Comparison of Machine Learning Methods for Classification of Alexithymia in Individuals With and Without Autism from Eye-Tracking Data

Furkan Ilgin
Old Dominion University

Megan A. Witherow
Old Dominion University, mwith010@odu.edu

Khan M. Iftekharuddin
Old Dominion University, kiftekha@odu.edu

Original Publication Citation
Furkan Ilgin, Megan A. Witherow, and Khan M. Iftekharuddin "Comparison of machine learning methods for classification of alexithymia in individuals with and without autism from eye-tracking data", *Proc. SPIE* 12675, Applications of Machine Learning 2023, 126750P (October 4, 2023). https://doi.org/10.1117/12.2682724

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Comparison of machine learning methods for classification of alexithymia in individuals with and without autism from eye-tracking data

Furkan Ilgin, Megan Witherow, Khan Iftekharuddin

Furkan Ilgin, Megan A. Witherow, Khan M. Iftekharuddin, "Comparison of machine learning methods for classification of alexithymia in individuals with and without autism from eye-tracking data," Proc. SPIE 12675, Applications of Machine Learning 2023, 126750P (4 October 2023); doi: 10.1117/12.2682724

SPIE.

Event: SPIE Optical Engineering + Applications, 2023, San Diego, California, United States

Comparison of Machine Learning Methods for Classification of Alexithymia in Individuals with and without Autism from Eye-Tracking Data

Furkan Ilgin, Megan A. Witherow, Khan M. Iftekharuddin*
Vision Lab, Dept. of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA USA 23529

ABSTRACT

Alexithymia describes a psychological state where individuals struggle with feeling and expressing their emotions. Individuals with alexithymia may also have a more difficult time understanding the emotions of others and may express atypical attention to the eyes when recognizing emotions. This is known to affect individuals with Autism Spectrum Disorder (ASD) differently than neurotypical (NT) individuals. Using a public data set of eye-tracking data from seventy individuals with and without autism who have been assessed for alexithymia, we train multiple traditional machine learning models for alexithymia classification including support vector machines, logistic regression, decision trees, random forest, and multilayer perceptron. To correct for class imbalance, we evaluate four different oversampling strategies: no oversampling, random oversampling, SMOTE, and ADASYN. We consider three different groups of data: ASD, NT, and combined ASD+NT. We use a nested leave-one-out cross validation strategy to perform hyperparameter selection and evaluate model performance. We achieve F1 scores of 90.00% and 51.85% using decision trees for ASD and NT groups, respectively, and 72.41% using SVM for the combined ASD+NT group. Splitting the data into ASD and NT groups improves recall for both groups compared to the combined model.

Keywords: Alexithymia, autism, eye-tracking, machine learning, imbalanced data, oversampling, decision tree, support vector machine

1. INTRODUCTION

Alexithymia, a personality trait that affects how an individual experiences emotions, is characterized by difficulties in processing one's own emotions, including identifying one's feelings and vocalizing them to others[1]. About ten percent of the general population is affected by clinically relevant alexithymia[1], though it is more prevalent among individuals with anxiety and mood disorders, neurodegenerative disorders, and neurodevelopmental disorders[2]. It has been estimated that approximately 50% of individuals with Autism Spectrum Disorder (ASD) have co-occurring clinically relevant alexithymia[3]. Recently, the 'alexithymia hypothesis'[4], i.e., that differences in emotional processing may be attributed to co-occurring alexithymia rather than ASD per se, has been gaining increasing attention.

For over a decade, eye-tracking has shown to be a reliable and objective instrument for ASD research[5]. While few historical studies account for co-occurring alexithymia[2], several recent studies suggest that differential gaze to social stimuli in ASD may be modulated by alexithymia[6,7]. However, other studies incorporating dynamic social stimuli have found that ASD-related traits still contribute to differences in emotional processing even when controlling for alexithymia[8,9]. Such studies suggest that alexithymia is only part of the full explanation of differences in gaze between neurotypical (NT) individuals and individuals with ASD[8,9].

Given the negative impact that alexithymia may have on emotional health outcomes[2,10], machine learning approaches have been introduced to predict alexithymia in general and patient populations. For example, machine learning has been applied to predict alexithymia in fibromyalgia patients using socio-demographic data and self-report questionnaires[10] and in the general population using physiological signals[11], cognitive behavioral therapy concepts[12], and a facial emotion recognition task[13]. Although machine learning applied to eye-tracking data has provided multiple contributions to ASD research including supporting diagnosis and intervention efforts[5], to our knowledge, machine learning of eye-tracking features has not been used to predict alexithymia in individuals with and without ASD.

