Evaluating the performance of simulated IMU data for animal activity recognition

W. Vanwinsen
University of Twente
w.vanwinsen@student.utwente.nl

***Abstract*—Labeled data for IMU-based animal activity recognition (AAR) is difficult to collect due to major time- and resource constraints. As a result, the datasets which can be used for machine learning are often small and mostly cover domesticated animals such as cattle and horses. This scarcity of data hinders the development of robust predictive models for animal activity.**

To overcome this issue, researchers have proposed methods to generate virtual IMU-sensor streams from videos by using pose estimation and forward kinematics. The underlying idea being that this simulated IMU data is easier to obtain and label. This paper describes a theoretical framework for simulating IMU data from videos of animals.

This paper determines the viability of using simulated IMU data for activity recognition. Acceleration and angular rate estimates are derived from optical motion capture data of horses and compared to real IMU data through signal-level evaluation. The real and simulated data is pre-processed to increase their correlation. Additionally, several machine learning models (SVM, Random Forest, KNN and Naive Bayes) are trained using real, simulated and a combination of both to assess the effects on model performance. In the latter case, augmentation of real data with simulated data can increase the average F1-score by more than 20% when little real data is available for training. Classifiers trained using only simulated data show predictive performance that is inferior to that of classifiers trained using real data. Nevertheless, competitive recognition accuracy can be achieved.

***Index Terms*—machine learning, animal activity recognition, pose estimation, computer vision, IMU, optical motion capture, support vector machine, random forest, decision tree, knn, naive bayes, data augmentation**

I. Introduction

Tracking animal activity over time can provide us with a rich source of information regarding their behaviour and well-being [12, 13, 17, 38]. For this reason, it can potentially serve many applications in areas such as wildlife monitoring and preservation [2, 24, 31], smart farming [1, 7, 29], veterinary care [33] and e-health of pets [9].

Advances in sensing technology have made it possible to attach lightweight, non-intrusive sensors to animals for the purpose of tracking their activity [12, 38]. Machine learning is commonly employed to recognize behavioural patterns from collected sensor data.

These developments have shifted the environment in which we can study animal activity from controlled laboratory environments, to the natural habitats of the observed species. However, this shift has proven to be difficult to realize. Collecting data for activity recognition of wildlife is a tedious process that requires a lot of time and resources.

A commonly applied sensing modality is the inertial measuring unit (IMU), a device that measures acceleration, angular velocity and magnetic field intensity over time. While collecting IMU data in the wild is feasible, labelling this data truthfully is problematic. There are many impracticalities bound to taking IMU measurements while filming animals simultaneously. The need for synchronization between sensors and camera's [14], the vast areas that animals cover in the wild and the potential of scaring animals away through human presence are all examples of this. Due to these reasons, in-the-field data collection is very time- and resource intensive and yields little labelled data.

State of the art techniques in human activity recognition (HAR) have shown that it is feasible to simulate IMU data using optical motion capture data (OMC) [28] and video frames [8, 19]. This is a promising notion for activity recognition of non-domestic animals. Pose estimation and forward kinematics can be applied to existing video footage of animals to generate simulated IMU streams at multiple locations of the body (figure 1). This new method for data collection is potentially more convenient than current procedures as it eliminates the need for attaching IMU sensors to animals.

The results obtained during the evaluation of said methods is not representative for animal activity recognition due to several reasons. First of all, they were developed for HAR problems. Generalization to other species is not straightforward because they either rely on human pose estimation models [4, 6] or 3D human mesh models [21] which are not available for animals. Additionally, the scale of most HAR datasets is considerably larger than that of most AAR datasets. Finally, the differences between simulated and real IMU data are shortly discussed in current literature but there is a lack of signal-level evaluation.

The goal of this research is to fill these gaps by providing an evaluation of using simulated IMU data in the context of AAR. The main question to be answered is: *To what extent can simulated IMU data be used for animal activity recognition?*. A horse dataset containing IMU and OMC measurements is used for this evaluation. Acceleration and angular rate estimates are derived from the OMC annotations and the resulting signals are compared to real IMU data in both the time- and frequency domain. Furthermore, filtering is applied to increase correlation between both signals.

