A Decision Tree Classification Model for detection Of Tourette Syndrome Using Cognitive and Olfactory Testing Data

# Harshi Jain, Kiran Pandey, Bhupendra Verma

*Technocrats Institute of Technology, Bhopal*

Email id[:*harshijain554@gmail.com*,](mailto:harshijain554@gmail.com)

*Technocrats Institute of Technology, Bhopal*

Email [id:*kiran.mishra.cse07@iitbhu.ac.in*,](mailto:kiran.mishra.cse07@iitbhu.ac.in)

*Technocrats Institute of Technology, Bhopal*

Email id[:*bkverma3@gmail.com*](mailto:bkverma3@gmail.com)

*Abstract*- *Tourette's Syndrome (TS for short) is normal neurodevelopment disorder that becomes active during childhood or adolescence. This is described by multiple motor tics (motor tics) and at least one vocal (vocal) tic. Examples of tics include blinking, coughing, clearing the throat, sniffling, and facial movements. Two aspect of Tic disorder are Motor and vocal which depends on level of the complication. The demographic, cognitive and olfactory testing data can be used to develop a model to predict the Tourette’s syndrome in the population and measures can be taken to treat the identified patients. This research paper attempts to build up a machine learning based model to predict TS based on demographic and cognitive testing and Olfactory testing data sets. Results show that Decision Tree model is able to predict the syndrome with 100% accuracy, 1 Mathew’s correlation coefficient, 100% F1-score, 100% precision and 100% recall on the limited available dataset.*

***Keywords:***

***Tourette Syndrome, Machine learning, demographic data, cognitive testing, olfactory testing, accuracy..***

# **Introduction**

Tourette syndrome (TS), a neurodevelopment condition with childhood onset, is characterized by varying motor and vocal tics. Tics typically start before the age of 10, go through a steady decline, and get better over time [1].Simple forms of tics include blinking, mouth twitching, and head twitching. The prevalence of TS is significantly higher than previously believed, according to several studies that have been published since 2000 [2, 3].The number of children with tic disorders is gradually rising as our understanding of the disorder deepens, although most instances do not receive prompt clinical treatment in the early phases of their ailment Additionally, 20% of TS sufferers are ignorant of their tic problem [4]. Tourette syndrome, a chronic tic disorder is characterized by motor and vocal tics, comorbidities such as obsessive-compulsive disorder and attention deficit hyperactivity disorder have been documented[5]. Clinical diagnosis of tic disorders typically involves several complex processes that require time and effective collaboration between physician  and patient to observe and assess patient behavior. Over time, patient diagnosis and treatment have improved due to the development of technology in healthcare. The healthcare industry is undoubtedly the most significant of all the sectors that have profited from technological adoption. As a result, it has helped people live quality life and saved many lives [3].Therefore, we look for an automated technique for detection of Tourette syndrome based on demographic and cognitive testing data to assist in diagnosis. We have used four classifiers i.e. Decision tree, K nearest neighbors, SGD (stochastic gradient descent) and Mathews correlation coefficient 1, F1 score 100% ,precision 100% and recall 100%. Section 1 contains introduction. Section 2 contains related work. Section 3 contains materials and method. Section 4 contains a diagram representing flow of our proposed work. Section 5 contains details of machine learning algorithms. .Section 6 contains Experimental Result and Section 7 contains Conclusion.

# **RELATED WORK**

In this Section we have mentioned the work done by other researchers for Tourette syndrome detection using machine learning and deep learning algorithms. To identify tick actions Geng et al. (Geng et al., 2022) [6] designed the slow-fast andlight-efficient channel attention network (SFLCA-Net). An optically efficient channel attention module (LCA) and two fast and slow branching sub-networks were adopted by the entire network to address the issue of the lack of complementarities of spatiotemporal channel information. The experimental findings show that the SFLCA-Net method is effective; the same was verified on the TD dataset.

