RESEARCH ARTICLE OPEN ACCESS

# An Optimized Feature Selection Technique Based on Feature Analysis for Bio-metric Gait Data Classification

KM Nivedita Singh
*M.Tech Scholar*
*Computer Science & Engineering*
*Eshan College of Engineering,*
*Mathura, India*
niveditas336@gmail.com

Vyom Kulshreshtha
*Associate Professor*
*Computer Science & Engineering*
*Eshan College of Engineering,*
*Mathura, India*
Vyom19@gmail.com

***Abstract—***

For bipedal robots to be stable and successful—especially in human terrain or external environments—optimized feature selection is essential. This work uses statistical techniques to extract important characteristics from bidirectional gait data, which are necessary to accurately simulate human gait patterns. Personalized health evaluation is made possible by the ability to classify irregularities based on a variety of gait data that indicate unique health status.

But feature selection is hard, especially when dealing with high-dimensional feature vectors, because handmade methods are insufficient and result in computational inefficiencies. A methodology that first selects features and identifies principal features is presented as a solution to this problem. The best feature set derived from Analysis of Variance (ANOVA) is then utilized by machine learning techniques for data classification.

In principal features verification, seventeen features are extracted, which serve as the foundation for the optimized feature set. Three-fold cross-validation is used to assess the model's effectiveness, guaranteeing its resilience and generalizability.

This technology can be used to provide bipedal robots with features that enable stable movement and precise human gait imitation. Furthermore, applications in healthcare and rehabilitation may benefit from the capacity to categorize gait data according to health-related factors.

**Keywords**— Bipedal locomotion, Gait cycle, Biometric identification

## I. INTRODUCTION

The integration of machine learning and bipedal robotics is a noteworthy development in computer science that seeks to reduce the disparity in intelligence and behavior between humans and machines. Bipedal robots are those that can walk on two legs and attempt to mimic the complex gait cycles of humans. This mimicry makes it easier for them to blend in with human-made surroundings and allows them to work in dangerous, boring, or unhygienic settings—essentially acting as a stand-in for human labor and maybe saving valuable human lives from dangerous circumstances [2].

The intrinsic nonlinearity of human gait patterns makes it extremely difficult to comprehend and use. Understanding and evaluating the nonlinear characteristics included in human gait patterns—especially when it comes to classification—requires the application of machine learning techniques. In this thesis, we analyze different abnormalities-related departures from the typical human gait cycle and compare them to the typical human gait pattern. In doing so, we want to identify markers of a person's health status and advance the field of individualized health monitoring and healthcare diagnostics.

### A. *Motivation*

Studying human gait behavior and incorporating it into bipedal robotics has enormous potential applications in a wide range of industries, far beyond the scope of conventional research domains. Here are a few unique uses:

**1. Enhancing Prosthetic Limbs:** By creating prosthetic limbs that replicate normal gait, scientists are able to better serve amputees by enhancing their quality of life.

**2. Examining Dangerous Environments:** Inspired by human gait, bipedal robots are excellent in investigating dangerous situations that are off-limits to people, such as nuclear reactors and bomb disposal sites.

**3. Reducing Uncertainty:** Examining human walking patterns aids in locating unstable areas in bipedal robots, improving their maneu verability and steadiness over irregular ground.

**4. Healthcare Diagnostics:** Disturbances from typical gait patterns provide information on early illness identification and prognosis, enabling prompt treatment and intervention.

**5. Revolutionizing Entertainment and Agriculture:** Bipedal robots' humanoid structure makes them perfect for compelling performances in entertainment and precision jobs in agriculture, all without causing crop damage or decreasing audience interest.

## II. REVIEW OF LITERATURE

Utilizing deep learning, Semwal et al. provide a novel approach to feature selection in humanoid push recovery and categorization. Their research looks on push recovery processes in different subjects while taking things like open and closed eyes into account. They use joint angle information obtained by empirical mode decomposition to classify pushes into various intensities. Their approach is more credible due to validation using 5-fold cross-validation and ANOVA, which also improves understanding of

humanoid push recovery and classification.

