Executive Summary
Of
the thesis entitled

Optimizing and Analyzing Feature Selection for GAIT cycle detection in pathological GAIT using machine learning

Submitted by

Desai Meghna Tarun

Under the guidance of

Dr. Viral Kapadia

सत्यं शिवं सुन्दरम्

Department of Computer Science and Engineering Faculty
of Technology and Engineering
The Maharaja Sayajirao University of Baroda Vadodara
390 001

December 2025

Table of Contents

Table of Contents of Thesis

Introduction

The general practice followed in gait and motion analysis laboratory after conducting walking gait trials of patients with pathological gait is manual annotation of key gait events (Foot Strike and Foot Off) by expert biomechanists. Events marked by ground reaction force plates are considered gold standard[3] but due to expense considerations only few force plates can be installed in laboratories. Moreover, due to high inter-individual and intra-individual gait variability in pathological gait it may not be possible to detect Foot Strike and Foot Off using force plates many times. Hence, the biomechanical expert visually reviews the trials using softwares such as Mokka, Visual 3D and manually annotates the key gait events.

It is a crucial initial step to annotate key gait events in trial data. The detection of these events helps distinguish gait cycle, define the boundaries of gait phases and provide a framework for further analysis.

After key event detection/ annotation, the calculation of spatio temporal gait parameters like step length, stride length, cadence, speed and other kinematic factors like joint angles at specific points is done. Comparision of the extracted parameters to normative data or reference values helps derive resultant graphs which are used by pediatric expert doctors to identify underlying pathology, decide on treatment plans, and even to assess the effectiveness of interventions or rehabilitation programs.

Research Gap

1. Results of validation studies conducted, for some popular position-based and velocity-based algorithms (position and velocity of 3d markers), for a larger group of patients with different gait contact type patterns and pathologies, indicate that there is a scope for improvement in the accuracy of automated gait detection of key events.

~~2.~~ Existing popular methods like Sequence to Sequence LSTM lack interpretability and are complex to train. Moreover, with existing popular methods, training on smaller datasets can result in overfitting.

3. Approaches that result in less number or no false positives are desirable.

4. Machine learning/ deep learning approaches generally use 3D position of markers and their linear velocities for gait event detection. There is scope for optimizing feature selection and selecting parameters which can speed up training, increase the accuracy and reduce the complexity of the machine learning model.

Problem Statement

Pathological gait patterns vary widely between individuals and have underlying complexities. Experts manually mark foot strike and foot-off events by visually reviewing the gait trials of patients that undergo three-dimensional gait analysis. This task is time-consuming, affected by inter-rater variability, and prone to errors. Due to this, the results of three-dimensional gait analysis are affected in the context of speed and accuracy. Moreover, the availability of the data produced by gait analysis of pathological gait is limited. Additionally, the data format produced at different laboratories and sampling rates may also vary.

This research aims towards accurate automated foot strike detection using effective kinematic features, in pathological gait, which eventually can help gait cycle identification.

Objectives

The objectives of this research are :

1. To design and develop a machine learning model for foot strike detection in pathological gait using **effective kinematic features** with the following characteristics:

   (i) which is fast to train

   (ii) is preferably interpretable

   (iii) is trainable on datasets of smaller size (less samples) and scalable to larger size

   (iv) detects the event accurately and results in minimum false positives

Research Methodology for work done

1) An extensive literature study was conducted for automated key gait event detection algorithms in pathological gait with different approaches, coordinate based methods,

velocity threshold based methods and data driven machine learning approaches currently used for solving the problem.

2) A paediatric pathological gait data of more than 200 individuals were collected from Jupiter Hospital, Thane, Mumbai, India and kinematic data (3d trajectories and joint angles) was extracted from gait trials data, other features like linear velocities and accelerations were calculated. All the trials were visually inspected and divided intro three broad categories (i) heel strike foot contact pattern (ii) mid foot strike foot contact pattern (iii) front foot strike foot contact pattern.
3) Sequence to Sequence LSTM network (with peak detection for event frame identification) for automated key gait event detection (Foot Strike and Foot Off detection). LSTM requires large dataset in order to train. We applied different combinations of kinematic features to seq-to-seq LSTM to observe the improvement in results with smaller and effective features as compared to providing 3D trajectories of markers and their linear velocities and joint angles.
4) In this research, the combination of shapelet based classifiers, specifically Random Dilated Shapelet Transform, with a combination of simple rule based detection algorithm was used for automated and accurate detection of Foo Strike events. Other state of the art time series classifiers like ROCKET were also applied with a rule based algorithm for detection of foot strike events.
5) We applied three feature selection algorithms namely Random Forest , CLeVeR , Merit based feature selection to identify optimal feature set for Foot Strike event detection, in heel strike foot contact type pattern.

Conclusion

Outcome and Key Findings

1) With a limited dataset size and high variability in pathological gait patterns, the Shapelet Transform demonstrated effectiveness in identifying a small temporal window containing the event frame while remaining fully interpretable. The precise identification of the exact event frame was achieved using a simple rule-based detection algorithm applied within the identified window. Other state of the art time series classification algorithms like ROCKET, multi ROCKET hydra and Inception Time Classifier also worked efficiently for the classification of the small temporal window.

