Indonesian Journal of Electrical Engineering and Computer Science
Vol. 32, No. 3, December 2023, pp. 1605~1613
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v32.i3.pp1605-1613

Ataxia severity classification using enhanced feature selection and ranking optimization through machine learning model

Pavithra Durganivas Seetharama[1], Shrishail Math[2]
[1]Department of Computer Science and Engineering, Canara Engineering College, Affiliated to Visvesvaraya Technological University, Belagavi, India
[2]Department of Computer Science and Engineering, Rajeev Institute of Technology, Affiliated to Visvesvaraya Technological University, Belagavi, India

Article Info

Article history:

Received Apr 4, 2023
Revised Jul 18, 2023
Accepted Sep 21, 2023

Keywords:

Ataxic severity classification
Class imbalance
Deep learning
Feature extraction-selection
Machine learning
Multi-label classification

ABSTRACT

The examination of neurological disorders and the monitoring of ataxic gait are major scientific topics that benefit from digital signal processing techniques and machine learning (ML) technologies. In this research, an ML approach is optimized with the use of Spatio-temporal data obtained from a kinect-sensor to differentiate between normal gait and ataxic. The current ML-based approaches perform very poorly because they cannot build feature-correlation among many gait characteristics. Furthermore, current ML-based techniques generate more false-positive whenever data is imbalanced in nature; especially for performing multi-label classification. This work presents a feature selection and ranking (FSR) based on extreme gradient boost (XGB) for ataxia severity classification. The FSR-XGB introduce an enhanced misclassification minimization error optimization and presents a novel feature selection and ranking to introduce feature importance using new cross-validation mechanism, both of which are aimed at solving the multi-label classification research problems. Results from experiments demonstrate that the presented FSR-XGB approach outperforms other ML-based and deep learning-based approaches.

Corresponding Author:

Pavitra Durganivas Seetharama
Department of Computer Science and Engineering, Canara Engineering College
Affiliated to Visvesvaraya Technological University
Belagavi-590018, Karnataka, India
Email: researchj23@gmail.com

1. INTRODUCTION

There are several fields, including physiology, rehabilitation, physiotherapy, orthopedics, arthritis, and neuroscience, in which movement abnormalities serve [1] as crucial diagnostic indicators [2]. Approximately 70% of neurological hospitalized patients have abnormal gait activity, Alsaif *et al.* [2], suggesting that the gait based evaluation method [3] can be used to make an earlier identification of neurological disease [4]. The primary goal of this study is to use gait analysis to diagnose ataxic neurological-disorder [5]. Developing a gait based method capable of automated identification and monitoring of neurological conditions improves the effectiveness of the treatment, facilitates treatment strategies, and helps to ease the burden on the healthcare-management system [6]. Effective diagnosis and detection of neurological-disorders require the use of appropriate tools and approaches [7]. Wearable gadgets [8], video, and thermal-camera devices [9], microelectromechanical-sensor-units [10], as well as the kinect-sensor [11], are just some examples of the sensor technologies that have become increasingly common due to advancements in wireless technology as well as sensing technologies. Wearable-gait sensors, as recently shown in [12], are useful for

investigating scale-for-the-assessment and rating-of-ataxia (SARA) behavior. The significance of investigating the gait characteristics of ataxic patients [13] with multiple-sclerosis was also demonstrated in [14].

Flat-fall prediction [15], Parkinson's [16], and Friedreich's-ataxia [17] are just a few of the neurological-disorders for which machine-learning techniques have now been implemented for diagnosis and detection. Concerning cerebellar-ataxic human-identification [18] provided a deep learning-based solution and evaluated it to other machine learning (ML) models. Current deep-learning (DL)-based methods [19] are all applied to the problem of predicting motion disorders [20] by making sense of the spatio-temporal-kinematics [21] of a typical human normal movement; when applied to ML-based, DL methodologies helps with both feature-selection for spatio-temporal information and make decisions. The DL-based algorithms [22] have several drawbacks, including a prerequisite higher computational resource, a large number of training data, and poor results of prediction of ataxia type/severity (i.e., multi-label) classification and also when the ataxia-type classification data is imbalanced in nature. This fact serves as motivation for the proposed research, which seeks to develop a better feature selection optimization employing ML method for ataxia type classification utilizing spatio-temporal-kinematic gait characteristics. This paper introduces a feature selection and ranking (FSR) based on extreme gradient boosting (XGB). The FSR-XGB model first provides the working of XGB and then optimize the loss function to reduce the misclassification. Then, describes standard cross-validation and provide a two-level cross validation (CV) for better selection of feature and ranking with high importance. Finally, the multi-label descriptor is constructed with maximum feature importance score and ataxia severity classification is done. The significance of FSR-XGB for ataxia severity classification is as follows:

- The FSR-XGB is very efficient in reducing misclassification using new loss function; better specificity and sensitivity assures the efficiency of new loss function.
- The FSR-XGB is efficient in distinguishing different severity level of ataxia considering imbalanced data; this is due to adoption of new feature selection and ranking method using novel two-level cross validation methods.
- The FSR-XGB achieves better accuracy, specificity, and sensitivity than existing ML and DL-based ataxia severity classificaon approaches.
- The FSR-XGB provides higher weight in comparison with XGB-based method thus, aided FSR-XGB to attain better accuracy, specificity, sensitivity, precision, and F-measure than XGB-based ataxia severity classification approaches.

The paper is organized as follows: in section 2, various existing ataxia severity and neurodegenerative disease identification using ML and DL method is given along with benefits and limitations. The section 3, the proposed method of FSR-XGB based ataxia severity classification is provided. The section 4, the outcome of FSR-XGB based ataxia severity classificaon method over various ML and DL-based ataxia severity classification method is provided. Lastly, the research significance is highlighted and future enhancement is provided.

2. RELATED WORK

This section studies various methodologies for detecting ataxia patient using gait characteristics. The purpose of [23] was to look at the feasibility of using gait to: i) identify people who have ataxia-related gait-characteristics (threat prediction) and ii) evaluate the degree of ataxia (severity-assessment). They gathered 155 films of 89 individuals, 24 individuals who had controlled the spinocerebellar-ataxia (SCAs), and 65 individuals with SCAs (or at risk for developing) doing the gait-task of the SARA at 11 clinical locations in eight states across the united states. In addition, they devised a strategy for isolating the subjects from their environment and built several features to record aspects of gait such as step-width, step-length, speed, swing, and stability. Their proposed method for predicting risk is 83.06% accurate and has an F1 score of 80.23%. Further, their proposed method for evaluating severity has an mean absolute error (MAE) of 0.6225 as well as a pearson's-correlation-coefficient-score (PCCS) of 0.7268, both statistically significant results. When tested on data from non-training sources, their proposed method maintained superior results. Moreover, the feature-importance evaluation revealed that their proposed method correlates higher ataxia-severity with broader steps, slower walking speed, and more instability. Zhang *et al*. [24], ataxia is a symptom that arises whenever the human body experiences problems with balance and coordination. Although there are a variety of possible internal causes of ataxia, the condition is often diagnosed based on external characteristics and the physician's own clinical experience. Here, they employ a contactless sensing method to identify cases of sensory-ataxia and separate them from cases of cerebellar-ataxia. Using a microwave sensing system, they initially gather data on romberg's tests as well as gait-analysis; then, following some preprocessing, then training of the method is done using machine-learning techniques. Since time series parameters are taken into account for romberg's test, all three methods achieve an accuracy of 96% or better in this task; for gait identification, principal-

component-analysis (PCA) is employed for reducing the dimension, as well as the accuracy rates of back-propagation neural-network, support-vector-machine (SVM), as well as random-forest (RF) were 97.8%, 98.9%, and 91.1%, correspondingly.

Having trouble walking is a common symptom of a wide range of debilitating neurologic and orthopedic conditions, as demonstrated in [25]. The use of accelerometers makes accurate simulation of gait patterns possible. However, they generate a large amount of data that might be difficult to analyze. In this study, they evaluated various techniques for clinical-data reduction as well as the categorization of the resulting data. From data collected on 43 subjects (20 healthy subjects, 23 ataxic subjects), resulting in 418 sequences of normal gait pattern, the maximum accuracy of 98% was attained using an RF classifier pre-processed by t-distributed-stochastic-neighbor segmentation to distinguish among healthy persons and individuals with ataxic-gait. Researchers have used motion sensor information to analyze how individuals with neurological disorders, like hereditary-ataxias (HA), walk over time [26]. By collecting data across 14 HA participants and 14 healthy individuals via iPhone motion sensors attached to their ankles, this study seeks to determine the minimum needed gait traits necessary for effective and less invasive HA patient recognition. To decrease the number of gait characteristics and sensor systems, two proposals were constructed: i) a local-minimum significant peak requirement to determine the beginning of every step, yielding a 10-stride frame from where 56 features were derived; as well as ii) a searching method depending mostly on hill climbing algorithm. The primary outcomes were that the k-nearest neighbor (KNN) and multi-layer perception (MLP) methods both obtained a 96% classification performance utilizing two gait sequences, as well as that for the MLP method just the right ankle sensor variations were necessary, hence reducing the intrusion.