Wu et al.[7] proposeda deep learning architecture that learns features from tic motion detection video and blends supervised and unsupervised learning techniques. This model is trained using leave-onesubject-out cross-validation for both binary and multiclass classification tasks. They have achieved accuracy of 94.87%. Barua et al. [8]described the use of a convolution neural network (CNN) model to categorize image data based on wireless channel information (WCI) to diagnose tic disorders. They created a dataset to train the CNN model using WCI data of symptoms from the simple and complex group of motor features. Experimental results showed that the CNN provided satisfactory classification results, with an accuracy of over 97%. [9] integrated multi-modal image features using multiple kernel learning (MKL).Classifying 48 healthy children who were age and gender matched and 44 TS youngsters served to validate the performance of their framework. Utilizing nested cross-validation and combining the features with MKL, the classification accuracy was 94.24%. The cortico-basal ganglia and frontal cortico-cortical circuits, which are assumed to be closely associated to TS disease, are where the majority of the discriminative brain regions were found. Greene et al [10]tested if patterns in brain network activity, measured with resting state functional connectivity (RSFC) MRI, might predict diagnostic group membership for individuals using a multivariate approach called support vector machine (SVM) classification. They reported a novel adaptation of SVM binary classification that, in addition to the overall accuracy for the SVM, provides a confidence measure to the accurate classification of each individual.

*A. Figures and table*

**Table 1.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S. No.** | **Author** | **Modality of Data set** | **Models** | **Accuracy** |
| 1 | Geng et al., 2022 [6] | Tic disorder video dataset | Slow-fast and  light-efficient channel attention network (SFLCA-Net) | >96% |
| 2 | Wu et al ,2021 [7] | Video data | LSTM network | 94.87%, |
| 3 | Barua et al., 2021 [8] | wireless channel information (WCI) based image data | Convolutional neural network | 97% |
| 4 | Wen et al, 2017 [9] | Multi modal image features | Multiple kernel learning | 94.2% |
| 5 | Greene et al., 2016 [10] | MRI | Support vector machine | ~70% |

From Table 1 it is evident that various researches have applied machine learning algorithms for detection of Tourette syndrome. Some [6,7] have used video dataset for detection, Some [8,9] have used image dataset and some [10] have used MRI dataset for training of machine learning algorithm for Tourette syndrome detection. We have not found any study in literature that has used demographic, cognitive testing and olfactory testing data for training machine learning algorithms for detection of Tourette syndrome. So this study proposed a machine learning model to detect Tourette syndrome based on demographic, cognitive and olfactory testing data. We have obtained better accuracy in comparison to all other studies mentioned in Table 1.

# **MATERIALS AND METHODS**

Although, Few studies have been published on the automatic diagnosis of motor abnormalities related to TS, despite the fact that numerous studies have recently concentrated on the pathological, genetic, and therapeutic therapy of TS. For the experimentation and testing of the algorithm we have obtained data for our work from Kronenbuerger et al, 2018 [5]. The twenty-eight adult subjects with TS were included and 28 healthy controls were studied. Healthy controls were matched for age, sex, education, and smoking habits to the TS participants. Healthy controls had no tics, ADHD, OCD, depression, anxiety, or any psychopathology based on the clinical interviews. Statistical analysis of various features is shown in Table 2 and Table 3. Table 2 shows statistical measures for Tourette patient’s and Table 3 shows statistical measures for Healthy control.

**Table 2.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S. No** | **Features** | **Tourette Patient** | | | |
|  |  | **Mean** | **Standard Deviation** | **Skewness** | **Kurtosis** |
| **Demographics** | | | | | |
| **1** | **Age** | 33.07 | 9.24 | 0.60 | -0.39 |
| **2** | **Education(years)** | 11.64 | 1.47 | -0.48 | -1.07 |
| **3** | **Smoking burden in smokers** | 9.57 | 9.36 | 0.09 | -1.90 |
| **Cognitive testing Data** | | | | | |
| **4** | **TMT-A** | 24.32 | 7.02 | 0.29 | -1.03 |
| **5** | **TMT-B** | 47.61 | 18.40 | 1.28 | 2.29 |
| **6** | **Digital Span Forward** | 9.18 | 1.91 | -0.27 | -0.57 |
| **7** | **Digital Span backward** | 6.61 | 2.21 | 0.11 | -0.22 |
| **Olfactory function testing Data** | | | | | |
| **8** | **TDI** | 31.88 | 5.01 | -0.37 | -0.82 |
| **9** | **Odor threshold** | 7.30 | 2.67 | -0.17 | 0.27 |
| **10** | **Odor discrimination** | 12.14 | 2.12 | -0.53 | 1.26 |
| **11** | **Odor identification** | 12.43 | 1.97 | -1.03 | 0.94 |

