Ataxic person prediction using feature optimized based on machine learning model

Pavithra Durganivas Seetharama¹, Shrishail Math²

¹Department of Computer Science and Engineering, Canara Engineering College, Affiliated to Visvesvaraya Technological University, Mangalore, India
²Department of Computer Science and Engineering, Rajeev Institute of Technology, Affiliated to Visvesvaraya Technological University, Hassan, Karnataka, India

ABSTRACT

Ataxic gait monitoring and assessment of neurological disorders belong to important areas that are supported by digital signal processing methods and artificial intelligence (AI) techniques such as machine learning (ML) and deep learning (DL) techniques. This paper uses spatio-temporal data from Kinect sensor to optimize machine learning model to distinguish between ataxic and normal gait. Existing ML-based methodologies fails to establish feature correlation between different gait parameters; thus, exhibit very poor performance. Further, when data is imbalanced in nature the existing ML-based methodologies induces higher false positive. In addressing the research issues this paper introduces an extreme gradient boost (XGBoost)-based classifier and enhanced feature optimization (EFO) by modifying the standard cross validation (SCV) mechanism. Experiment outcome shows the proposed ataxic person identification model achieves very good result in comparison with existing ML-based and DL-based ataxic person identification methodologies.

Keywords: Ataxic person identification, Binary classification, Class imbalance, Deep learning, Feature extraction, Feature selection, Machine learning

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

Motion disorders plays a key indicator in identifying many diseases as described in [1], [2] such as physical activity, rehabilitation, physical therapy, orthopedics, rheumatology, and neurology. As stated in [1], around 70% of neurological inpatients exhibit abnormal gait activity; thus, gait-based assessment model can be leveraged for early diagnosis [3] of neurological disorder as demonstrated in [4], [5]. The research work mainly focusses on identifying ataxic neurological disorder through gait analysis as demonstrated in [6]. Designing efficient gait-based mechanism that is skilled for automatic detection and nursing of neurological condition aid in enhancing treatment efficiency, diseases management and also further aid in reducing load of healthcare management environment as stated in [7], [8]. Tool and methodologies are needed for efficient detection and diagnosis of neurological disorders as shown in [9], [10]. The significant growth of wireless communication and sensor technology have led to usage of different sensors such as wearable devices [11], video, depth and thermal camera systems [12], microelectromechanical sensor units, and Kinect sensor [13] for monitoring different neurological disorder [14]. Recently, in [15] presented wearable gait sensor are efficient in studying the behavior of scale for the assessment and rating of ataxia (SARA). Similarly, in [16] showed importance of studying gait characteristic of ataxic patient with multiple sclerosis [17], [18].

Journal homepage: http://ijece.iaescore.com
Ataxic person prediction using feature optimized based on … (Pavithra Durganivas Seetharama)
classify hereditary ataxia using gait features collected from motion sensors. They focused in establishing minimum gait feature required to perform hereditary ataxia classification using ankle-based motions sensors. They considered total 28 participant out of which 14 are health and remaining 14 are hereditary ataxia patient. The spatio-temporal gait features are extracted considering stride window size of 10. Then, for reducing the gait features using ankle sensors Hill Climbing model is used are trained with different ML models such as multi-layer perceptron (MLP) and k-nearest neighbors (KNN) model; the model achieved an accuracy of 96%.

In [23] studied the benefit of using accelerometer sensor for extraction of gait features and train it with deep learning model such as convolution neural network for classifying whether a person is ataxic or healthy. In conducting experiment total 35 patient are considered out of which 19 people are health and 16 are ataxic person. The DL-based classification algorithm is compared with standard ML-based model such as two-layer neural network (TLNN), SVM, and Bayesian network. They identified accuracies varies with respect to placement of sensors and DL-based method achieves an accuracy of 95.8. However, the model performs badly considering limited training sample and when data is imbalanced in nature. In addressing the research limitation, the following research methodology is presented.