Accelerometric data may help optimize DL convolution neural network (CNN) systems in [18], which might then be used to differentiate between normal and ataxic gait. There is a total of 860 signal segmentation from 16 ataxic subjects and 19 control subjects, with the average life expectancy of the two groups being 38.6 and 39.6 years, correspondingly, inside the experimental dataset. The technique relies on decomposing accelerometric data captured at multiple body locations at once and sampling at 60 hertz into their frequency components. All of the parameters between 0 and 30 hertz are utilized by the DL algorithm. SVM, Bayesian methodologies, as well as two-layer neural networks having characteristics evaluated as that of the relative-power in specified frequencies are among the conventional techniques whose results are contrasted with those achieved in this classification experiment. The outcomes demonstrate that picking the right spots for the sensors can boost accuracy from 81.2% at the foot to 91.7% at the spine. The accuracy was 95.8% achieved by integrating the input data using a five-layer DL algorithm. However, the model performs badly considering limited training sample considering imbalanced data; further, for performing multi-label classificaon require effective feature weight optimization and its importance for selecting during training; thus, the current approaches failed attain good accuracies. In addressing the research limitation, the following research methodology is presented.

3. PROPOSED METHOD

In this section a novel feature selection and ranking method is introduced for improving classification of ataxia severity prediction such as progressive ataxia (PA), chronic ataxia (CA), and healthy (H) participant. The architehture of proposed ataxia severity classification is provided in Figure 1. The classification task is said to be a multi-label classification problem. The ataxia severity classification dataset can be described as (1).

$$E = \{(a_1, b_1), (a_2, b_2), \dots, (a_m, b_m)\} \quad (1)$$

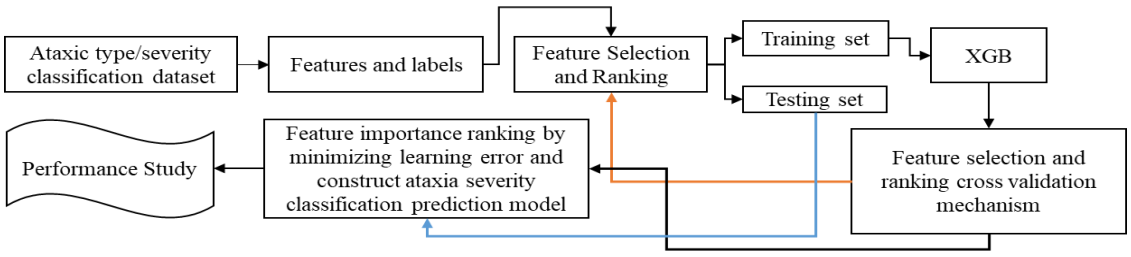

Figure 1. Proposed ataxia type/severity classification model using FSR-XGB

In (1) $j = 1,2,3, \dots, m$, which represent total participant size, $b_j \in \{-1,1\}$ defines j^{th} participant output, and a_j defines observer participant j's independent feature considering n −dimensional vector. The ataxia severity classification dataset is smaller in size. Therefore, building ataxia severity prediction model $\hat{G}$, using ML model is challenging. In this work, the ataxia severity classification prediction model $\hat{G}$ is used for predicting the actual class G considering multi-label classification using XGB as expressed in (2).

$$g: A \to B \tag{2}$$

3.1. XGBoost prediction algorithm

The XGB algorithm is very efficient in the field of classification task across different domain such as education, agriculture, and 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 in (3):

$$M^{(u)} = \sum_{j=1}^{o} m\left(z_j, A_j^{(u-1)} + g_u(y_j)\right) + \rho(g_u) + d \tag{3}$$

where the loss function is defined through parameter $m(.,.)$, the parameter ρ 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 (4):

$$g(y + \delta y) \approx g(x) + g'(y)\delta y + \frac{1}{2} g''(y)\delta y^2 \tag{4}$$

then substituting (4) into (3) we get:

$$M^{(u)} \approx \sum_{j=1}^{o} \left[m\left(z_j + A_j^{(u-1)}\right) + h_j g_u(y_j) + \frac{1}{2} i_j \left(g_u(y_j)\right)^2 \right] + \rho(g_u) + d \tag{5}$$

where h_j is computed as (6):

$$h_j = \frac{\partial M}{\partial a_j}, \tag{6}$$

and i_j is computed as (7):

$$i_j = \frac{\partial^2 M}{\partial a_j^2} \tag{7}$$

in (5) the constant term is removed for simplifying the computation at instance u as (8).