To assess whether simulated IMU data can be used to recognize real IMU data, four well-known machine learning models (SVM, Random Forest, KNN and Naive Bayes) are trained using real, simulated and a combination of both types

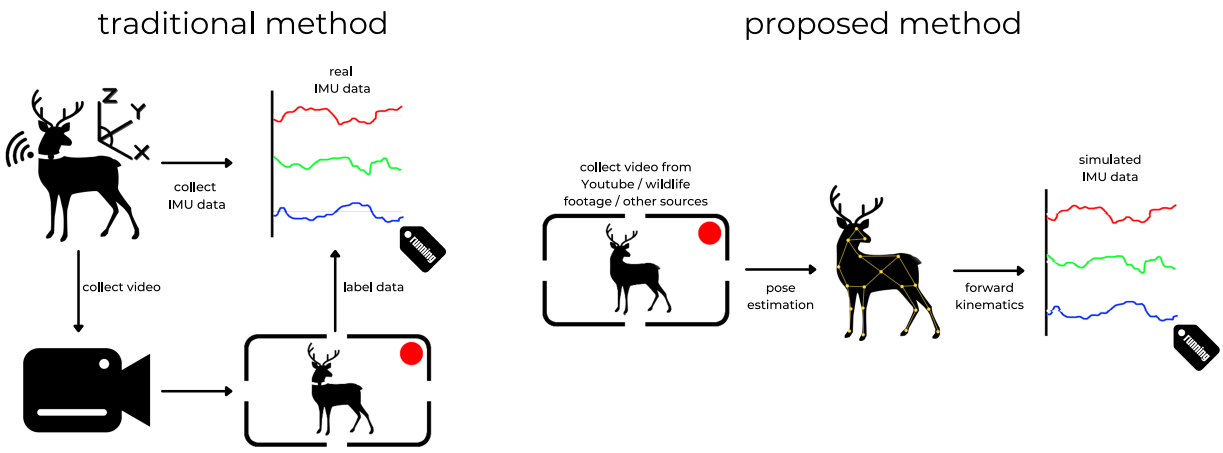


Fig. 1: High-level overview of the traditional method for IMU data collection versus the proposed method.

of data. Several experiments are conducted in which a varying amount of real training data is augmented using simulated data. Lastly, the performance of classifiers trained using solely simulated data is evaluated and compared with that of classifiers trained with real data.

Section 2 provides an overview of the state of the art research for simulating IMU streams from various sources. In section 3, the contents of the horse dataset are described and in section 4 and 5 the different experiments are described and evaluated respectively. Finally, the outcome of the research is discussed briefly in section 6.

II. Related work

The potential benefits of using simulated motion data for activity recognition has only appeared in academic literature since 2019. The available literature can be divided into two categories based on the used sensing modality: optical motion-capture-based and video-based methods. The state of the art and their limitations will shortly be discussed.

A. OMC-based methods

IMU data has been augmented with acceleration and angular rate estimations derived from optical motion capture. The majority of these approaches rely on the SMPL body model proposed by Loper et al [21]. Virtual IMU sensors are positioned on the body mesh of the model and forward kinematics are applied to simulate IMU data [10, 28, 35].

Pei et al. [28] evaluate this simulated data using a variety of machine learning algorithms. Each classifier is trained on solely real IMU data and compared to a classifier that has been trained using a combination of both real and simulated data. They report higher accuracies for classifiers trained with the mixed data.

While this demonstrates the benefits of using simulated data, the experiments did not assess the performance of which classifiers are trained using only simulated data. Moreover, the simulated data is not compared to real data at a signal-level. Additionally, this method has some limitations that make it difficult to generalize to animal activity recognition. First of all the method relies upon a realistic 3D model of the human body which to our knowledge is non-existent for animals. Secondly, the approach uses a large-scale OMC dataset for evaluation. The currently available motion capture dataset for animals are of much smaller size.

Takeda et al. [34] take a less sophisticated approach to get acceleration estimates by taking the second derivative of position annotations made by OMC systems. The derived accelerometer data is then used to train a random forest and support vector machine classifier and compared with these same models that have been trained with real accelerometer data. Varying results are achieved during prediction of the test set. In certain cases the models trained using real data vastly outperform the models trained with simulated data. However, cases in which the opposite holds are observed as well. The conducted experiments do not take the orientation of the IMU sensors into account which could be a possible explanation for these observations. Sensor orientation changes constantly during movement and is dependent on sensor placement. A change in sensor orientation can alter sensor readings over time which could result in classification error. In this research, the extracted features are made orientation independent to avoid this (Section 4).

B. Video-based methods

Methods based on video have an apparent advantage over other proposed methods due to the wide availability and accessibility of data on video repositories such as YouTube.