3. PROPOSED METHODOLOGY

The section introduces a methodology of feature optimization model for designing effective ataxic person identification model. In this work first provide mathematical representation of gradient Boosting tree; then, provide the detailed discussion of feature weight optimization to reduce misclassification and feature optimization process to handle data imbalance. The architecture of proposed ataxic person identification model using enhanced ML model is given in Figure 1.

3.1. XGBoost prediction algorithm

The XGBoost algorithm are very efficient in the field of classification task across different domain such as education, agriculture, healthcare. The XGB takes the input \( y_j \) and the corresponding actual label is
defined using parameter $z_j$ and the initial prediction outcome prior to sigmoid is defined through parameter $a_j$ and the objective function of XGBoost algorithm is expressed through following equation.

$$M^{(u)} = \sum_{j=1}^{o} m\left(z_j A_j^{(u-1)} + g_u(y_j)\right) + \rho(g_u) + d \quad (1)$$

where the loss function is defined through parameter $m$, the parameter $\rho$ is used for penalizing the model complexities, $u$ defines respective tree, $\rho(g_u)$ defines regularization operation penalty parameter and constant are represented by parameter $d$. The second-order Taylor expansion are expressed using (2),

$$g(y + \delta y) \approx g(x) + g'(y)\delta y + \frac{1}{2} g''(y)\delta y^2$$

then substituting (2) into (1) we get,

$$M^{(u)} \approx \sum_{j=1}^{o} \left[ m\left(z_j + A_j^{(u-1)}\right) + h_j g_u(y_j) + \frac{1}{2} i_j \left(g_u(y_j)\right)^2 \right] + \rho(g_u) + d \quad (3)$$

where $h_j$ is computed as (4),

$$h_j = \frac{\partial M}{\partial a_j} \quad (4)$$

and $i_j$ is computed as (5),

$$i_j = \frac{\partial^2 M}{\partial a_j^2} \quad (5)$$

in (3) the constant term are removed for simplifying the computation at instance $u$ as (6)

$$M^{(u)} \approx \sum_{j=1}^{o} \left[ h_j g_u(y_j) + \frac{1}{2} i_j \left(g_u(y_j)\right)^2 \right] + \rho(g_u) \quad (6)$$

In fitting the gradient boosting tree model the parameter such as $h_j$ and $i_j$ need to be established. The standard loss function of gradient boosting tree in solving binary classification problem is obtained through cross entropy loss function as (7),

$$M = -\sum_{j=1}^{o} [z_j \log(z_j) + (1 - z_j) \log(1 - z_j)] \quad (7)$$

In (9), the parameter $\hat{z}_j$ is computed as (8)

$$\hat{z}_j = \frac{1}{1 + \exp(-a_j)} \quad (8)$$

and for activation sigmoid function is used and we obtain (9).

$$\frac{\partial \hat{z}_j}{\partial a_j} = \hat{z}_j (1 - \hat{z}_j) \quad (9)$$

### 3.2. Misclassification minimization aware weight optimization for XGBoost algorithm

The traditional XGB-based classification is primarily focussed in minimizing the loss function; however, the weights for both correctly and wrongly classified class remain same. However, in this work feature weight optimized (FWO) learning method is presented for minimizing misclassification. In FWO, a larger weight is given to misclassified class for error induced by positive sample; thus, higher importance is given to positive samples. In Table 1, binary classification weight matrix is provided.

<table>
<thead>
<tr>
<th>Actual ataxic person (Positive)</th>
<th>Actual non-ataxic person (Negative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predict Ataxic person (Positive)</td>
<td>TP ($T_{11}$)</td>
</tr>
<tr>
<td>Predict non-ataxic person (Negative)</td>
<td>FN ($T_{01}$)</td>
</tr>
</tbody>
</table>

Ataxic person prediction using feature optimized based on … (Pavithra Durganivas Seetharama)
The feature weight optimized XGB (FWO-XGB)-based ataxic person identification model considers the misclassification instance only. Thus, let $O_{00} = O_{11}$, $O_{10} = b(b > 0)$, and $O_{01} = 1$, the loss function using feature weight optimized influence is established using (10),

$$M_b = -\sum_{j=1}^{o}[bz_j \log(\hat{z}_j) + (1 - z_j)\log(1 - \hat{z}_j)]$$ (10)

where $b$ defines feature weight optimized influence. In above equation additional loss will be considered on false positive if $b$ is lesser than 1; however, $b$ is greater than 1, no additional loss will be considered for false negative. The first-order derivative $h_j$ is computed using (11):

$$h_j = \frac{\partial M_b}{\partial a_j}$$ (11)

