



Synthetic Data for Foot Strike Angle Estimation

Christoph Schranz¹^a, Stefan Kranzinger¹^b and Stephanie R. Moore²

¹Human Motion Analytics, Salzburg Research Forschungsgesellschaft mbH, 5020 Salzburg, Austria

²Department of Sport and Exercise Science, University of Salzburg, 5400 Hallein/Rif, Austria

Keywords: Data Augmentation, Human Running, GAN, Autoencoder, Foot Strike Angle.

Abstract: A runner's foot strike angle (FSA) can be relied on to assess performance, comfort, and injury risk. However, the collection of FSA datasets is time-consuming and costly, which may result in small datasets in practice. Therefore, the creation of synthetic FSA datasets is of great interest to researchers to improve the performance of machine learning models while maintaining the same effort in data collection. We evaluate data augmentation (jittering, pattern mixing, SMOTE) and synthetic data generation (Generative Adversarial Networks, Variational Autoencoders) methods with four subsequent machine learning models to estimate the FSA on a dataset involving 30 runners across a range of FSAs. The results show promising results for the SVM and MLP, as well as for the jittering and pattern mixing augmentation methods. Our findings underscore the potential of data augmentation to improve FSA estimation accuracy.


1 INTRODUCTION


Running is a widespread activity around the world, largely due to its limited equipment and facility requirements. It also has a positive impact on physical and mental health (Mikkelsen et al., 2017; Oswald et al., 2020). However, due to the physical forces acting on the joints, it is important to use proper footwear and running techniques to improve comfort and reduce the risk of injury and long-term joint health issues (Nigg et al., 2015). Therefore, the foot strike pattern (FSP) is a significant consideration, particularly in choosing suitable footwear (Zrenner et al., 2018).

Previous works have employed machine learning techniques for the estimation of foot strike angle (FSA) and FSP classification from pressure sensors (Moore et al., 2020). FSA is the angular degree of the foot at the moment of ground contact, and it is of importance because it affects numerous performance-related outcomes, such as vertical compliance, ankle and knee stiffness, vertical impact force, and instantaneous loading rates (Lieberman et al., 2010; Hamill et al., 2014; Cheung and Davis, 2011). Moore et al. (2020) compared the accuracy and precision of continuous FSA prediction and FSP classification

using multiple regression, conditional inference tree, and Random Forest (RF) (Breiman, 2001), employing data derived from Loadsol™ pressure insoles. The results have led to significant insights; however, the quest for enhanced accuracy in FSA estimation necessitates further investigation.

This study extends the work of Moore et al. (2020) who demonstrated the feasibility of two-sensor pressure insoles for detecting foot strike patterns and achieving over 90% FSP classification accuracy using multiple regression, conditional inference tree, and Random Forest. Moreover, the same methods were applied on the regression task of FSA estimation. However, the study is limited by the amount of mid-foot steps, types of evaluated machine learning models, and an ungrouped cross-validation scheme. Moore et al. (2020) proposed in their discussion that over- or under-sampling techniques and more complex machine learning algorithms may lead to an increased performance. Thus, we decided to employ state-of-the-art machine learning methods and applying data augmentation and synthetic data generation techniques to investigate the potential for enhanced FSA model accuracy when synthetic data is used. These techniques offer promise for enhancing the performance of machine learning models in FSA

^a <https://orcid.org/0000-0002-5786-7807>

^b <https://orcid.org/0000-0002-4014-7846>

estimation, thereby facilitating an even more nuanced understanding of running biomechanics and providing a tool for running shoe development and recommendation processes.

Data augmentation involves artificially expanding the dataset by applying transformations such as jittering (JIT), pattern mixing (PM), and Synthetic Minority Oversampling Technique (SMOTE) to the existing data points, thus enhancing the robustness of the model without the need for additional data collection. Synthetic data generation, on the other hand, utilizes methods like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) to create entirely new, yet realistic, instances based on the patterns learned from the existing data (Shorten and Khoshgoftaar, 2019; Iwana and Uchida, 2021; Jorge et al., 2018). Such techniques have shown potential in various fields, notably in scenarios with limited datasets, by enhancing model generalizability and preventing overfitting. In sports science, the application of data augmentation has been identified as a necessity to bridge the lab-to-field gap, however, only few approaches exist yet (Mundt, 2023).