  

Wang et al. investigate the use of several machine learning techniques in bipedal robot control, including supervised, unsupervised, and reinforcement learning. Their research sheds light on how these algorithms have developed and how they might be used to improve bipedal robot adaptive control techniques. Through an examination of cutting-edge methods, they illuminated the effectiveness, consequences, and constraints of every strategy.

Matthew D. Zeiler introduced ADADELTA, an adaptable learning rate technique for gradient descent, in his article. This approach eliminates the need for manual intervention in determining the learning rate. It computes the necessary statistics with approximations over time using just first order information. It seems resilient to noisy gradient descent, diverse architectures, a range of data models, and hyper-parameter selection. The MNIST dataset was tested on a single machine, while the voice data set was tested in a distributed clustered environment. The outcomes were then compared to those obtained using alternative techniques.

Hybrid automata are a type of language tool for real-time systems that were developed by S. Shankar etal. [5]. They are capable of modeling and analyzing both digital and analog input. In order to provide a stability criterion, their study addresses dynamical systems and outlines the necessary requirements for the existence of a unique solution. Two basic theorems are extended in this work: LaSalle's invariance principle and Lyapunov's theorem on stability through linearization.

In order to assess Complex Regional Pain Syndrome (CRPS) treatment, Yang et al. looked at the gait cycle [6]. Limb pain and a decline in function are brought on by CRPS. Over brief walking distances, they used accelerometer-recorded gait data from ten CRPS patients and ten controls. They used a Multilayer Perceptron (MLP) neural network to extract thirty-three features from each recording and classified abnormal and normal gait patterns, obtaining an 85.7% prediction accuracy in testing.

In their study 'Biologically-inspired Push Recovery Capable Bipedal Locomotion', Semwal, Katiyar, and colleagues presented a novel eight-phase hybrid automata model for the human gait cycle [7]. This method allows for a more thorough study because it takes into consideration all 8 stages of human walking, in contrast to conventional 2-phase models. Their research concentrated on the dynamic factors necessary for strong push recovery mechanisms and steady bipedal mobility. The process of validation comprised utilizing their exclusive controller and OpenSim data to compare different gait locomotion stages. Furthermore, they validated the hybrid automata model using actual human gait dynamics, which represents a noteworthy breakthrough in the modeling of bipedal locomotion.

## III. METHODOLOGY

The proposed work is based on following methodolgy.

### A. *Acquring the dataset*

The National Institutes of Health (NIH) has compiled gait data from the Sim Toolkit website and Open Sim models into the "Robita Gait Data" collection. There are four classes in it: Normal, MO, SE, and MI. The Normal class depicts healthy people, whereas the first three classifications indicate those with varied gait impairments. There are three individuals each class, for a total of 12 individuals in the dataset.

One gait cycle is collected for every individual, yielding 100 rows of data for every individual. There are eighteen columns in each row; the class designation appears in the last column, and the preceding seventeen columns show joint angle features. As a result, there are 1200 rows in the dataset overall, with 17 features each row.

In particular, the crouch dataset in the MI, MO, and SE classes characterizes people who have crouching gait patterns, most likely as a result of motor control problems, sensory impairments, or muscle imbalance, respectively.

Table 1: Information About Dataset

| For each subject their age, height, weight, speed and Min KFA is given.Subject | Age (years) | Height (cm) | Weight (kg) | Min KFA (deg) | Speed (m/s) |
|---|---|---|---|---|---|
| MI | 9.4 | 131.0 | 28.2 | 15.5 | 0.94 |
| MI | 7.9 | 112.5 | 21.3 | 17.7 | 0.88 |
| MI | 9.1 | 127.6 | 23.1 | 21.1 | 0.67 |
| MO | 12.4 | 133.5 | 78.5 | 35.7 | 0.7 |
| MO | 8.7 | 131.0 | 21.1 | 33.1 | 0.9 |
| MO | 11.0 | 143.0 | 28.7 | 32.6 | 1.2 |
| SE | 12.2 | 167.0 | 37.9 | 46,5 | 1.2 |
| SE | 13.2 | 144.0 | 35.9 | 60.3 | 0.8 |
| SE | 16.3 | 160.0 | 50.8 | 86.0 | 0.7 |