*kiftekha@odu.edu

Applications of Machine Learning 2023, edited by Michael E. Zelinski, Tarek M. Taha, Barath Narayanan Narayanan, Proc. of SPIE Vol. 12675, 126750P
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2682724

In this work, we study multiple machine learning methods for the classification of alexithymia in an eye-tracking data set collected from 70 individuals with and without ASD[6]. Given the relatively small sample size, we focus on traditional machine learning methods including support vector machines (SVMs), logistic regression (LR), decision trees (DTs), random forests (RFs), and multilayer perceptron models (MLPs). Based on the literature, we anticipate some distributional differences in the data collected from individuals with and without ASD, which may affect the performance of the machine learning models[14]. Thus, we perform experiments on three selections of the data: individuals with ASD, NT individuals, and combined (ASD and NT). The remainder of this paper is organized as follows. Section 2 covers relevant background. Section 3 describes the methods. Section 4 presents and discusses the results. Section 5 concludes and suggests future work.

2. BACKGROUND

The traditional machine learning pipeline begins with data preprocessing and feature extraction. For eye-tracking studies, the raw data takes the form of time-sampled x and y coordinates that encode the locations of study participants' eye fixations, i.e., where the participants are looking at any given moment[5]. These fixations may fall into areas of interest (AOIs) that have been defined for the given stimuli[5]. For example, AOIs for videos of dynamic facial expression stimuli may include the eyes and mouth, among others[6]. Standard fixation measures include the number and duration of fixations to different stimuli and the AOIs within these stimuli[5,6]. These fixation measures have also been considered as features for machine learning[5].

Traditional machine learning approaches, including SVM, LR, DT, RF, and MLP, have been applied to eye-tracking data for ASD research[5]. For classification tasks, SVM aims to find the best hyperplane in the high-dimensional feature space to maximally separate different classes of data points. In feature spaces where the classes are not linearly separable, the kernel trick, e.g., using the Radial Basis Function (RBF) kernel or others, allows for efficient mapping into a higher dimensional space to improve the class separability. LR predicts the probability of a binary outcome using the logistic function and a linear combination of input features. DT is a non-parametric supervised learning method that builds a flowchart-like model, with nodes representing features and branches representing decisions, to classify or predict outcomes based on input data. RF is an ensemble learning method that combines multiple decision trees to make predictions by aggregating the results of individual trees. MLP is a type of artificial neural network that consists of multiple layers of interconnected artificial neurons, enabling it to learn complex patterns.

Class imbalance occurs when a data set has an unequal number of samples for each of its constituent classes. Class imbalance poses a challenge for machine learning and may negatively impact performance. Resampling the training data is a common approach to address imbalance. To preserve as much information from the training set as possible, oversampling the minority class(es) is typically preferred to undersampling the majority class(es) for small data sets. Oversampling can be achieved by randomly replicating minority training samples (i.e., random oversampling) or generating synthetic samples for the minority class. For the latter approach, the synthetic minority oversampling technique (SMOTE)[15] and the adaptive synthetic (ADASYN)[16] technique are two well-established methods. SMOTE works by generating synthetic samples between each of the samples and its nearest neighbors[15]. ADASYN uses a weighted distribution to generate more synthetic samples based on samples that are more difficult to classify using k-nearest neighbors[16].

Classification seeks to predict labels y from features X based on the conditional probability $p(y|X)$ which relies on the assumption that the samples are independent and identically distributed (i.i.d.). Predictive performance may suffer when this assumption is violated, such as when data are drawn from different domains or distributions[14].

3. METHODS

3.1 Data

We consider the publicly available eye-tracking data set published in Ref. 6. The data set consists of 70 adult participants, including 45 NT participants (29 female) and 25 diagnosed with ASD (11 female)[6]. All participants have completed the Toronto Alexithymia Scale (TAS)[6]. Each participant has time-sampled x and y eye fixation coordinate data collected using a screen-mounted Tobii TX300 eye-tracker while the participant viewed validated dynamic facial expression stimuli taken from Ref. 17. The dynamic facial expression stimuli is made up of videos of four male and four female actors displaying six emotions (neutral, anger, disgust, fear, happy, sad)[17]. Each emotion is presented eight times for each of four conditions: free-gaze, cued, emotion recognition, and intensity judgement[6]. For the cued and free-gaze conditions, the participant is instructed to gaze freely at the videos with or without, respectively, being told the emotion beforehand[6]. For the emotion recognition and intensity judgment tasks, the participant is asked to determine the emotion or intensity of emotion, respectively, that is being shown[6]. For all videos, fixations fall into AOIs defined for the eyes, mouth, or nose, or no AOI[6].