Our research aims to utilize these innovative methods to augment the existing dataset, thereby enriching the input for subsequent machine learning models and further improving the estimation of FSA. The objective of this paper is to investigate to what extent data augmentation methods can compensate for the impact of a reduced number of participants. A secondary objective is to employ multiple downstream models in order to enhance the quality of the FSA estimations and to establish a more robust evaluation metric for the augmentation methods. We aspire to elevate the precision and reliability of FSA estimation. Ultimately, our goal is to provide a method that could support the processes of running shoe development and athlete training to improve performance and reduce the risk of injury.

2 MATERIALS AND METHODS

Our study included 30 injury-free male recreational runners (Mean \pm SD; 1.79 \pm 0.07 m; 80.1 \pm 9.6 kg; 34.0 \pm 6.9 yr). Participants were instructed to perform six foot strike conditions (extreme fore-foot, fore-foot, mid-foot, rear-foot, extreme rear-foot, and natural) at a comfortable speed in a randomized counterbalanced order. The vertical force of the insoles of each participant were captured using the Loadsol™ wearable sensors (Loadsol™; Novel GmbH; Munich, Germany) (Seiberl et al., 2018). In total, data were recorded for 3,489 steps.

2.1 Data Collection and Preprocessing

The Loadsol™ wearable sensors were utilized to measure insole forces during running at a sampling rate of 100 Hz. The captured time-series data were split into separate steps for analysis. The same insole outcome variables were used in the current study as in Moore et al. (2020); ten features were extracted for each step including four impulse ratios, two peak force ratios, and four ratios from the rate of force development.

In conjunction with kinetic data, a three-dimensional (3D) motion capture system (Qualysis system, 13-camera setup; 2019.3, Göteborg, Sweden) was used to optically measure the ground truth FSA, i.e., the angle of the foot at the initial contact on the ground. Six anatomical markers were applied to the left foot segment for kinematic data capture. For more information on the data collection and features, refer to Moore et al. (2020).

2.2 Downstream Models and Validation

Our study extended the original modeling approach by applying multiple machine learning models to estimate the FSA at ground contact. These models included RF (Breiman, 2001), Support Vector Machine (SVM) (Boser et al., 2001), XGBoost (XGB) (Chen and Guestrin, 2016), and a Multi-Layer Perceptron (MLP) (Hornik et al., 1989). A grouped cross-validation approach with k=10 folds was used (i.e., instances of the same participants were grouped into the same fold). For SVM and MLP, the features and target FSA were normalized.

Each model's hyperparameters were optimized through 200 iterations on the original data using a Tree-structured Parzen Estimator (TPE) (Bergstra et al., 2022). The estimation result from our RF was consistent with the approach from Moore et al. (2020) using basic cross-validation.

2.3 Data Augmentation and Synthetic Data Generation

We used data augmentation techniques to extend our dataset. For features measured within defined intervals (e.g., ratio values on the interval [0,1]), a Fisher's z-transformation was applied to prevent generating values outside the plausible range.

Data augmentation methods employed include:

- JIT (Iwana and Uchida, 2021): Gaussian noise was added, where noise intensity was

proportional to each feature’s standard deviation.

- PM (Iwana and Uchida, 2021): New instances were generated as a linear combination of two instances. Here, an alpha (α) was sampled from a normal distribution, and new instances were generated by $\alpha * X1 + (1-\alpha) * X2$.
- SMOTE (Chawla et al., 2002): A method used to balance class distribution in an unbalanced dataset by creating “synthetic” examples in the feature space, effectively combining aspects of jittering and pattern mixing techniques.
- VAE (Kingma and Welling, 2019): An encoder-decoder network that applies the “reparameterization trick” to sample the latent variable from a normal distribution with encoded parameters.
- GAN (Goodfellow et al., 2020): two separate networks are employed; One generates instances as realistic as possible, while the other distinguishes whether an instance is original or not. This results in a generative network able to create realistic instances.

For each combination of the five data augmentation and four downstream model, an optimization of their hyperparameters with 200 iterations was conducted. Each optimization included synthetic data for the training of the downstream model which was limited to five times the number of original samples.

Each combination of augmentation method and downstream model was trained on any number of participants. For this purpose, in the experiment, a varying number of participants was randomly sampled from the training fold of the cross-validation. Data augmentation was then applied to this subset before training the downstream model. The number ranged from only one randomly sampled participant to all available in the training fold which was at least 24 using a 10-fold cross-validation.

The Root Mean Square Error (RMSE) values of the estimations are aggregated and compared for a high number of participants ($n = 20-24$; Table 1) and for a reduced size ($n = 6-10$; Table 2) to investigate the effects of data augmentation for a significantly smaller dataset. All validations were performed solely on the original data of disjunct participants. No test-time augmentation, as described in Shorten and Khoshgoftaar (2019), was applied.