2) The study further shows that selecting effective kinematic features using suitable data-driven feature selection algorithms significantly improves the performance of machine learning models for automated foot strike detection, particularly when the dataset size is relatively small. Feature optimization reduced model complexity, improved generalization, and contributed to minimizing false positives compared to using all available marker trajectories.

Future Work

1) Feature selection methods can be applied for the other foot contact type pattern and accuracy for foot strike detection using machine learning algorithm, for those contact type patterns can also be increased.
2) After identifying a small window containing event frame, another algorithm like LSTM or a suitable algorithm can be used for increasing accuracy of foot strike detection.
3) Existing suitable feature selection algorithms can be tailored to design a feature selection algorithm specific to the problem.

References

1. J. Perry and J. Burnfield, Gait Analysis: Normal and Pathological Function. SLACK Incorporated, 2010
2. A. Esquenazi, &M. Talaty,. Gait analysis, Technology and clinical applications. Physical Medicine and Rehabilitation. , pp.99-116, 2011.
3. A.H.M AtaUllah& O. Jesus, Gait Disturbances, Stat Pearls, NCBI BookShelf, Jan 2022.
4. Hreljac, A., Marshall, R.N.: Algorithms to determine event timing during normal walking using kinematic data. Journal of Biomechanics. 33, 783–786 (2000).
5. Zeni, J.A., Richards, J.G., Higginson, J.S.: Two simple methods for determining gait events during treadmill and Overground walking using Kinematic Data. Gait & Posture. 27, 710–714 (2008).
6. Ghoussayni, S., Stevens, C., Durham, S., Ewins, D.: Assessment and validation of a simple automated method for detection of gaite vents and intervals. Gait & Posture. 20, 266–272 (2004)
7. Salazar-Torres J-D-J, Validity of an automated gait event detection algorithm in children with cerebral palsy and non-impaired children. Gait Posture, 2006
8. O'Connor, C.M., Thorpe, S.K., O'Malley, M.J., Vaughan, C.L.: Automatic detection of gait events using Kinematic Data. Gait & Posture. 25, 469–474 (2007).
9. Jasiewicz, J.M., Allum, J.H.J., Middleton, J.W., Barriskill, A., Condie, P., Purcell, B., Li, R.C.: Gait evento detection using linear accelerometersor angular velocity transducers in able-bodied and spinal-cord injured individuals. Gait & Posture. 24, 502–509 (2006).
10. E. Desailly, D. Yepremian, P. Sardain, P. Lacouture, Foot Contact event detection using kinematic data in cerebral palsy children and normal adults gait, Gait Posture, vol. 29, pp. 76-80 ,June 2008,
11. B-J Hsue, F Miller, F-C Su, J Henley, C Church, Gait timing event determination using kinematic data for toe walking children with cerebral palsy, Journal of Biomechanics, 2000
12. A.R. de Asha, M.A. Robinson, GJ Barton, A marker based kinematic method of identifying initial contact during gait suitable for use in real-time visual feedback applications, Gait and Posture, 2012:36(3):650-2
13. Bruening, D.A., Ridge, S.T.: Automated evento detection algorithms in pathological gait. Gait & Posture. 39, 472–477 (2014).
14. C. Gómez-Pérez, J. Martori, A.Josep M. Casanovas, J.Samsó, Josep M. Font-Llagunes,Gait event detection using kinematic data in children with bilateral spastic cerebral palsy, Clinical Biomechanics, Volume 90 ,2021, 105492
15. Visscher, R. M. S., Sansgiri, S., Freslier, M., Harlaar, J., Brunner, R., Taylor, W. R., & Singh, N. B. (2021). Towards validation and standardization of automatic gait event identification algorithms for use in paediatric pathological populations. Gait & Posture, 86, 64–69