$$M^{(u)} \approx \sum_{j=1}^{o} \left[h_j g_u(y_j) + \frac{1}{2} i_j \left(g_u(y_j)\right)^2 \right] + \rho(g_u) \tag{8}$$

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:

$$M = -\sum_{j=1}^{o} \left[z_j \log\left(\hat{z}_j\right) + \left(1 - z_j\right) \log\left(1 - \hat{z}_j\right) \right] \tag{9}$$

in (9) the parameter $\hat{z}_j$ is computed as:

$$\hat{z}_j = \frac{1}{[1 + \exp(-a_j)]}, \tag{10}$$

and for activation sigmoid function is used and we obtain (11).

$$\frac{\partial \hat{z}_j}{\partial a_j} = \hat{z}_j\left(1 - \hat{z}_j\right) \tag{11}$$

3.2. Feature selection and ranking under multi-label imbalanced data

In traditional ML model for feature selection uses iterative cross-validation as defined in (12).

$$CV(\sigma) = \frac{1}{M}\sum_{k=1}^{K}\sum_{j\in G_{-k}} P\left(b_j, \hat{g}_{\sigma}^{-k(j)}(y_j, \sigma)\right) \quad (12)$$

Using (12), the ataxia severity dataset is segmented randomly into K-subset of equal size. Then ataxia severity classification predictive model is constructed using reaming $K-1$ subset. Finally, by optimizing feature that minimized the validation error an taking average value of grid l as defined in (13).

$$\hat{\sigma} = \underset{\sigma\in\{\sigma_1,\ldots,\sigma_l\}}{\arg\min}\ CV_s(\sigma) \quad (13)$$

The above cross-validation model fails to establish feature correlation between different attributes and participant. The proposed model models a new cross-validation mechanism for better feature selection and ranking considering multilabel ataxia dataset. In proposed cross-validation encompasses two-level feature optimization. The level 1 considers the ideal features are feature subsets; in level 2, the ataxia severity classification prediction model is constructed using ideal feature subset obtained through level 1. The proposed two-level cross-validation is defined as:

$$CV(\sigma) = \frac{1}{SM}\sum_{s=1}^{S}\sum_{k=1}^{K}\sum_{j\in G_{-k}} P\left(b_j, \hat{g}_{\sigma}^{-k(j)}(y_j, \sigma)\right) \quad (14)$$

where, $P(\cdot)$ defines loss function, $\hat{g}_{\sigma}^{-k(j)}(\cdot)$ defines coefficient computation function, and M are total size of ataxia severity classification dataset. The optimization process in choosing ideal $\hat{\sigma}$ is computed using (13). The process of feature selection and ranking for attaining better ataxia severity classification is given in:

Step 1. The ataxia severity classification dataset E is randomly segmented into K-folds of equal size. Let assume $k = 1$ to K we define parameter E^{-k}, where k^{th} ataxia sample data rows are removed for outer-level training sets and E^{k} with remaining ataxia dataset k^{th} segments are utilized for validating outer-level testing sets. Using pre-configured S, the below steps are repeated. Then, the ataxia severity classification dataset E^{-k} is divided random manner into H-folds of equal size, $\forall\ h = 1$ to H. A). Conclude H diverse ataxia severity classification dataset E^{-kh} with h^{th} part removed for inner-level training datasets and E^{kh} with h^{th} portion are utilized during inner-level testing dataset. $\forall\ l = 1$ defining size considered for grid formation during feature selection and ranking process. Ataxia severity classification prediction model $\hat{g}_{\sigma_l}$ construction through as shown in:

$$\hat{g}_{\sigma_l} = \hat{g}\left(b_j, \hat{g}(E^{-kh}; \sigma_l)\right). \quad (15)$$

then, using loss function estimate error using inner-level testing data through optimization of $\hat{g}_{\sigma_l}$ on inner-level testing data E^{kh} through as shown in (16).

$$\mathcal{E}_{\sigma_n} = \sum_{j\in E^{-kh}} P\left(b_j, \hat{g}(E^{-kh}; \sigma_l)\right) \quad (16)$$

For every l compute H-folds cross-validation errors; thus, each row of M_h in layer 1 for k^{th} part will have different cross-validation as (17).

$$CV(\hat{g}; \sigma_l) = \frac{1}{M_h}\sum_{h=1}^{H}\sum_{j\in E^{-kh}} P\left(b_j, \hat{g}(E^{-kh}; \sigma_l)\right) \quad (17)$$