### B. *Graphical representation of joint angle of all classes*

- Class MI

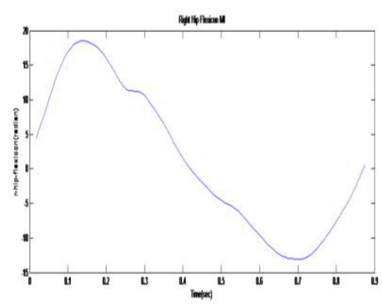

Figure 2.1: Hip joint angle

  

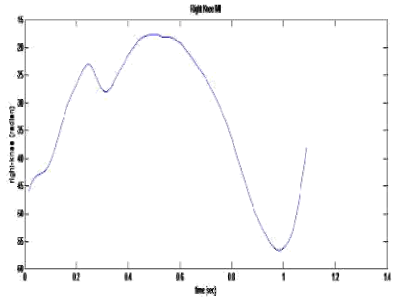

Figure 2.2: Knee joint angle

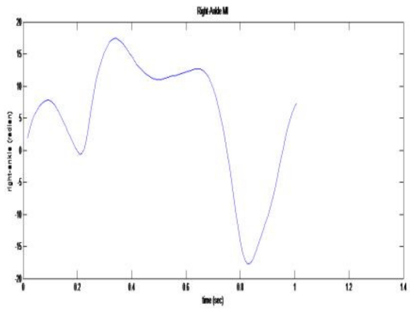

Figure 2.3: Ankle joint angle

- Class MO

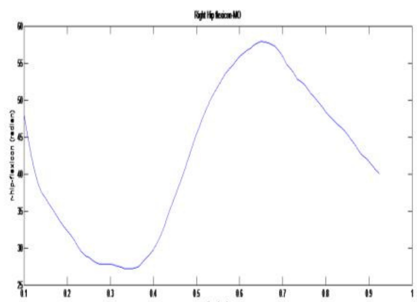

Figure 2.4: Hip joint angle

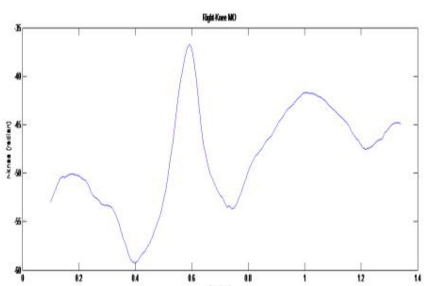

Figure 2.5: Knee joint angle

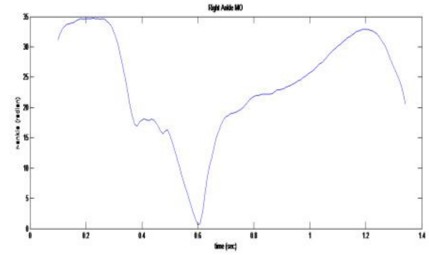

Figure 2.6: Ankle joint angle

- Class SE

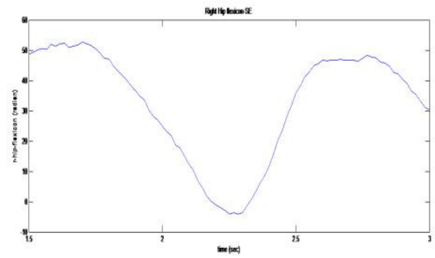

Figure 2.7: Hip joint angle

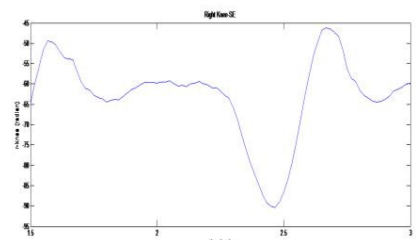

Figure 2.8: Knee joint angle

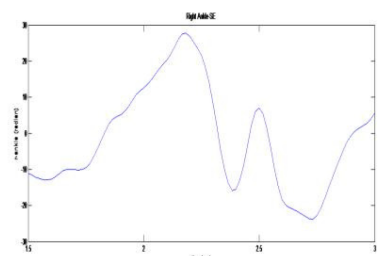

Figure 2.9: Ankle joint angle

- Class Normal

  