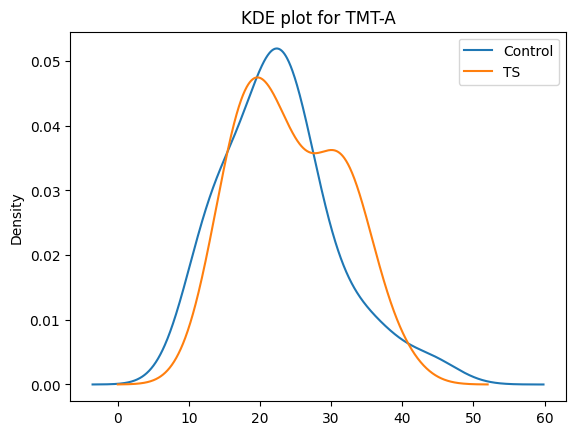
Statistical measures of Demographic and cognitive testing and Olfactory testing data of Tourette patients.

**Table 3.**

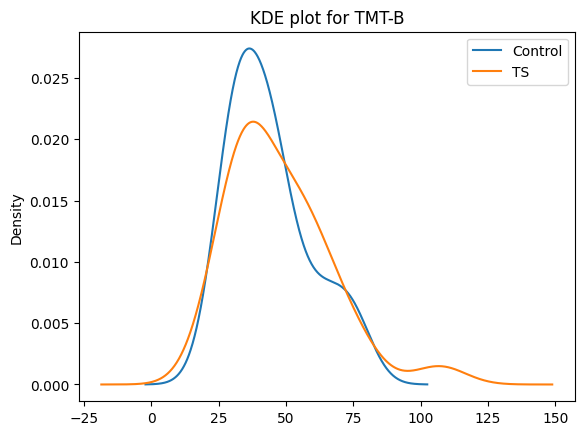
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S. No** | **Features** | **Healthy Control** | | | | |
|  |  | **Mean** | **Standard**  **Deviation** | | **Skewness** | **Kurtosis** |
| **Demographics** | | | | | | |
| **1** | **Age** | 31.93 | 8.61 | 0.93 | | -0.18 |
| **2** | **Education(years)** | 11.25 | 1.72 | 0.12 | | -1.66 |
| **3** | **Smoking burden in smokers** | 7.75 | 8.26 | 0.29 | | -1.82 |
| **Cognitive testing Data** | | | | | | |
| **4** | **TMT-A** | 22.69 | 7.60 | 0.84 | | 0.93 |
| **5** | **TMT-B** | 43.94 | 14.51 | 0.88 | | -0.15 |
| **6** | **Digital Span Forward** | 8.32 | 1.85 | 0.21 | | -0.67 |
| **7** | **Digital Span backward** | 6.82 | 1.73 | -0.32 | | 0.59 |
| **Olfactory function testing Data** | | | | | | |
| **8** | **TDI** | 35.04 | 3.07 | | -0.28 | -0.31 |
| **9** | **Odor threshold** | 8.12 | 2.15 | | 0.19 | -0.32 |
| **10** | **Odor discrimination** | 13.21 | 1.61 | | -0.70 | 0.62 |
| **11** | **Odor identification** | 13.71 | 1.39 | | -0.47 | 0.18 |

Statistical measures of Demographic and cognitive testing and Olfactory testing data of Healthy Control.