The above equation can be refined as (12),

$$h_j = \hat{z}_j(1 - z_j + bz_j) - bz_j$$ (12)

and second-order derivative $I_j$ is computed using (13)

$$I_j = \frac{\partial^2 M_b}{\partial^2 a_j}$$ (13)

The above equation can be refined as (14),

$$I_j = \hat{z}_j(1 - \hat{z}_j)(1 - z_j + bz_j)$$ (14)

The performance of featured weight optimized XGB can be seriously impacted when data is imbalanced in nature; thus, required effective feature optimization mechanism for designing ideal ataxic person identification.

3.3. Feature optimization for imbalanced data

In the standard K-fold cross validation, the feature subset for training classification model is created in by dividing the dataset in random manner with equal size. Then, remaining $K - 1$ are used for construction of ataxic person identification model. Finally, the model is selected that minimize the prediction error considering grid $l$. The standard cross validation is obtained as (15),

$$CV(\sigma) = \frac{1}{M} \sum_{k=1}^{K} \sum_{j \in G_{-k}} P \left( b_j, \hat{\sigma}^{-k}(y_j, \sigma) \right)$$ (15)

The proposed feature optimized cross validation is obtained as (16)

$$CV(\sigma) = \frac{1}{SM} \sum_{s=1}^{S} \sum_{k=1}^{K} \sum_{j \in G_{-k}} P \left( b_j, \hat{\sigma}^{-k}(y_j, \sigma) \right)$$ (16)

A two-step cross validation mechanism is defined. First, main feature is chosen from feature subsets; later, feature extracted in previous steps is used for building high accuracy ataxic person identification model. In. (15) and (16), for choosing good $\hat{\sigma}$ for optimization of ataxic person identification model obtained through following (17),

$$\hat{\sigma} = \arg \min_{\sigma \in \{\sigma_1, ..., \sigma_l\}} CV_s(\sigma)$$ (17)

In (16), $M$ represent dataset size used during training, $P(\cdot)$ are the loss function, and $\hat{\sigma}^{-k}(\cdot)$ represent method used for computing its coefficients. After extracting good features in iterative manner using (16), the features are ranked $r(\cdot)$ to attain high accurate ataxic person identification model using (18),

$$r(a) = \begin{cases} 0 & \text{if } n_j \text{ is not selected} \\ 1 & \text{if } n_j \text{ is selected as optimal prediction model } j = 1, 2, 3, ..., n \end{cases}$$ (18)

The feature subset is obtained using (19)
Ataxic person prediction using feature optimized based on 
(Pavithra Durganivas Seetharama)

\[ F_k = \{r(n_1), r(n_2), ..., r(n_n)\} \quad (19) \]

For different \( K \)-folds instance good features with maximal score are computed as using (20)

\[ F_{k} = \{r(n_1), r(n_2), ..., r(n_n)\} \quad (20) \]

After that how many a respective feature has been chosen for \( K \) feature subsets have been identified for construction of final features is obtained through (21)

\[ F_{\text{final}} = \{f_{p_1}, f_{n_2}, ..., f_{n_n}\} \quad (21) \]

where \( f_i() \) defines instance where \( n^{th} \) feature is chosen or not and is established using (22).

\[ F_i(a) = \begin{cases} 0 & \text{if } q_j \text{ is chosen less than } \frac{K}{2} \text{ times, } j = 1,2,3,...,n \\ 1 & \text{if } q_j \text{ is chosen greater or equal to } \frac{K}{2} \text{ times, } j = 1,2,3,...,n \end{cases} \quad (22) \]

Using above equation feature importance of different gait features are first established. Then, based on extracted features the proposed ataxic person identification model is constructed. This allows the proposed FWO-XGB based ataxic person identification model in significantly improving the overall prediction accuracy in comparison with other DL-based and ML-based ataxic person identification schemes. In next section the performance evalution is done using standard ataxia dataset.