In iterative manner considering S-times for different l the cross-validation error is computed for M_h in layer 1 for k^{th} part using:

$$CV_S(\hat{g}; \sigma_l) = \frac{1}{M_h S}\sum_{s=1}^{S}\sum_{h=1}^{H}\sum_{j\in E^{-kh}} P\left(b_j, \hat{g}(E^{-kh}; \sigma_l)\right) \quad (18)$$

the model obtain a more realistic optimized feature with different combination of l as:

$$\hat{\sigma}_n = \underset{\sigma\in\{\sigma_1,\sigma_l\}}{\arg\min}\ CV_S(\hat{g}; \sigma_l) \quad (19)$$

then, using gradient decent function the realistic optimization feature is selected through minimization goal. The final feature subset to build ataxia severity classification model is constructed employing feature ranking function $r(\cdot)$ using:

$$r(a) = \begin{cases} 0 \; if \; n_j \; is \; not \; chosen \\ 1 \; if \; n_j \; is \; chosen \; as \; final \; predictive \; model \; j = 1,2,3, \dots, n \end{cases} \quad (20)$$

the total feature subset using above ranking function is established as:

$$F_s = \{r(n_1), r(n_1), \dots, r(n_n)\}, \quad (21)$$

to provide effective accuracy, only feature that has higher rank/weights (i.e., higher importance) considering different K-folds are selected through (22).

$$F_{s_k} = \{r(n_1), r(n_1), \dots, r(n_n)\} \quad (22)$$

Step 2. The work further computes number of instances a particular feature is selected within K feature subsets with higher weights based on which the final good quality feature subset are selected as defined:

$$F_{s_{final}=\{f_s(p_1), f_s(n_2), \dots, f_s(n_n)\}}, \quad (23)$$

where $f_s(\cdot)$ defines if n^{th} feature is selected or not; the process is given in (24).

$$F_s(a) = \begin{cases} 0 \; if \; q_j \; is \; chosen \; lesser \; than \; \frac{K}{2} \; times, j = 1,2,3, \dots, n \\ 1 \; if \; q_j \; is \; chosen \; greater \; or \; equal \; to \; \frac{K}{2} \; times, j = 1,2,3, \dots, n \end{cases} \quad (24)$$

Using (24), the model establishes subset of n' selected features, where n^{th} represent total feature selected considering providing feature importance for building ataxia severity classification prediction model. In the next section, the process involved in multi-label classifier construction for ataxia type classification.

3.3. Multi-label classifier cosntruction

The steps involved in ataxia severity classification prediction model are as follows: Step 1. In this work ataxia severity classification dataset E is reduced to E' by retaining feature selected in stage 1 as defined:

$$E' = (E; n') \quad (25)$$

Step 2. In similar manner with respect to stage 1, the K-folds is used $\forall \; k = 1$ to K. 1). Provide detail information of ataxia severity classification dataset ${E'}^{(-k)}$ with k^{th} part removed for performing the training process and ${E'}^{(k)}$ part remaining are utilized for testing purpose. Next, in iterative manner considering configured S, the below steps are executed $\forall \; l = 1$ to L, where L represent the grid size used during optimization process. Ataxia severity classification prediction model construction through (26);

$$\hat{g}_{\sigma_l} = \hat{g}\left({E'}^{(-k)}; \sigma_l\right) \quad (26)$$

then, for inner-level testing data ${E'}^{(-k)}$ applies $\hat{g}_{\sigma_l}$ and compute error for different l using (27);

$$\mathcal{E}_{\sigma_l} = P\left(b_j, \hat{g}\left({E'}^{(-k)}; \sigma_l\right)\right) \quad (27)$$

then, for different L compute the K-fold cross-validation error for optimization process using (28):

$$CV(\hat{g}; \sigma_l) = \frac{1}{M}\sum_{k=1}^{K}\sum_{j \in {E'}^{(-k)}} P\left(b_j, \hat{g}\left({E'}^{(-k)}; \sigma_l\right)\right) \quad (28)$$

In iterative manner the cross-validation error is computed as (29).

$$CV_S(\hat{g};\sigma_l) = \frac{1}{KS}\sum_{s=1}^{S}\sum_{k=1}^{K}\sum_{j\in E'^{(-k)}} P\left(b_j, \hat{g}\left(E'^{(-k)};\sigma_l\right)\right) \quad (29)$$

Step 3. Then, considering different l the realistic optimizations parameter is obtained as (30).6

$$\hat{\sigma} = \underset{\sigma\in\{\sigma_1,\ldots,\sigma_l\}}{\arg\min}\ CV_S(\hat{g};\sigma_l) \quad (30)$$

Step 4. Using gradient decent the realistic value is optimized through minimization function. In reducing randomness of proposed ataxia severity classificaon predictive model in level-1, H-folds are built and iterated S times; further, the layer-2 only utilizes feature subset for ataxia severity classification prediction model; thus, the FSR-XGB attain better performance than existing ML-based ataxia severity classification.