16. R.V. Goncalves, S. T. Fonseca, P. A. Araujo, V. L. Araujo, T.M. Barboza, G. A. Martins, M. C. Mancini, Identification of gait events in children with cerebral palsy: comparison between force plate and algorithms, Brazilian Journal of Physical Therapy, 2019, https://doi.org/10.1016/j.bjpt.2019.05.007
17. Miller, A.: Gait event detection using a multilayer neural network. Gait & Posture. 29, 542–545 (2009).
18. Kidziński, Ł., Delp, S., Schwartz, M.: Automatic real-time gait evento detection in children using Deep Neural Networks. PLOS ONE. 14, (2019).
19. Lempereur, M., Rousseau, F., Rémy-Néris, O., Pons, C., Houx, L., Quellec, G., Brochard, S.: A new Deep learning based method for the detection of gait events in children with gait disorders: Proof-of-concept and concurrent validity. Journal of Biomechanics. 98, 109490 (2020).
20. Kim, Y.K., Visscher, R.M., Viehweger, E., Singh, N.B., Taylor, W.R., Vogl, F.: A deep-learning approach for automatically detecting gait-events based on foot-marker kinematics in children with cerebral palsy—which markers work best for which gait patterns? PLOS ONE. 17, (2022).
21. Ye, L., Keogh, E.: Time series shapelets: A new primitive for data mining. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p. 947–956. KDD ’09, Associationfor Computing Machinery, New York, NY, USA (2009)
22. Josif Grabocka, Nicolas Schilling, Martin Wistuba, Lars Schmidt-Thieme, Learning time-series shapelets. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2014) doi:10.1145/2623330.2623613.
23. Bostrom, A., & Bagnall, A. (2017). Binary Shapelet Transform for Multiclass Time Series Classification, Transactions on Large-Scale Data- and Knowledge-Centered Systems XXXII
24. Antoine Guillaume et al. “Random Dilated Shapelet Transform: A New Approach for Time Series Shapelets”, Pattern Recognition and Artificial Intelligence. ICPRAI 2022.
25. Guillemé, M., Malinowski, S., Tavenard, R., Renard, X. (2019). Localized random shapelets. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2019) Workshops (pp. 85-97). Springer, Cham.
26. Dempster, A., Petitjean, F., & Webb, G. I. (2020). ROCKET: Efficient and accurate time series classification with random kernels. Data Mining and Knowledge Discovery.
27. Monica Arul, Ahsan Kareem,Applications of shapelet transform to time series classification of earthquake, wind and wave data, Engineering Structures, Volume 228, 2021, 111564, ISSN 0141-0296,https://doi.org/10.1016/j.engstruct.2020.111564.
28. Ghalwash MF, Obradovic Z. Early classification of multivariate temporal observations by extraction of interpretable shapelets. BMC Bioinformatics. 2012 Aug 8;13:195. doi: 10.1186/1471-2105-13-195. PMID: 22873729; PMCID: PMC3475011.
29. Al Dhanhani, Ahmed, et al. "Framework for traffic event detection using Shapelet Transform."Engineering Applications of Artificial Intelligence82 (2019): 226-235.
30. Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems (NIPS).

31. Shajina T. and P. B. Sivakumar, "Human Gait Recognition and Classification Using Time Series Shapelets," 2012 International Conference on Advances in Computing and Communications, Cochin, India, 2012, pp. 31-34, doi: 10.1109/ICACC.2012.8.

32. Li Liu, Yuxin Peng, Ming Liu, Zigang Huang, Sensor-based human activity recognition system with a multi layered model using time series shapelets, Knowledge-Based Systems, Volume 90,2015, Pages 138-152,ISSN 0950 7051, https://doi.org/10.1016/j.knosys.2015.09.024.

33. Liu, L., Peng, Y., Wang, S., Liu, M., & Huang, Z. (2016). Complex activity recognition using time series patterndictionary learned from ubiquitous sensors. Information Sciences, 340, 41–57

34. Dempster, A., Schmidt, D.F. & Webb, G.I. Hydra: competing convolutional kernels for fast and accurate time series classification. *Data Min Knowl Disc* **37**, 1779–1805 (2023). https://doi.org/10.1007/s10618-023-00939-3

35. Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P.-A., & Petitjean, F. (2020). InceptionTime: Finding AlexNet for time series classification. Data Mining and Knowledge Discovery, 34(6), 1936–1962.

36. Breiman, L. Random Forests. Machine Learning **45**, 5–32 (2001) https://doi.org/10.1023/A:1010933404324

37. Yang, K., Yoon, H., & Shahabi, C. (2005). CLeVer: A feature subset selection technique for multivariate time series. In T. B. Ho, D. Cheung, & H. Liu (Eds.), Advances in Knowledge Discovery and Data Mining (PAKDD 2005) (pp. 516–522). Springer. https://doi.org/10.1007/11430919_60

38. Kathirgamanathan, B., & Cunningham, P. (2021). A feature selection method for multi-dimension time-series data. Pattern Recognition, [Article ID], https://doi.org/10.1016/j.patcog.2020.107525

39. Dhariyal, B., Nguyen, T. L., & Ifrim, G. (2021). Fast channel selection for scalable multivariate time series classification. In Advanced Analytics and Learning on Temporal Data (pp. 36–54). Springer. https://doi.org/10.1007/978-3-030-91445-5_3

40. Chollet, F., et al. (2015). Keras. https://keras.io

41. Middlehurst M, Ismail-Fawaz A, Guillaume A, Holder C, Guijo-Rubio D, Bulatova G, Tsaprounis L, Mentel L, Walter M, Schäfer P, Bagnall A (2024) aeon: a Python Toolkit for Learning from Time Series. Journal of Machine Learning Research 25:1–10

42. A. Mannini, D. Trojaniello, U. Della Croce and A. M. Sabatini, "Hidden Markov model-based strategy for gait segmentation using inertial sensors: Application to elderly, hemiparetic patients and Huntington's disease patients," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 2015, pp. 5179-5182, doi: 10.1109/EMBC.2015.7319558.

43. Talaei Khoei, T., OuldSlimane, H. &Kaabouch, N. Deep learning: systematic review, models, challenges, and research directions. Neural Comput & Applic 35, 23103–23124 (2023). https://doi.org/10.1007/s00521-023-08957-4