Column 2 of Table 2 represents various features. Row 1, 2 and 3 represents demographic data of TS patients and healthy control i.e. age, education and smoking burden in smokers respectively. Rows 4-7 represents cognitive testing data. The Trail Making Test (TMT) is a popular cognitive test for assessing executive and attention function in older individuals. (Hroyuki et al., 2022) [11]. It has two parts TMT –A and TMT B. Row 4 contains data about TMT-A and Row 5 contains data for TMT-B. Digital span tasks are commonly used to assess short-term verbal and Visio spatial memory. There are two versions of this test i.e. digital span forward and digital span backward tests. Rows 6 and Row 7 contain data related to these tests .Rows 8-11 contain data related to olfactory function testing. Row 8 contains TDI (Threshold, Discrimination and Identification) Score: the composite sum score of odor testing, Row 9, row 10 and row 11 contain odor threshold, odor discrimination and odor identification testing score respectively.

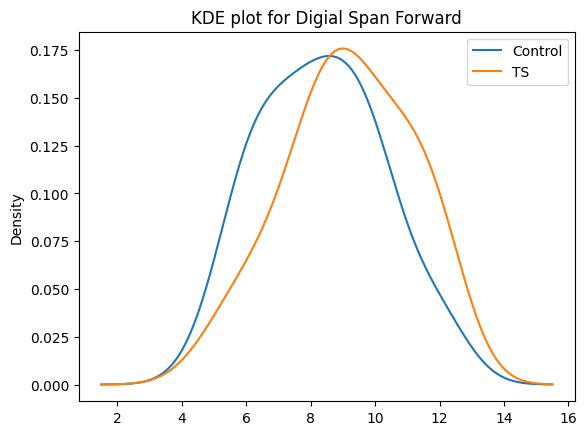


**Fig.1:** Analysis of TMT-A for Healthy control and Tourette syndrome group of patients.

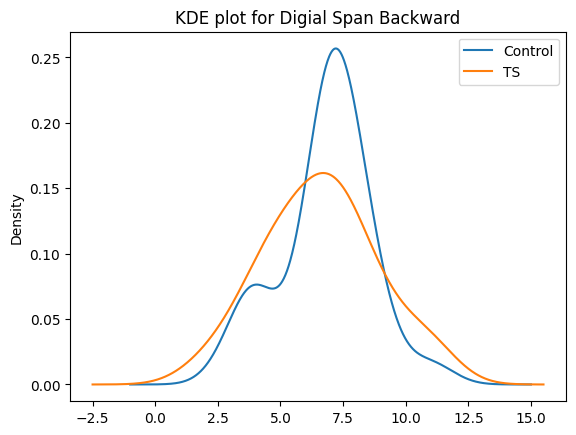


**Fig. 2:** Analysis of TMT-B for Healthy control and Tourette syndrome group of patients.

Figure 1 to Figure 7 shows Kernel Distribution Estimation (KDE) density plots for various cognitive testing data and olfactory testing data. From Figure 1 and Figure 2 it is evident that Tourette patients have lower TMT-A and TMT-B than the healthy control.