4. RESULT AND ANALYSIS

This section studies the performance achieved using proposed FSR-XGB based ataxic severity classification model over standard XGB-based classification model. Further, the model is compared with existing ML [25] and DL-based [14], [18] ataxia severity classification model. The proposed and other existing ataxic severity classification model is implemented using Anaconda Python framework. The experiment is conducted using dataset collected from [27], [28]. The experiment is conducted using ataxia severity classification dataset [29] collected from [28] which has three classes such as healthy, PA, and CA; thus, classificaon task becomes a multi-label classificaon problem. The gait characteristic dataset is generated using kinect v2, tripod, microsoft kinect software devolopemnt kit and more detail of dataset can be obtained from [28]. The accuracies, sensitivity, specificity, precision, and F-measure are metrics used for validating the classification algorithm performance.

4.1. Specificity vs sensitivity measure

This section provides performance study of specificity and sensitivity metrics in classifying whether a person is ataxic or healthy. The proposed FSR-XGB is compared with a total of five existing ML algorithm [25] such as random forest (RF), ADABoost, radial basis function (RBF) support vector machine (RBF-SVM), linear regression (LR), and XGB and one DL algorithm [14], [18]. The graphical outcome of specificity and sensitivity is given in Figure 2. From Figure 2 we can interpret that FSR-XGB outperforms all existing ML and DL-based ataxic person identification methods.

4.2. Accuracy performance

This section provides performance study of accuracies metrics in classifying whether a person is ataxic or healthy. The proposed FSR-XGB is compared with a total of eight existing ML algorithms [25] such as RF, ADABoost, RBF-SVM, LR, XGB, Bayes, SVM, two layer neural network (TLNN), and one DL neural network algorithm [14], [18]. The graphical outcome of accuracy is given in Figure 3. From Figure 3 we can interpret that FSR-XGB outperforms all existing ML and DL-based ataxic person identification methods.

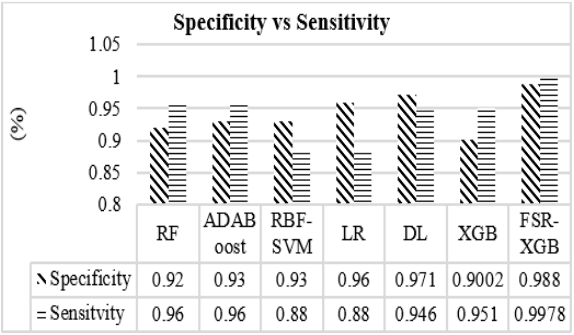

Figure 2. Specificity vs sensitivity performance of different ML and DL algorithms

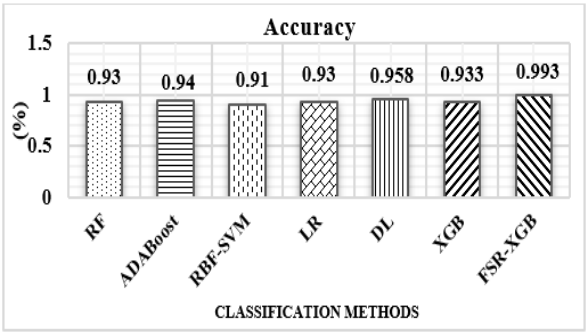

Figure 3. Accuracy performance of different ML and DL algorithms

4.3. Classification performance and feature importance analysis

This section provides performance analysis of feature importance i.e., selection and ranking of both XGB and FSR-XGB and their role in multi-label classification outcomes. Figure 4 shows the graphical study of feature selection and ranking and their importance score of both XGB and FSR-XGB. The figure shows the proposed FS-XGB provides a higher feature importance score in comparison with XGB. Figure 5 shows graphical representation of classification performance of XGB and FSR-XGB. From Figures 4 and 5 we can interpret that FSR-XGB outperform standard XGB-based ataxic person identification methods.