4. PERFORMANCE EVALUATION

This section studies the performance achieved using proposed FWO-XGB based ataxic person identification model over standard XGB-based classification model. Further, the model is compared with existing machine learning [32] and deep learning-based [17], [23] ataxic person identification model. The proposed and other existing ataxic person identification model is implemented using Anaconda python framework. The experiment is conducted using dataset collected from [34], [35]. The dataset composed of spatio-temporal data measured through Kinematic v2 sensor of 19 cerebellar ataxic person and 65 healthy persons. The accuracies, sensitivity, specificity, precision, and F-measure are metrics used for validating the classification algorithm performance. The specificity is computed as (23),

\[ \text{Specificity} = \frac{TN}{TN+FP} \quad (23) \]

where \( TP \) defines true positive, \( FP \) defines false positive, \( TN \) defines true negative, and \( FN \) defines false negative. The sensitivity is computed as (24),

\[ \text{Sensitivity} = \frac{TP}{TP+FN} \quad (24) \]

The accuracy is computed as (25),

\[ \text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (25) \]

The precision is computed as (26),

\[ \text{Precision} = \frac{TP}{TP+FP} \quad (26) \]

The F-measure is computed as (27),

\[ F - \text{measure} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (27) \]

4.1. Specificity and sensitivity performance

This section provides performance study of specificity and sensitivity metrics in classifying whether a person is ataxic or healthy. The proposed FWO-XGB is compared with total five existing machine learning algorithm [32] such as random forest (RF), AdaBoost, radial basis function support vector machine
(RBF-SVM), linear regression (LR), and XGB and one deep learning (DL) neural network algorithm [17], [23]. The graphical outcome of specificity is computed using (23) is given in Figure 2. Similarly, the graphical outcome of sensitivity is computed using (24) is given in Figure 3. From both Figures 2 and 3 we can interpret that FWO-XGB outperform all existing ML and DL-based ataxic person identification methods.

![Specificity](image1)

Figure 2. Specificity performance of different ML and DL algorithms for ataxic person identification

![Sensitivity](image2)

Figure 3. Sensitivity performance of different ML and DL algorithms for ataxic person identification

4.2. Accuracy performance

This section provides performance study of accuracies metrics in classifying whether a person is ataxic or healthy. The proposed FWO-XGB is compared with total eight existing machine learning algorithm [32] such as RF, AdaBoost, RBF-SVM, LR, XGB, Bayes, SVM, two-layer neural network (TLNN), and one DL neural network algorithm [17], [23]. The graphical outcome of accuracy is computed using (25) is given in Figure 4. From Figure 4 we can interpret that FWO-XGB outperforms all existing ML and DL-based ataxic person identification methods.

4.3. Classification performance of XGB vs FWO-XGB

This section provides performance study of standard XGB and proposed improved XGB version i.e., FWO-XGB in terms of accuracy, sensitivity, specificity, precision, and F-measure. The graphical outcome of classification performance is given in Figure 5. From Figure 5 we can interpret that FWO-XGB outperform all standard XGB-based ataxic person identification methods.
Ataxic person prediction using feature optimized based on ... (Pavithra Durganivas Seetharama)


BIOGRAPHIES OF AUTHORS

Pavithra Durganivas Seetharama received M.Tech. in Computer Science and Engineering from Oxford College of Engineering, Bangalore under Visvesvaraya Technological University. Currently she is working as assistant professor in Canara Engineering College, Benjanapadavu, Karnataka, India. She can be contacted at email: ds.pavithra88@gmail.com.

Shrishail Math holds obtained his Ph.D. from Indian Institute of Information Technology, Allahabad. Currently working as professor in Rajeev Institute of Technology, Hassan, Karnataka, India. Has guided various scholar in the field of machine learning, deep learning, artificial intelligence technique. Also published various paper in international indexed journal. He can be contacted at email: shri_math@yahoo.com.