3 RESULTS

Figure 1 illustrates the influence of the number of participants on the RMSE for each augmentation method. The results are averaged across the four downstream models. For each augmentation model, the error decreases and converges at about 15 participants in the training fold.

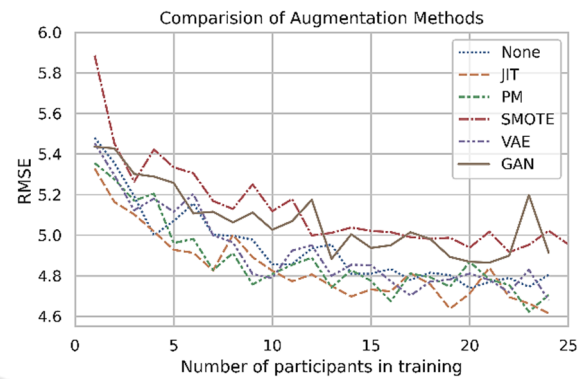


Figure 1: Comparison of augmentation methods, averaged across all downstream models. A higher number of participants used for augmentation and training decreases the RMSE.

JIT (orange) and PM (green) yield the lowest RMSE across all numbers of participants. VAE (violet) shows promising behavior for a higher number of participants.

Table 1 aggregates the obtained results from the grouped cross-validation experiment with higher participant numbers. The results represent the average RMSE within the range of 20 to 24 participants prevalent in each training fold to get a more robust measure for comparison. We tested four machine learning models (MLP, RF, SVM, XGB) using different data augmentation techniques and a control case without any augmentations ('None'). The 'Mean' column represents the average RMSE across the four downstream models for each augmentation technique. Bold numbers indicate the augmentation method with the lowest RMSE for each downstream model.

Table 1: Mean RMSE for 20-24 participants with 10 folds.

Method	MLP	RF	SVM	XGB	Mean ^a
None	4.751	4.984	4.449	4.892	4.769
JIT	4.524	4.785	4.739	4.771	4.705
PM	4.812	4.932	4.556	4.685	4.746
SMOTE	4.781	5.230	4.873	4.991	4.969
VAE	4.529	4.987	4.758	4.778	4.763
GAN	4.921	4.996	4.891	5.018	4.957

^a Mean of all downstream models in the same row.

The SVM model achieved the best results without any data augmentation (RMSE = 4.449). Following data augmentation, we observed the lowest RMSE with the MLP downstream model and the JIT and VAE, with an RMSE of 4.524 and 4.529, respectively. The SVM was the only downstream model that did not perform better after data augmentation.

The results summarized in Table 2 are obtained from our cross-validation experiment involving the average RMSE values across six to ten participants in each training fold to depict the effect of data augmentation on a low number of participants.

The SVM achieved the best results using PM augmentation with an average RMSE of 4.684 for the reduced training subsample ($n = 6-10$). Following data augmentation, PM resulted in the lowest mean RMSE across all downstream models (4.864), improving the score by 2.8% compared to no augmentation method.

Table 2: Mean RMSE for 6-10 participants with 10 folds.

Method	MLP	RF	SVM	XGB	Mean ^a
None	4.884	5.099	4.924	5.081	4.997
JIT	4.877	4.937	4.819	4.929	4.891
PM	4.800	5.115	4.684	4.830	4.857
SMOTE	5.202	5.352	5.135	5.086	5.194
VAE	4.875	5.092	4.882	4.961	4.953
GAN	5.137	5.108	5.060	5.072	5.090

^a Mean of all downstream models in the same row.

The more complex methods SMOTE and GAN failed to improve the average RMSE. VAE yielded only minor but consistent improvements. Despite the simplicity of JIT and PM, these results suggest that they performed best in improving the estimation accuracy of the FSA across all models tested in this study, especially for a lower number of participants.

4 DISCUSSION

Our study aims to enhance the accuracy of estimating FSA by using a suite of multiple machine learning models and data augmentation techniques. The best-performing approach of Moore et al. (2020), i.e., RF without augmentation, was replicated for the same ungrouped cross-validation scheme. This baseline was then enhanced by both employing preceding data augmentation and by selecting other machine learning methods.

Across varying numbers of participants, both JIT and PM augmentation methods consistently led to the lowest RMSE, indicating the highest accuracy in FSA

estimation. On the other hand, SMOTE appears to be less effective for this particular task, presumably as it was originally designed to tackle imbalanced classification problems.