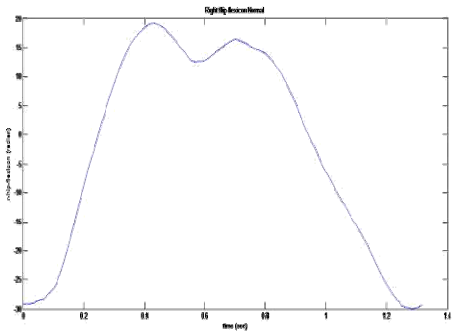

Figure 2.10: Hip joint angle

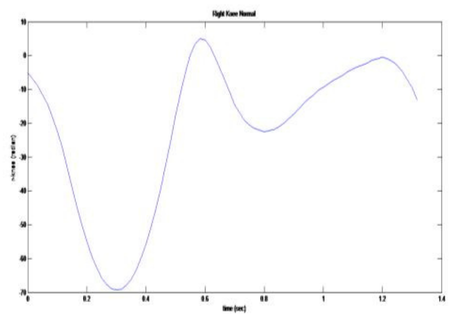

Figure 2.11: Knee joint angle

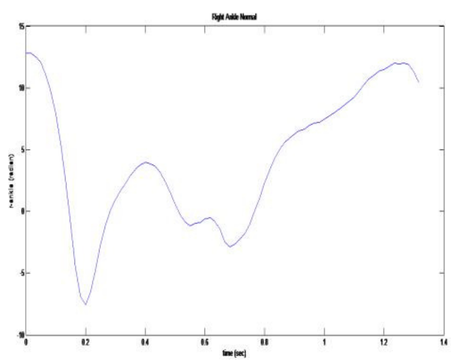

Figure 2.12: Ankle joint angle

### *C. Selection of features*

After preparing our database we have analyzed the feature of our data sets to get the actual feature that have greater effect on our datasets. So we apply data analysis technique of statistics called ANOVA (Analysis of Variance) to get how important each feature is. The method calculate the probability and F-value for each feature for all category or classes that is MI, MO, SE and Normal and based on that we have selected our features.

### *D. ANOVA (Analysis of Variance)*

Analysis of Variance is a technique of statistical analysis of the data. It is a collection of various methods of statistics that are used to analyze the difference of means among group and their associated methods. Analysis of variance Ronald Fisher, a renowned statistician and biologist, created (ANOVA). ANOVA generalizes the t-test for more than two groups by offering a statistical test to determine if the means of several groups are equal or not. The t-test is only helpful when there are only two groups. ANOVA is helpful for comparing and testing the statistical significance of three or more means.

## IV. PROPOSED MODEL

Learning a new programming language involves two components: comprehension of previously written code and code writing ability. Choosing the right programming language is inherently a multi-criteria decision-making process since it requires learning the fundamentals of programming at a level where one can both comprehend and write code for new problems. One of the better methods for handling uncertainty information in an MCDM

problem is fuzzy logic.

The right programming language will be chosen using a recurrent deep learning neural network model. Every programming language will be evaluated by the model using the MCDM parameters. The RNN will get the weights assigned to each criterion via MCDM techniques. We'll use machine learning algorithms.

The supervised learning needs to be completed in the subsequent step: Choose the right kind of training example first. That is, before moving on to another task, the user needs to know what kind of data to utilize as a training set for the specific issue at hand.

The training example and its matching label or output are then gathered once the type of data has been determined in order to train our system or model.

figuring out how many features to include in our training set so that our machine model is trained. The feature that best captures the input-output relationship of our challenge should be carefully selected. A huge number of features is not recommended due to the dimensionality curse.

figuring out the learning framework and the associated learning methods.