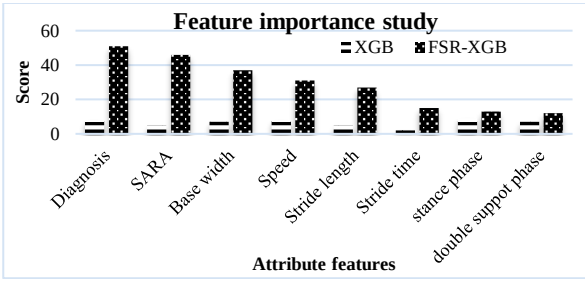

Figure 4. Feature importance analysis of XGB and FSR-XGB

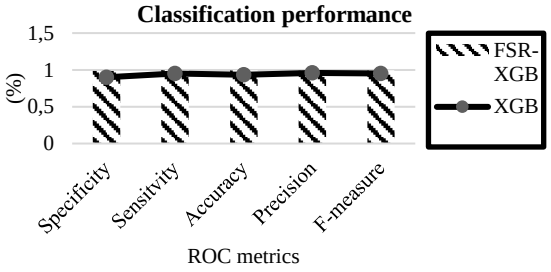

Figure 5. Classification performance for different ROC metrics

5. CONCLUSION

This research aims to use ML to identify ataxic individuals and compare them to control groups by analyzing gait characteristics. This paper conducted an in-depth analysis of the advantages and disadvantages of several ML and DL approaches used for ataxic-person-identification based on gait characteristics. The research indicates that more extensive data sets are needed for the DL approach to obtain optimal accuracy. However, when data is imbalanced, ML-based algorithms produce inaccurate results due to an increase in false positives. This paper introduces a novel feature weight optimization method for XGB to decrease false positives, and it also introduces a novel feature selection process by changing the cross-validation mechanism that is effective even with data that is imbalanced in its true form. In comparison to traditional XGB-based ataxic person identification, the FSR-XGB achieves higher levels of accuracy, sensitivity, specificity, precision, and F-measure performance by efficiently identifying which features affect classification accuracies by decreasing objective error. In future additional classes of ataxia will be considered and perform classificaon; Further, other nondegenerative like Parkinson will be considered and perform classification and validate the performance of FSR-XGB.

REFERENCES

[1] M. I. M. Razi, M. R. C. Beson, S. N. Azemi, and S. A. Aljunid, "FBG sensor strain performance analysis using optisystem software tools," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 14, no. 2, pp. 564–572, May 2019, doi: 10.11591/ijeecs.v14.i2.pp564-572.

[2] O. I. Alsaif, S. Q. Hasan, and A. H. Maray, "Using skeleton model to recognize human gait gender," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 2, pp. 974–983, Jun. 2023, doi: 10.11591/ijai.v12.i2.pp974-983.

[3] R. S. Tanna and C. H. Vithalani, "Classification of lower limb rehabilitation exercises with multiple and individual inertial measurement units," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 28, no. 2, pp. 840–849, Nov. 2022, doi: 10.11591/ijeecs.v28.i2.pp840-849.

[4] O. Dostál, A. Procházka, O. Vyšata, O. Ťupa, P. Cejnar, and M. Vališ, "Recognition of motion patterns using accelerometers for ataxic gait assessment," *Neural Computing and Applications*, vol. 33, no. 7, pp. 2207–2215, Apr. 2021, doi: 10.1007/s00521-020-05103-2.

[5] A. Gouelle, S. Norman, B. Sharot, S. Salabarria, S. Subramony, and M. Corti, "Gauging gait disorders with a method inspired by motor control theories: a pilot study in friedreich's ataxia," *Sensors (Switzerland)*, vol. 21, no. 4, pp. 1–18, Feb. 2021, doi: 10.3390/s21041144.

[6] E. J. Harris, I. H. Khoo, and E. Demircan, "A survey of human gait-based artificial intelligence applications," *Frontiers in Robotics and AI*, vol. 8, Jan. 2022, doi: 10.3389/frobt.2021.749274.

[7] Y. Guo, J. Yang, Y. Liu, X. Chen, and G. Z. Yang, "Detection and assessment of Parkinson's disease based on gait analysis: a survey," *Frontiers in Aging Neuroscience*, vol. 14, Aug. 2022, doi: 10.3389/fnagi.2022.916971.

[8] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: a survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, Jul. 2019, doi: 10.1016/j.patrec.2018.02.010.

[9] H. A. A. A. Al-Khamees, N. Al-A'araji, and E. S. Al-Shamery, "Enhancing the stability of the deep neural network using a non-constant learning rate for data stream," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 2, pp. 2123–2130, Apr. 2023, doi: 10.11591/ijece.v13i2.pp2123-2130.