VAE yielded only minor but consistent improvements, offering improvements comparable to those of JIT for MLP and XGB downstream models, as illustrated in Table 1. VAE might profit from an increased number of training instances to learn the inherent data distribution. A combination of VAE with a preceding JIT or PM might further improve the results by providing VAE with more data (Shorten and Khoshgoftaar, 2019). GAN was not successful in improving the RMSE of the FSA estimation. Similar to VAE (but more pronounced), GAN might require more data for training (Iwana and Uchida, 2021). The ineffectiveness of GAN could be due to too little data. Furthermore, GANs are designed to produce data that appear realistic such as images, and not to improve the quality of a subsequent downstream model applied on mixed data. Nevertheless, further investigations would be necessary to fully clarify the cause.

The improvements by employing data augmentation are small but consistent, therefore improving results without additional expensive data acquisition. Future work could explore augmenting time-series data for enhanced performance in synthetic data generation. Incorporating biomechanical constraints and more domain knowledge into augmentation methods has the potential to further improve the quality of the estimations. Additionally, the implementation of test-time augmentation methods (Shorten and Khoshgoftaar, 2019) could contribute to enhancing estimation accuracy, which is a research avenue that warrants further exploration.

Interestingly, SVM performed best without any data augmentation. This is possibly due to the fact that SVM minimizes in addition to the main objective, i.e., the MSE, also a regularization term. This regularization term penalizes the function implemented by an SVM to be as flat as possible to avoid overfitting for unseen instances. We, therefore, hypothesize that this regularization term helps the SVM to better represent the inherent data distribution than preceding augmentation methods.

We chose SVM for FSA estimation due to its strong performance on small to medium-sized datasets. Moreover, SVM can handle sparse high-dimensional feature spaces and is effective in dealing with non-linearly separable data using kernel features (Guido et al. 2024; Cyran et al. 2013). Furthermore, SVM has already been extensively validated in

biomechanical applications (see e.g. Begg et al. 2005; Halilaj et al. 2018), making it a reliable choice where sensor data often have complex relationships.

Mixed data augmentation strategies, unexplored in our comparisons, may yield improvements, particularly for complex methods like VAEs and GANs, that require larger datasets. An initial experiment has shown that the RMSE of GAN with SVM could be improved from 5.06 to 4.78 (for 6-10 participants) by applying JIT and PM prior to the training of the GAN, yielding better results than JIT alone.

One limitation of the experiments might be the setup for the hyperparameter optimization. The decision to use 200 iterations may be too restrictive, particularly given the complexity of models with up to 20 hyperparameters, such as GAN-XGB. Conversely, models with fewer hyperparameters, like the SVM downstream model, as well as the JIT, PM, and SMOTE data augmentation methods, might have been favored. A more comprehensive optimization could potentially enhance the performance of the other methods, in particular VAE and GAN.

The work established a preliminary step into synthetic data generation in the context of FSA estimation from mobile sensorics, focusing primarily on the comparison of methods. Future research should build upon these findings to explore new dimensions in augmentation and synthetic data generation, aiming to maximize the accuracy and utility of FSA prediction in real-world running scenarios. Ultimately, our goal is to provide a data generation method that supports the development of running shoes and athlete training for improved performance and injury prevention.

5 CONCLUSION

In conclusion, our work represents a step forward in the quest to incorporate data augmentation and synthetic data generation into the domain of wearable sensor development. We evaluated different combinations of methods for varying numbers of participants to estimate the FSA, with SVM improving the RMSE by more than 10 % compared to RF. The success of the simple JIT and PM method underscores the value of revisiting and adapting methods for more specific biomechanical constraints. Data augmentation methods adapted for specialized problems may have the potential to generate realistic synthetic data and therefore facilitate the development of more cost-effective algorithms for wearable sensors, thus enabling researchers to move

to field-based data collections with less intensive lab-based back-end development.

ACKNOWLEDGMENT

This work has been supported by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology under Contract No. 2021-0.641.557.