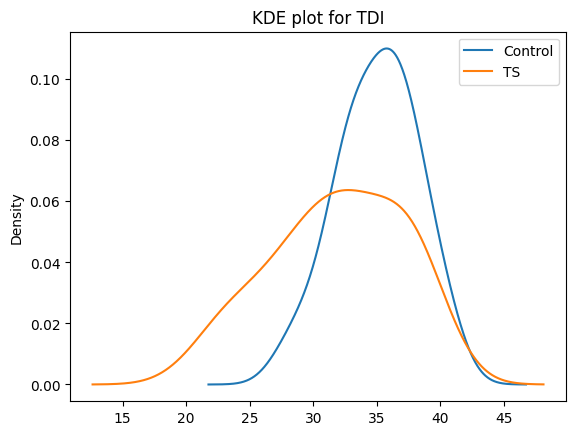
****

**Fig. 3:** Analysis of Digital Span Forward for Healthy control and Tourette Syndrome group of patients.

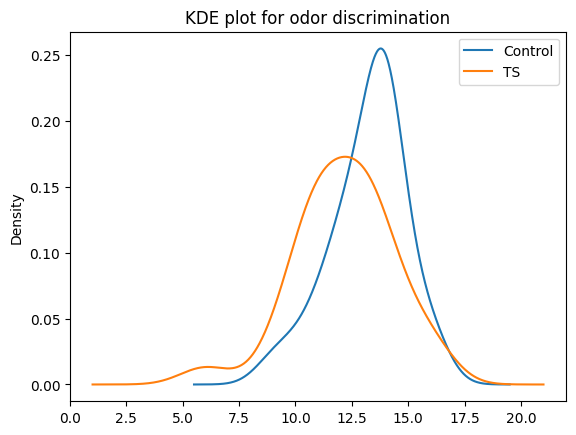
****

**Fig. 4:** Analysis of Digital Span backward for Healthy control and Tourette syndrome group of patients

From Figure 3 it is evident that both healthy control and TS patients have almost similar Digital span forward. From Figure 4 it is evident that TS patients have lower Digital span forward than healthy control.

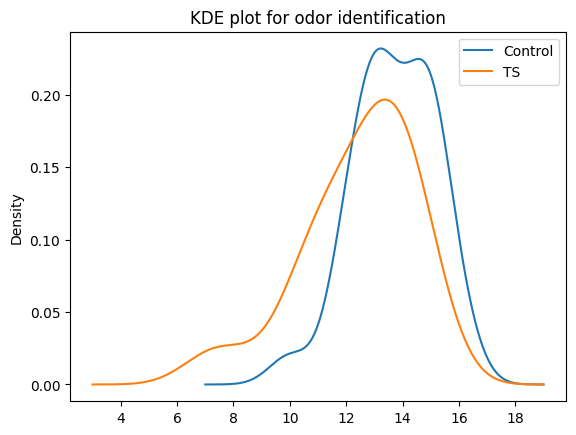


**Fig. 5:** Analysis of TDI for Healthy control and Tourette syndrome group of patients.



**Fig. 6:** Analysis of Odor Discrimination for Healthy control and Tourette Syndrome group of patients.

From Figure 5, Figure 6 and Figure 7it is evident that TS patients have lower TDI and lower odor discrimination and lower odor identification than healthy control.



**Fig. 7**: Analysis of TMT-B for Healthy control and Tourette syndrome group of patients.

**IV. PROPOSED WORK:**

This Section presents the flow chart of our proposed work. Figure 8 presents workflow for proposed model for detection of Tourette syndrome.

**Fig. 8**: Proposed workflow

First we have obtained Tourette syndrome data from[5]. Then we did preprocessing for finding the missing value if any in data. We found that there were no null or missing values in the data. Then we split the preprocessed data in train and test split set. We have used four machine learning algorithms (classifiers) i.e. Decision tree, K nearest neighbors, SGD (stochastic gradient descent) and MLP (Multi Layer Perceptron) for detection of Tourette syndrome.. We trained and tested these classifiers and then we have calculated five performance measures i.e. accuracy, Mathew’s correlation coefficient, F1-score, precision and Recall for the evaluation of these machine learning models.

**V. MACHINE LEARNING ALGORITHMS**

In this Section we have mentioned machine learning algorithms which we have trained for detection of Tourette syndrome.

*A. Decision Tree*

The Decision Trees overview provides a tree-based model for iteratively splitting data based on feature cutoff values. . Split creates subsets by splitting instances into subsets. The intermediate subsets are called interior nodes and the leaves are called leaf nodes.  Decision trees are most useful when there are significant interactions between features and goals.

*B. K-Nearest Neighbour*

Evelyn Fix and Joseph Hodges developed the nonparametric classification method known as K-nearest neighbour classification (KNN) in 1951. Both classification and regression analysis are appropriate uses for KNNs. The result of KNN classification is class membership. Items are categorized using a voting system. The distance between two data samples is calculated using Euclidean distance methods. The average of KNN values serves as the anticipated value in the regression analysis [12]. (Fix E et al.,1989).

