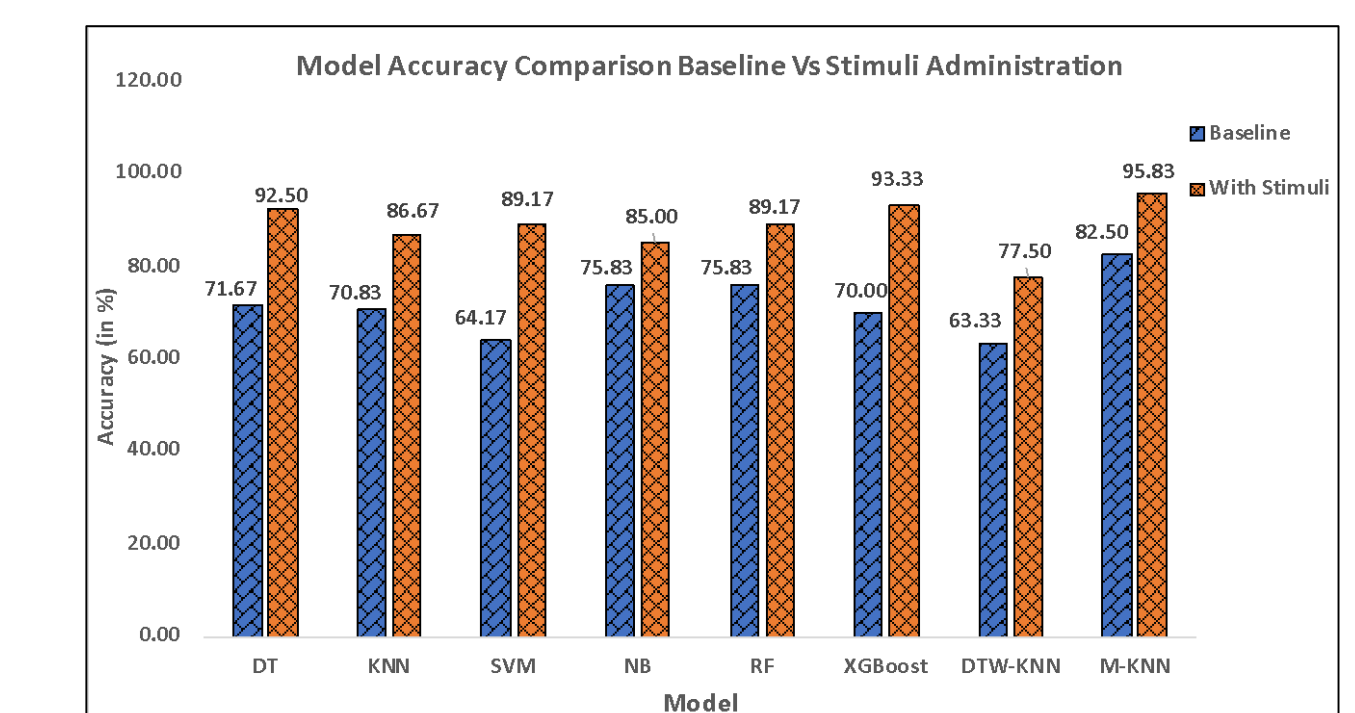


Fig8: Baseline vs Stimuli Application Comparison

- Stimuli contributing towards best performing model are: **Baseline, Tones 84dB, Visual Cues, Siren 78dB, Olfactory & Vestibular.**