3.2 Feature Extraction and Oversampling Strategies

Following Ref. 6, we discard the first 150 milliseconds of each trial representing rapid reorienting of gaze to stimuli. Then, for each emotion (neutral, anger, disgust, fear, happy, sad) within each condition (free-gaze, cued, emotion recognition, and intensity judgement), we accumulate the number and duration of fixations to each AOI (eyes, mouth, nose, or no AOI). This yields a total of $4\,conditions \times 6\,emotions \times 4\,AOIs \times 2\,(number\ of\ fixations\ or\ duration) = 192\,features$. Examples of these 192 features include 'number of fixations to the eyes AOI for cued anger' and 'duration of gaze to the mouth AOI for happy intensity judgement'. For each feature, we perform min-max normalization based on the training data, which scales the features to the range (0,1), and we apply the same transformation to validation and test sets. Using the TAS scores, we generate the binary class labels for alexithymia classification as follows. We assign '0' if the score is ≤51, which corresponds to the published TAS cutoff for an absence of alexithymia, and we assign '1'otherwise, which corresponds to ranges for 'possible alexithymia' (TAS score between 52 - 60) or 'alexithymia' (TAS score >=61)[18].

In addition to no oversampling, we consider random oversampling, SMOTE, and ADASYN as implemented in the imbalanced-learn library (https://imbalanced-learn.org/) to correct for class imbalance in the training data. For SMOTE and ADASYN, we use three nearest neighbors to define the neighborhood for generating the synthetic samples.

3.3 Machine Learning Models

We consider five different traditional machine learning methods as implemented in the scikit-learn library (https://scikit-learn.org/): SVM, LR, DT, RF, and MLP. Using validation data, we perform hyperparameter selection for each of the five methods. For SVM, we consider linear and RBF kernel functions. To tune the l2 regularization parameter 'C' and kernel coefficient 'gamma', we consider a logarithmic search space from 10^{-3} to 10^{3}. For LR, we consider four different options for the norm of the penalty: no penalty, l1, l2, and elastic net. For regularization parameter 'C', we consider a logarithmic search space from 10^{-3} to 10^{3}. When using the elastic net penalty for LR, we search [0.1, 0.2, …, 0.9] for the best l1 ratio to mix the contributions of the l1 and l2 penalties. We use the SAGA solver which supports all penalties for LR. For DT, we consider Gini impurity, log loss, and entropy for measuring split quality. We search [1, 2, …, 12] for the optimal maximum depth of the tree. For RF, we consider the same options as DT for split criterion and maximum tree depth. Additionally, we select among [2, 5, 10, 100, 1000] for number of estimators in the ensemble. For MLP, we consider four possible activation functions ('identity', 'logistic', 'tanh', and 'relu') and choose the best number of hidden units within [2, 4, 8, 16, 32, 64, 128, 256, 512, 1024]. Given the small size of the data set, it is feasible to use the second derivative in the optimization. Therefore, we choose to optimize MLP using L-BFGS as opposed to SGD based methods that utilize the first derivative only.

3.4 Experiments

We use a nested leave-one-out cross validation (LOOCV) strategy to evaluate our approach. For N samples, the outer LOOCV loop splits the data into N-1 training samples and 1 test sample. Then, the inner LOOCV loop splits the data again into N-2 training samples and 1 validation sample. The inner LOOCV loop is used to determine the best selection of hyperparameters for a particular outer LOOCV split. For each hyperparameter combination, the N-1 executions of the inner loop yield N-1 predictions, one for each validation sample. These predictions are compared with the ground truth labels to determine the number of true positives, false positives, true negatives, and false negatives, which are subsequently used to generate a validation F1 score for each combination of hyperparameters. The best selection of hyperparameters based on the validation performance is then used to retrain the model on the N-1 training samples and predict the label for the left-out test sample. The outer loop repeats N times, such that each sample is considered as the test sample once. The predictions for each test sample and corresponding ground truth labels are used to compute F1 score, precision, recall, and accuracy. We apply this procedure for three sets of data: ASD, NT, and combined ASD+NT.