REFERENCES

- Begg, R. K., Palaniswami, M., & Owen, B. (2005). Support vector machines for automated gait classification. *IEEE transactions on Biomedical Engineering*, 52(5), 828-838.
- Bergstra, J., Yamins, D., & Cox, D. D. (2022). Hyperopt: Distributed Asynchronous Hyper-Parameter Optimization. *Astrophysics Source Code Library*, ascl-2205.
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).
- Breiman, L. (2001). Random Forests. *Machine learning*, 45, 5-32. doi: 10.1023/A:1010933404324/METRICS.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of artificial intelligence research*, 16, 321-357. doi: 10.1613/JAIR.953.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). doi: 10.1145/2939672
- Cheung, R. T. H., & Davis, I. S., (2011). Landing Pattern Modification to Improve Patellofemoral Pain in Runners: A Case Series. *Journal of Orthopaedic & Sports Physical Therapy*, vol. 41, no. 12, pp. 914-919, doi: 10.2519/jospt.2011.3771.
- Cyran, K. A., Kawulok, J., Kawulok, M., Stawarz, M., Michalak, M., Pietrowska, M., Widlak, P., Polańska, J. (2013). Support vector machines in biomedical and biometrical applications. In *Emerging paradigms in machine learning* (pp. 379-417). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2020). Generative Adversarial Networks. *Communications of the ACM*, 63(11), 139-144. doi: 10.1145/3422622.
- Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information*, 15(4), 235.
- Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J. L., Hastie, T. J., & Delp, S. L. (2018). Machine learning in human

- movement biomechanics: Best practices, common pitfalls, and new opportunities. *Journal of biomechanics*, 81, 1-11.
- Hamill, J., Gruber, A. H., & Derrick, T. R. (2014). Lower extremity joint stiffness characteristics during running with different footfall patterns, *Eur J Sport Sci*, vol. 14, no. 2, pp. 130–136, doi: 10.1080/17461391.2012.728249.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359-366. doi: 10.1016/0893-6080(89)90020-8.
- Iwana, B. K., & Uchida, S. (2021). An empirical survey of data augmentation for time series classification with neural networks. *Plos one*, 16(7), e0254841. doi: 10.1371/JOURNAL.PONE.0254841.
- Jorge, J., Vieco, J., Paredes, R., Sanchez, J. A., & Benedí, J. M., (2018). Empirical Evaluation of Variational Autoencoders for Data Augmentation, doi: 10.5220/0006618600960104.
- Kingma, D. P., & Welling, M. (2019). An Introduction to Variational Autoencoders. *Foundations and Trends® in Machine Learning*, 12(4), 307-392. doi: 10.1561/2200000056.
- Lieberman, D. E. (2010). Foot strike patterns and collision forces in habitually barefoot versus shod runners, *Nature*, vol. 463, no. 7280, pp. 531–535, doi: 10.1038/nature08723.
- Mikkelsen, K., Stojanovska, L., Polenakovic, M., Bosevski, M., & Apostolopoulos, V. (2017). Exercise and mental health. *Maturitas*, 106, 48-56. doi: 10.1016/J.MATU RITAS.2017.09.003.
- Moore, S. R., Kranzinger, C., Fritz, J., Stöggel, T., Kröll, J., & Schwameder, H. (2020). Foot Strike Angle Prediction and Pattern Classification Using Loadsol™ Wearable Sensors: A Comparison of Machine Learning Techniques. *Sensors*, 20(23), 6737. doi: 10.3390/s20236737.
- Mundt, M. (2023), Bridging the lab-to-field gap using machine learning: a narrative review, pp. 1–20, doi: 10.1080/14763141.2023.2200749
- Nigg, B. M., Baltich, J., Hoerzer, S., & Enders, H. (2015). Running shoes and running injuries: mythbusting and a proposal for two new paradigms: ‘preferred movement path’ and ‘comfort filter,’ *Br J Sports Med*, vol. 49, no. 20, p. 1290, doi: 10.1136/bjsports-2015-095054.
- Oswald, F., Campbell, J., Williamson, C., Richards, J., & Kelly, P. (2020). A Scoping Review of the Relationship between Running and Mental Health. *International journal of environmental research and public health*, 17(21), 8059. doi: 10.3390/IJERPH17218059.
- Seiberl, W., Jensen, E., Merker, J., Leitel, M. & Schwirtz, A. (2018). Accuracy and precision of loadsol® insole force-sensors for the quantification of ground reaction force-based biomechanical running parameters, *Eur J Sport Sci*, vol. 18, no. 8, pp. 1100–1109, doi: 10.1080/17461391.2018.1477993.
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of big data*, 6(1), 1-48. doi: 10.1186/S40537-019-0197-0.
- Zrenner, M., Ullrich, M., Zobel, P., Jensen, U., Laser, F., Groh, B. H., Duemler, B., Eskofier, B. M. (2018). Kinematic parameter evaluation for the purpose of a wearable running shoe recommendation. *In 2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN)* (pp. 106-109). IEEE. doi: 10.1109/BSN.2018.8329670.