Run our learning algorithm on our training set or data set to discover the relationships and structures among the data pertaining to each class. The user may initialize or provide a parameter in this.

Next, we assess how well our learnt function performed using the data from our test set and compute its accuracy. We accept our model if the accuracy is sufficient. Cross-validation in k steps:

Rotation estimation, another name for cross-validation, is a process that verifies a model's generalization to separate datasets and its overall performance. It provides insight into how well a model will perform in real-world circumstances and is typically employed in classification jobs.

The original dataset is split into k equal-sized subsamples for k-fold cross-validation. K-1 subsamples are used as training data and the remaining subsample is used for validation in each cycle. To make sure that every subsample is used as validation data at least once, this process is repeated k times. One may then combine the k outcomes of the k-fold cross-validation to get a single estimate of the model's performance.

The process for a single k-fold cross validation run is as follows:

1. Put the training example in a randomly selected order.
2. Make k folds in the training example.(Roughly one sample per m/k)
3. From i = 1 to k:
• Use all of the examples that do not belong to I to train the classifier.
• Determine the number of cases in fold I that were incorrectly classified (ni). • Test the classifier on all of the examples in i.
4. Give the classifier error back the following estimate:

$$E = \frac{\sum_{i=1}^{k} n_i}{m}$$

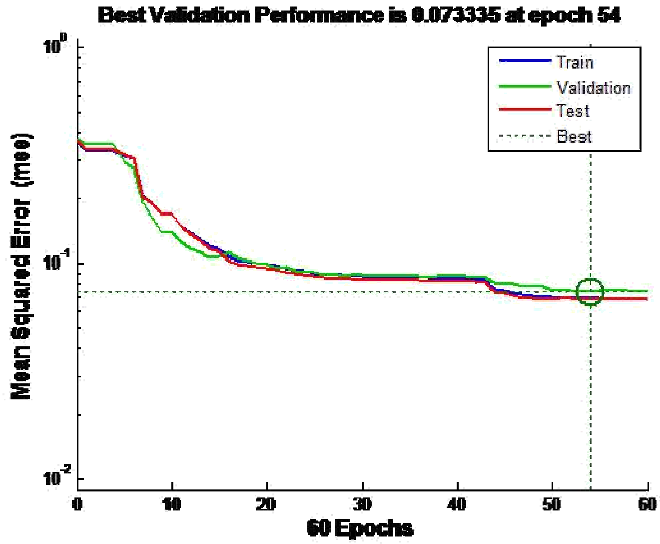


Figure 13: Performance

## V. RESULT ANALYSIS

The training, validation, and test data-set confusion matrices are shown in the accompanying figure. It provides 85.8% classification accuracy