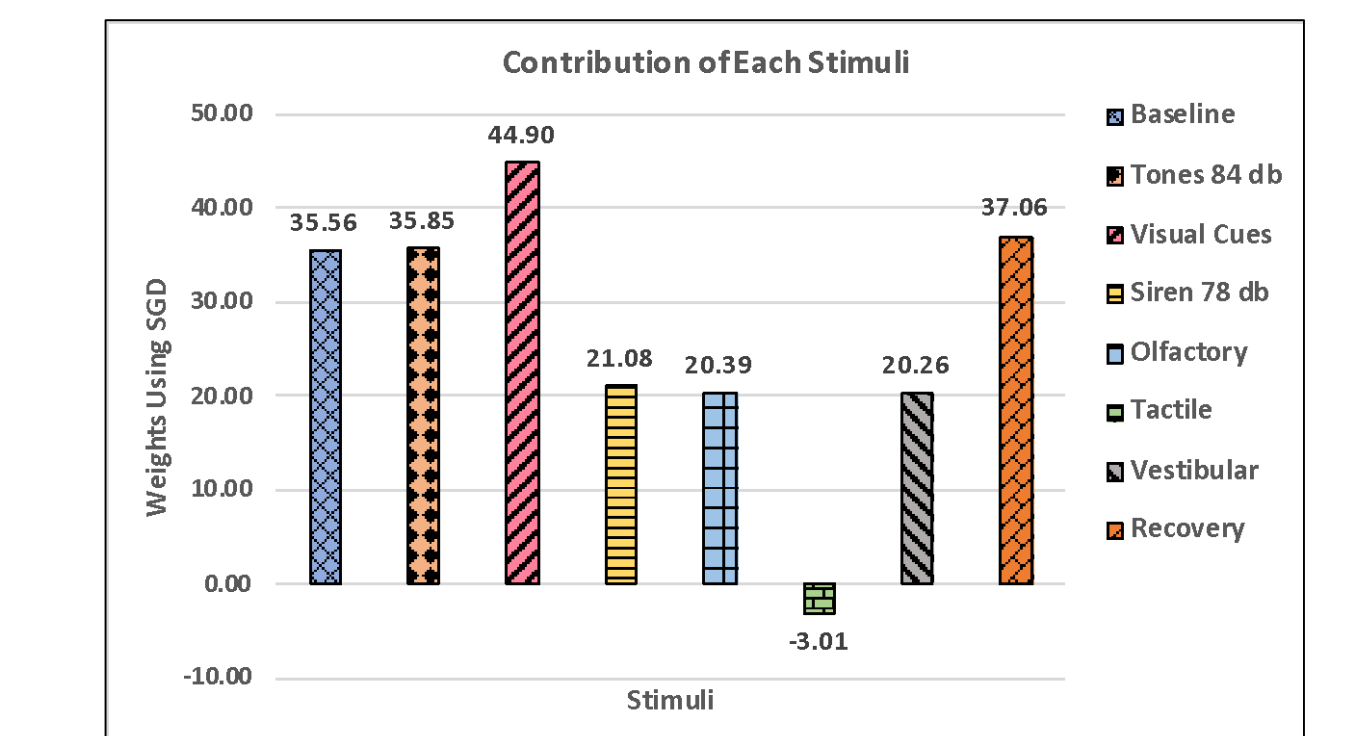


Fig9: Contribution of Each Stimuli Towards Best Performing Model

- DTW being one of the best approaches to compare time-series data in general, doesn't perform satisfactorily with huge time-series data and is computationally very expensive, taking **3x** the amount of time to train and **450x** time to predict as compared to the best model.

- The best performing model i.e. M-KNN takes on an average **18.8 hours** to train and approximately **30 seconds** to make predictions.

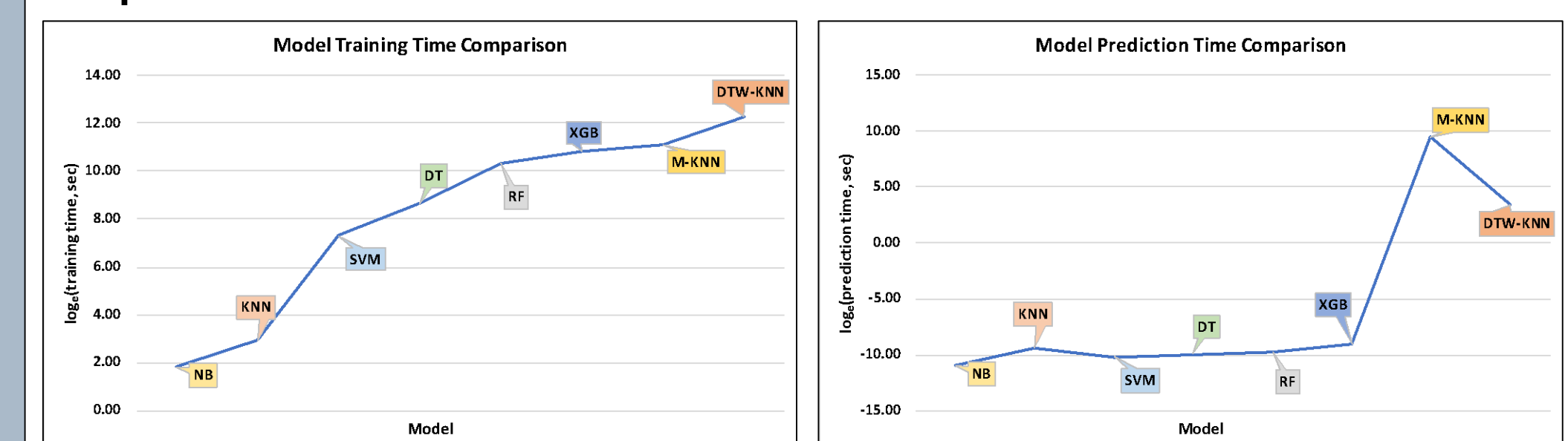


Fig10: Model Time Efficiency Comparison

## Key References

- [1] M. C. Chang, L. D. Parham, E. I. Blanche, A. Schell, C.-P. Chou, M. Dawson, and F. Clark, "Autonomic and behavioral responses of children with autism to auditory stimuli," *American Journal of Occupational Therapy*, vol. 66, no. 5, pp. 567–576, 2012, doi:10.5014/ajot.2012.004242.
- [2] Anastasiu, David C., Chang Megan C., & Kapoor Manika (2018). A Data-Driven Approach for Detecting Autism Spectrum Disorders. In *Big Data (Big Data), 2018 IEEE International Conference*. IEEE.(under review)