4. RESULTS AND DISCUSSION

Figure 1 shows the class distributions for the ASD, NT, and combined ASD+NT data sets. All three data sets exhibit class imbalance, with a minority of samples representing the alexithymia class. This imbalance is worse for the NT data set than the ASD data set.

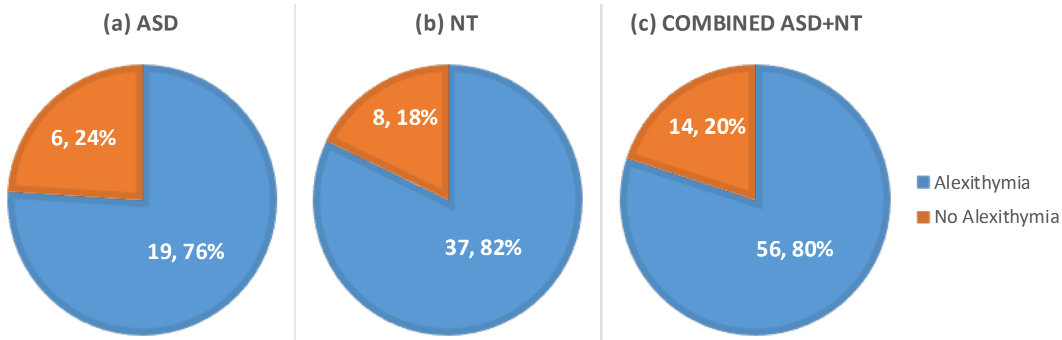

Figure 1. Number and percentage of samples for each class (alexithymia or no alexithymia) for three sets: (a) ASD, (b) NT, and (c) combined ASD+NT.

Figure 2 shows the comparative performance of five different traditional machine learning methods (SVM, LR, DT, RF, and MLP) and four oversampling strategies (no oversampling, random oversampling, SMOTE, and ADASYN) for classification of alexithymia in three sets of data (ASD, NT, combined ASD+NT). As shown in Figure 2, no oversampling performs the best for the ASD data set across all machine learning methods, while for the NT data set, all oversampling strategies perform equally well or better than no oversampling across all machine learning methods. For the combined ASD+NT group, the best oversampling strategy is model-dependent. We expect that the differences in performance of oversampling strategies across the three different sets of data may be attributed to differences in the class conditional distributions and separability of classes in the feature space.

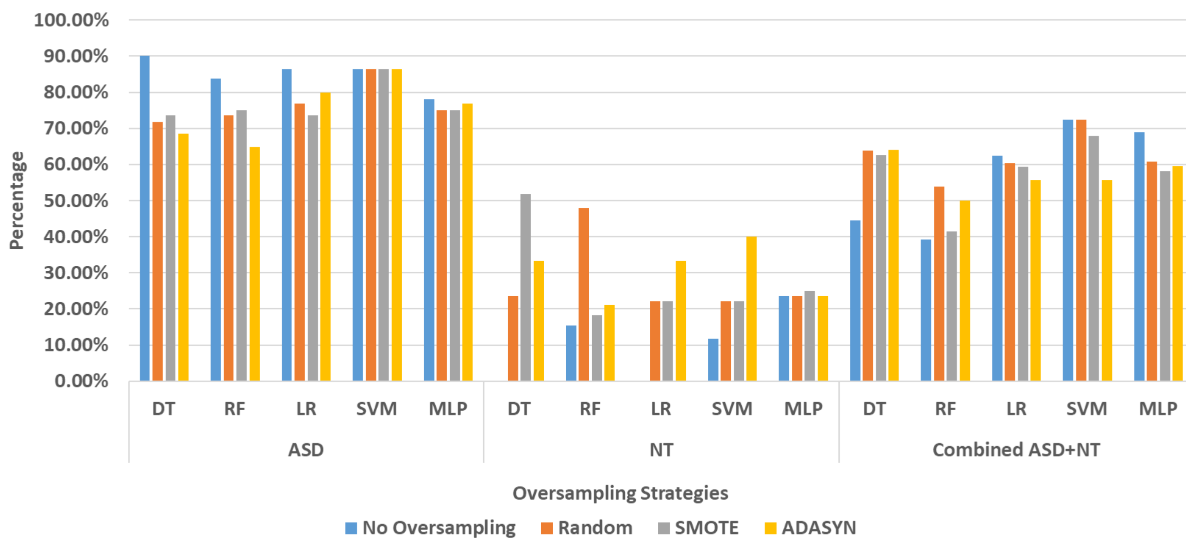

Figure 2. Comparison of LOOCV F1 scores for classification of alexithymia in ASD, NT, and Combined ASD+NT eye-tracking data sets with five machine learning methods and four oversampling strategies.