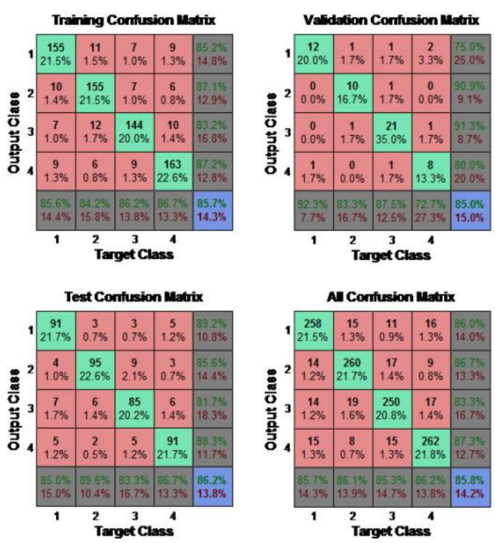


Figure 12: Confusion Matrix for ANN

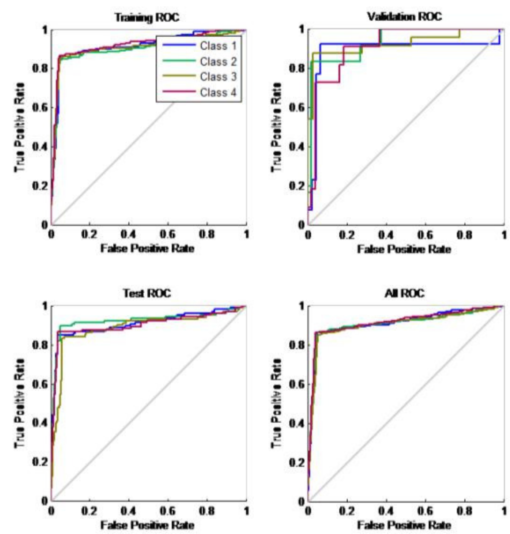


Figure 14: ROC curve

## VI. CONCLUSION

A gait detection system for bipedal humans has been developed, taking into account 12 different gaits. Ten features in all were chosen for classification from the available data set. This model is used to categorize gait data according to an individual's health; that is, if an individual has a deformity in their gait pattern or is ill, our model will classify them into one of three crouch categories: MI,

MO, or SE; if not, our model will identify them as normal.

Since we used relatively small data sets to generate the classification model, additional individual data was needed to build the model on a larger scale. Therefore, further gait data from four distinct categories is needed, and our model's accuracy could potentially increase. It would be beneficial to find more features or combinations of features to improve the accuracy of our model for classifying human gaits.

  

➢ References__

[1] Vijay Bhaskar Semwal, Kaushik Mondal, and GC Nandi. Robust and accurate feature selection for humanoid push recovery and classification: deep learning approach. *Neural Computing and Applications*, pages 1–10, 2015.

[2] V. B. Semwal, A. Bhushan, and G. C. Nandi. Study of humanoid push recovery based on experiments. In *Control, Automation, Robotics and Embedded Systems (CARE), 2013 International Conference on*, pages 1–6, Dec 2013. doi: 10.1109/ CARE.2013.6733741.

[3] Shouyi Wang, Wanpracha Chaovalitwongse, and Robert Babuska. Machine learning algorithms in bipedal robot control. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(5):728–743, 2012.

[4] Matthew D Zeiler. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

[5] J. Lygeros, K. H. Johansson, S. N. Simic, Jun Zhang, and S. S. Sastry. Dynamical properties of hybrid automata. *IEEE Transactions on Automatic Control*, 48(1): 2–17, Jan 2003. ISSN 0018-9286. doi: 10.1109/TAC.2002.806650.

[6] Mingjing Yang, Huiru Zheng, Haiying Wang, Sally McClean, Jane Hall, and Nigel Harris. A machine learning approach to assessing gait patterns for complex regional pain syndrome. *Medical engineering & physics*, 34(6):740–746, 2012.

[7] Vijay Bhaskar Semwal, Shiv A Katiyar, Rupak Chakraborty, and GC Nandi. Biologically-inspired push recovery capable bipedal locomotion modeling through hybrid automata. *Robotics and Autonomous Systems*, 70:181–190, 2015.

[8] De Zhang, Yunhong Wang, Zhaoxiang Zhang, and Maodi Hu. Estimation of view angles for gait using a robust regression method. *Multimedia Tools and Applications*, 65(3):419–439, 2013. ISSN 1573-7721. doi: 10.1007/s11042-012-1045-9. URL http: //dx.doi.org/10.1007/s11042-012-1045-9.

[9] Aaron D Ames. Human-inspired control of bipedal walking robots. *IEEE Transactions on Automatic Control*, 59(5):1115–1130, 2014.

[10] David Cunado, Mark S Nixon, and John N Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 93–102. Springer, 1997.

[11] William EH Harcourt-Smith. 5 the origins of bipedal locomotion. In *Handbook of paleoanthropology*, pages 1483– 1518. Springer, 2007.

[12] Loudon J et al. The clinical orthopedic assessment guide. 2nd ed. kansas: Human kinetics, 2008.

[13] Subotnick S. *Sports medicine of the lower extremity. Harcourt (USA):Churchill Livingstone*. Churchill Livingstone; 2nd edition, 1999.

[14] *http://www.biometricsinstitute.org/pages/types-of-biometrics.html*.

  