*C. Stochastic Gradient Descent*

As an iterative method for maximizing a differentiable objective function, stochastic gradient descent, commonly referred to as incremental gradient descent, is a stochastic approximation of gradient descent optimization. To determine the parameters or coefficients of functions that minimize a cost function, one can utilize the straightforward and effective optimization process known as stochastic gradient descent (SGD). SVM and logistic regression are two examples of linear classifiers that use it for discriminative learning under convex loss functions [13] (Pedregosa et al.*,* 2011)

*D. Multi-Layer Perceptron*

A feed forward artificial neural network model called a multilayer perceptron (MLP) transfers a set of input data to a set of outputs. An MLP consists of many layers, each wholly connected to next layer. Except for the input layer nodes, other layer nodes are neurons with nonlinear activation functions. Between the input and output layers, there may be one or more hidden nonlinear layers [13] (Pedregosa et al.*,* 2011).

**VI. EXPERIMENTAL RESULT**

We have done all experimental work using Python. We have trained four ML algorithms i.e. Decision tree (DT), K-nearest classifier, Stochastic Gradient Descent (SGD) and Multi-layer perceptron (MLP) for detection of Tourette syndrome. We have briefly mentioned about these algorithms in Section 4. The performances of models have been evaluated using accuracy, Mathew’s correlation coefficient, F1-score, Precision and Recall. Table 4 represents the evaluated performance measures. Column 1 shows the various machine learning models, column 2 shows accuracy, column 3 shows Mathews correlation coefficient, column 4 shows F1-score, column 5 shows precision and column 7 contains Recall score.

**Table 4:** Performance measure of various machine learning classifiers

| **Machine Learning Algorithms** | **Accuracy** | **MCC** | **F1- Score** | **Precision** | **Recall** |
| --- | --- | --- | --- | --- | --- |
| Decision Tree Classifier | 100% | 1 | 100% | 100% | 100% |
| K Neighbors Classifier | 70% | .81 | 81.91% | 100% | 81.91% |
| SGD Classifier | 90% | 1 | 90% | 100% | 100% |
| MLP Classifier | 91.66% | 1 | 91.66% | 100% | 100% |

From Table 4 it is evident that accuracies of Decision tree, KNN, SGD and MLP classifiers are 100%,70%, 90 % and 91.66% respectively. The Mathew’s correlation coefficients (MCC) of Decision tree, KNN, SGD and MLP classifiers are 1.81,1, and 1 respectively. The F1-score of Decision tree, KNN, SGD and MLP classifiers are 100%, 81.91%, 90% and 91.66% respectively. The Precision of all these models are 100%, and Recall of Decision tree, KNN, SGD and MLP classifiers are 100%, 81.9%, 100% and 100% respectively. Since Accuracy, F1-score, Precision and Recall of Decision Tree is 100% and MCC is 1, we are reporting this model as the Best model for the prediction of Tourette syndrome

b.

c.

d.

e.

Figure 9 presents various performance measures of the machine learning classifiers for detection of Tourette syndrome. Figure 2a shows accuracy, Figure 2b shows Mathew’s correlation coefficient, Figure 2c shows F1-score, Figure 2d shows Precision and Figure 2e shows Recall. From figure 9 it is evident that the Accuracy, F1-score, Precision and Recall of Decision Tree is 100% and MCC is 1 and we are reporting this model as the Best model for the prediction of Tourette syndrome.

**VII. CONCLUSION**

In this work, we have used the demographic and cognitive and olfactory testing data to develop a computational model to predict the Tourette’s syndrome in the population. The motivation behind the work is that early prediction of syndrome gives us the opportunity to take early measures for the treatment the identified patients. In this work we to develop a machine learning oriented model to try to predict TS based on demographic and cognitive test data sets. Results show that Decision tree model is able to predict the syndrome with 100% accuracy on the limited available dataset. The research could have clinical implications for early detection auxiliary diagnosis, and assessment of therapeutic efficacy. In the future, we can test the model with more available data and further tune it for the better prediction of TS.