In Figure 2, the best model and oversampling strategy is DT with no oversampling for the ASD set, DT with SMOTE for the NT set, and SVM with no oversampling for the combined NT+ASD set. The best hyperparameter selections for the best model and oversampling strategy for each data set are as follows. For the ASD set's best performing model, DT with no oversampling, Gini impurity is selected as the splitting criterion 19 times, log loss is selected 3 times, and entropy is selected 3 times. A max tree depth of 1 is selected 23 times and a max depth of 2 is selected 2 times. For the NT set's best performing model, DT with SMOTE, Gini impurity is selected as the splitting criterion 43 times and log loss is selected 2 times. The frequencies of selecting each possible value for max tree depth are '1': 25, '2': 0, '3':6, '4':1, '5': 0, '6': 1, '7': 2, '8':1, '9':2, '10':1, '11':4, and '12':2. The linear kernel is selected all 70 times for SVM with no oversampling, the best model for the combined ASD+NT set. A 'C' value of 10.0 is selected 69 times and 1.0 is selected once. We report the detailed accuracy, precision, recall, and F1 scores for these best models with their associated best performing oversampling strategy in Figure 3. As shown in Figure 3, splitting the combined ASD+NT data set into separate ASD and NT sets improves performance across all metrics for the ASD set. Splitting the data improves recall for the NT set compared to the combined set. However, the best model for the NT set does not perform as well as the combined model for other metrics.

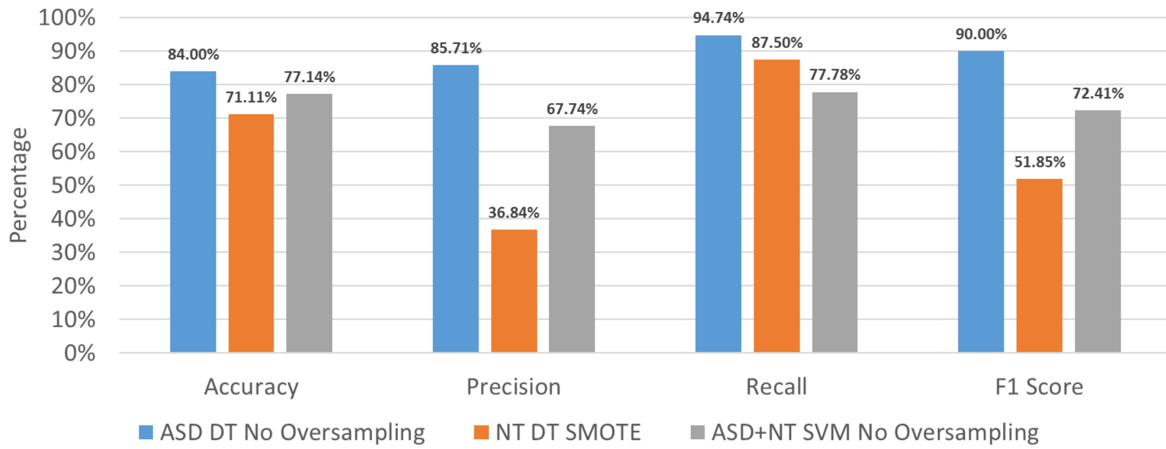

Figure 3. Comparison of LOOCV F1 scores for classification of alexithymia in ASD, NT, and Combined ASD+NT eye-tracking data sets with five machine learning methods and four oversampling strategies.

5. CONCLUSION AND FUTURE WORK

This work compares five different traditional machine learning methods and four different oversampling strategies for alexithymia classification based on eye-tracking data collected from NT individuals and individuals with ASD. We perform rigorous hyperparameter selection and evaluation to identify the best performing approaches. Furthermore, we find that partitioning data into ASD and NT sets improves recall and offers overall improved performance on data collected from individuals with ASD. In the future, we plan to analyze feature importance to identify and explain the features that contribute most to the classification. We also plan to further investigate the model decision boundaries and data distributions of eye-tracking data from individuals with ASD and NT individuals to explain how these may affect the selection of an appropriate correction for class imbalance and their affect on model classification performance.