[10] Y. Ma, K. Mithraratne, N. Wilson, Y. Zhang, and X. Wang, "Kinect v2-based gait analysis for children with cerebral palsy: validity and reliability of spatial margin of stability and spatiotemporal variables," *Sensors*, vol. 21, no. 6, pp. 1–17, Mar. 2021, doi: 10.3390/s21062104.
[11] V. Cimolin *et al.*, "Computation of gait parameters in post stroke and Parkinson's disease: a comparative study using RGB-D sensors and optoelectronic systems," *Sensors*, vol. 22, no. 3, p. 824, Jan. 2022, doi: 10.3390/s22030824.
[12] C. A. A. Izhar, Z. Hussain, M. I. F. Maruzuki, M. S. Sulaiman, and A. A. A. Rahim, "Gait cycle prediction model based on gait kinematic using machine learning technique for assistive rehabilitation device," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, pp. 752–763, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp752-763.
[13] F. Nabizadeh *et al.*, "Artificial intelligence in the diagnosis of multiple sclerosis: a systematic review," *Multiple Sclerosis and Related Disorders*, vol. 59, p. 103673, Mar. 2022, doi: 10.1016/j.msard.2022.103673.
[14] M. Kaya, S. Karakuş, and S. A. Tuncer, "Detection of ataxia with hybrid convolutional neural network using static plantar pressure distribution model in patients with multiple sclerosis," *Computer Methods and Programs in Biomedicine*, vol. 214, p. 106525, Feb. 2022, doi: 10.1016/j.cmpb.2021.106525.
[15] B. Chen *et al.*, "Computer vision and machine learning-based gait pattern recognition for flat fall prediction," *Sensors*, vol. 22, no. 20, p. 7960, Oct. 2022, doi: 10.3390/s22207960.
[16] A. Li and C. Li, "Detecting Parkinson's disease through gait measures using machine learning," *Diagnostics*, vol. 12, no. 10, p. 2404, Oct. 2022, doi: 10.3390/diagnostics12102404.
[17] T. A. Zesiewicz *et al.*, "Longitudinal gait and balance decline in Friedreich's ataxia: a pilot study," *Gait and Posture*, vol. 55, pp. 25–30, Jun. 2017, doi: 10.1016/j.gaitpost.2017.03.019.
[18] A. Prochazka *et al.*, "Deep learning for accelerometric data assessment and ataxic gait monitoring," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 360–367, 2021, doi: 10.1109/TNSRE.2021.3051093.
[19] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P. A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, Jul. 2019, doi: 10.1007/s10618-019-00619-1.
[20] N. Golestani and M. Moghaddam, "Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks," *Nature Communications*, vol. 11, no. 1, p. 1551, Mar. 2020, doi: 10.1038/s41467-020-15086-2.
[21] L. Sadouk and T. Gadi, "Convolutional neural networks for human activity recognition in time and frequency-domain," in *Advances in Intelligent Systems and Computing*, vol. 756, 2019, pp. 485–496.
[22] T. Aşuroğlu and H. Oğul, "A deep learning approach for parkinson's disease severity assessment," *Health and Technology*, vol. 12, no. 5, pp. 943–953, Sep. 2022, doi: 10.1007/s12553-022-00698-z.
[23] W. Rahman *et al.*, "Auto-gait," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Mar. 2022, vol. 7, no. 1, pp. 1–19, doi: 10.1145/3580845.
[24] Q. Zhang, X. Zhou, Y. Li, X. Yang, and Q. H. Abbasi, "Clinical recognition of sensory ataxia and cerebellar ataxia," *Frontiers in Human Neuroscience*, vol. 15, Apr. 2021, doi: 10.3389/fnhum.2021.639871.
[25] O. Vyšata *et al.*, "Classification of ataxic gait," *Sensors*, vol. 21, no. 16, p. 5576, Aug. 2021, doi: 10.3390/s21165576.
[26] F. D. Acosta-escalante, E. Beltrán-naturi, and M. Boll, "Gait pattern recognition in patients with hereditary ataxia's using supervised learning algorithms based on small subsets smartphone sensor data," *Preprints*, no. December, pp. 1–18, 2020, doi: 10.20944/preprints202012.0054.v1.
[27] M. Serrao *et al.*, "Dataset on gait patterns in degenerative neurological diseases," *Data in Brief*, vol. 16, pp. 806–816, Feb. 2018, doi: 10.1016/j.dib.2017.12.022.
[28] S. Summa *et al.*, "Spatio-temporal parameters of ataxia gait dataset obtained with the Kinect," *Data in Brief*, vol. 32, p. 106307, Oct. 2020, doi: 10.1016/j.dib.2020.106307.
[29] A. S. Nunes, N. Kozhemiako, C. D. Stephen, J. D. Schmahmann, S. Khan, and A. S. Gupta, "Automatic classification and severity estimation of ataxia from finger tapping videos," *Frontiers in Neurology*, vol. 12, p. 795258, Feb. 2022, doi: 10.3389/fneur.2021.795258.

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 and researchj23@gmail.com.

Dr. 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, 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.