**REFERENCES**

[1] M. M. Robertson, V. Eapen, H. S. Singer et al., “Gilles de la Tourette syndrome,” Nature Reviews Disease Primers, vol. 3,no. 1, 2017.

[2] L. Scahill, M. Specht, and C. Page, “The prevalence of tic disorders and clinicalcharacteristics in children,” Journal ofObsessive-Compulsive and Related Disorders, vol. 3, no. 4, pp. 394–400, 2014.

[3] Peng Jialin, Wu Junya, Zhou Tianshu, GuoYufan,Tian Yu, Lou Yuting, RuHua, FengJianhua, Li Jingsong, Tic Detection in Tourette Syndrome Patients Based on Unsupervised Visual Feature Learning, Journal of Healthcare Engineering, ISSN: 2040-2295, Vol 2021 June 07, 2021, Article ID 5531186, doi.org/10.1155 /2021/5531186

[4] M. Bernabei, G. Andreoni, M. O. Mendez Garcia et al., “Automatic detection of tic activity in the Tourette Syndrome,”inProceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology,pp. 422–425, Qu´ebec, Canada, July 2010.

[5] Kronenbuerger, Martin; Belenghi, Patrizia; Ilgner, Justus; Freiherr, Jessica; Hummel, Thomas; Neuner, Irene (2018): Raw data.. PLOS ONE.Dataset. https://doi.org/10.1371/journal.pone.0197598.s001

[6] GengFudi, Ding Qiang, Wu Wanyu, Wang Xiangyang, Li Yanping, Sun Jinhua, Wang Rui, Light-efficient channel attention in convolutional neural networks for tic recognition in the children with tic disorders,Frontiers in Computational Neuroscience, vol 16,2022 DOI=10.3389/fncom.2022.1047954

[7] Wu, Junya, Zhou, Tianshu, Guo, Yufan, Tian, Yu, Lou, Yuting, Ru, Hua, Feng, Jianhua, Li, Jingsong (2021) Tic Detection in Tourette Syndrome Patients Based on Unsupervised Visual Feature Learning doi.org/10.1155/2021/5531186

[8] BaruaArnab, Dong Chunxi, Yang, Xiaodong. A deep learning approach for detecting tic disorder using wireless channel information, Transactions on Emerging Telecommunications Technologies, 2021 July 1, Vol 32, DOI 10.1002/ett.3964

[9] Wen Hongwei,., Liu Yue,., RekikIslem,., Wang Shengpei,., Chen., Zhiqiang, Jishui, Yue, Zhang., Yun Zhang., , Peng., He. Huiguang, "Multi-modal multiple kernel learning for accurate identification of Tourette syndrome children." Pattern Recognition, 63 (2017).:601-611. doi: 10.1016/J.PATCOG. 2016.09. 039

[10] Greene DJ, Church JA, Dosenbach NU, Nielsen AN, Adeyemo B, Nardos B, Petersen SE, Black KJ, Schlaggar BL. Multivariate pattern classification of pediatric Tourette syndrome using functional connectivity MRI. Dev Sci. 2016 Jul;19(4):581-98. doi: 10.1111/desc.12407. Epub 2016 Feb 1. PMID: 26834084; PMCID: PMC4945470.

[11] Suzuki Hiroyuki, Sakuma Naoko, Kobayashi Momoko, Ogawa Susumu, Inagaki Hiroki, EdahiroAyako, Ura Chiaki, Sugiyama Mika, Miyamae Fumiko, Watanabe Yutaka, Shinkai Shoji, Awata Shuichi, Normative Data of the Trail Making Test Among Urban Community-Dwelling Older Adults in Japan, Frontiers in Aging Neuroscience Vol 14, 2022, DOI : 10.3389/fnagi.2022.832158

[12] Fix, E.; Hodges, J.L. Discriminatory analysis. Nonparametric discrimination: Consistency properties. Int. Stat. Rev. Int. De Stat. 1989, 57, 238–247.

[13] Pedregosa et al., 2011 Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.