ACKNOWLEDGEMENTS

This material is partially supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1753793 and by the Research Computing clusters at Old Dominion University under National Science Foundation Grant No. 1828593.

REFERENCES

[1] K. S. Goerlich, "The Multifaceted Nature of Alexithymia – A Neuroscientific Perspective," Frontiers in Psychology, 9, (2018).
[2] J. Hogeveen, and J. Grafman, "Alexithymia," Handb Clin Neurol, 183, 47-62 (2021).
[3] E. Kinnaird, C. Stewart, and K. Tchanturia, "Investigating alexithymia in autism: A systematic review and meta-analysis," Eur Psychiatry, 55, 80-89 (2019).
[4] G. Bird, and R. Cook, "Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism," Transl Psychiatry, 3(7), e285 (2013).

[5] K.-F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis, "The Contribution of Machine Learning and Eye-Tracking Technology in Autism Spectrum Disorder Research: A Systematic Review," Electronics, 10(23), 2982 (2021).
[6] H. C. Cuve, S. Castiello, B. Shiferaw, E. Ichijo, C. Catmur, and G. Bird, "Alexithymia explains atypical spatiotemporal dynamics of eye gaze in autism," Cognition, 212, 104710 (2021).
[7] H. C. Cuve, Y. Gao, and A. Fuse, "Is it avoidance or hypoarousal? A systematic review of emotion recognition, eye-tracking, and psychophysiological studies in young adults with autism spectrum conditions," Research in Autism Spectrum Disorders, 55, 1-13 (2018).
[8] C. T. Keating, D. S. Fraser, S. Sowden, and J. L. Cook, "Differences Between Autistic and Non-Autistic Adults in the Recognition of Anger from Facial Motion Remain after Controlling for Alexithymia," Journal of Autism and Developmental Disorders, 52(4), 1855-1871 (2022).
[9] K. G. Stephenson, S. G. Luke, and M. South, "Separate contributions of autistic traits and anxious apprehension, but not alexithymia, to emotion processing in faces," Autism, 23(7), 1830-1842 (2019).
[10] G. Orrù, A. Gemignani, R. Ciacchini, L. Bazzichi, and C. Conversano, "Machine Learning Increases Diagnosticity in Psychometric Evaluation of Alexithymia in Fibromyalgia," Frontiers in Medicine, 6, (2020).
[11] V. Filippou, N. Theodosiou, M. Nicolaou, E. Constantinou, G. Panayiotou, and M. Theodorou, "A Wavelet-based Approach for Multimodal Prediction of Alexithymia from Physiological Signals," Series A Wavelet-based Approach for Multimodal Prediction of Alexithymia from Physiological Signals. 177–184 (2022).
[12] M. K. Yöntem, and K. Adem, "Prediction of the Level of Alexithymia through Machine Learning Methods Applied to Automatic Thoughts," Otomatik Düşüncelere Makine Öğrenme Yöntemlerinin Uygulanması ile Aleksitimi Düzeyinin Tahmini., 11, 64-78 (2019).
[13] N. Farhoumandi, S. Mollaey, S. Heysieattalab, M. Zarean, and R. Eyvazpour, "Facial Emotion Recognition Predicts Alexithymia Using Machine Learning," Computational Intelligence and Neuroscience, 2021, 2053795 (2021).
[14] J. Dockès, G. Varoquaux, and J.-B. Poline, "Preventing dataset shift from breaking machine-learning biomarkers," GigaScience, 10(9), (2021).
[15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," J. Artif. Int. Res., 16(1), 321–357 (2002).
[16] H. Haibo, B. Yang, E. A. Garcia, and L. Shutao, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning." 1322-1328 (Year).
[17] N. Yitzhak, N. Giladi, T. Gurevich, D. S. Messinger, E. B. Prince, K. Martin, and H. Aviezer, "Gently does it: Humans outperform a software classifier in recognizing subtle, nonstereotypical facial expressions," Emotion, 17(8), 1187-1198 (2017).
[18] R. M. Bagby, J. D. Parker, and G. J. Taylor, "The twenty-item Toronto Alexithymia Scale--I. Item selection and cross-validation of the factor structure," J Psychosom Res, 38(1), 23-32 (1994).