Current research literature applies off-the-shelf human pose estimation models [4, 6] to obtain 2D and 3D pose estimates from video. Kwon et al. [19] applies forward kinematics on 3D poses to get acceleration and angular rate estimates. Rey et al. [8] apply similar methods but instead directly regress from 2D pose estimates obtained using OpenPose [4] to corresponding IMU sensor streams for several joints using a convolutional network architecture.

Both approaches found that the addition of simulated IMU data to real data for model training can be beneficial for model performance as higher F1-scores were achieved during evaluation. Another commonality between both papers is that training activity classifiers solely on simulated data leads to inferior performance when classifying real IMU data.

Several reasons can be given for this. First of all, current pose-estimation models are prone to produce erroneous poses due to factors such as camera ego-motion, occlusion, blurring and the subject moving within the frame. Getting an accurate pose estimation in 3D proves even more complex due to the apparent lack of depth information in common video frames. The inaccuracies of the estimated pose create discrepancies between real and simulated motion data.

C. Activity recognition

Machine learning models such as Support Vector Machines [15, 23, 32], Hidden Markov Models [20], Decision trees [18] and Naive Bayes [15, 16] are often appied to animal activity recognition problems. While these traditional machine learning methods can achieve high recognition accuracy they often tend to fail at generalizing to other problems.

In more recent work, artificial neural networks such as CNN's and LSTM's [11, 36] are used. The main advantage of these models is that they mitigate the need for careful feature selection and extraction by automatically exploiting the statistical relationships embedded in the data.

A recent point of critique for applying deep learning to activity recognition problems is that they tend to under-perform when not enough data is available for training [30]. Activity recognition datasets are often too small to fully reap the benefits of using deep neural networks. Due to the limited amount of data available during this research, traditional machine learning methods are favoured over deep-learning methods.

III. Dataset

The dataset used for evaluation was collected by Bragança et al. [3, 33] and was originally used as a means to better recognize abnormal gait in horses. The dataset contains IMU and OMC data as well as video recordings of 8 different horses performing two different activities (walking and trotting). These same activities were performed while abnormal gait was safely induced on either the fore- or hindlimbs to imitate lameness of the horse. The relevant aspects of the dataset will be described.

A. IMU Data

For each subject, measurements were taken at 8 different locations of the body (figure 2). Acceleration, angular rate and magnetic field intensity were measured at a sampling rate of 200 Hz. Approximately 37 minutes of labelled IMU data is available in total. The activity distribution of this data has been visualised in figure 3.

B. OMC Data

Optical motion capture annotations were recorded at 63 locations using reflective markers. The OMC system only tracks

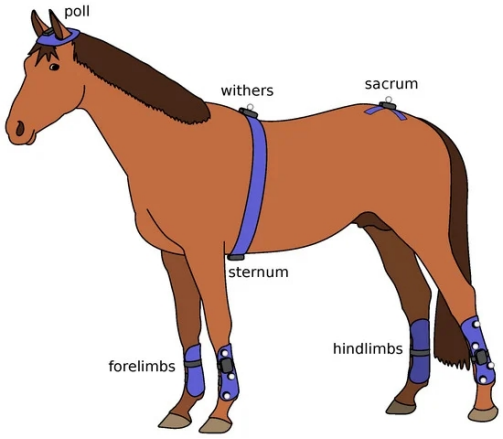


Fig. 2: Positioning of the IMU sensors

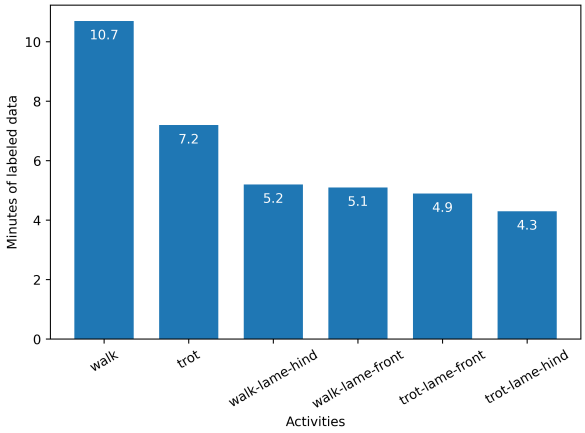


Fig. 3: Activity distribution of all available IMU data

the displacement. To obtain orientation data, multiple reflective markers were placed in close proximity to one another around the limbs of the horse. This cluster of reflective markers can be tracked as a single rigid body for which orientation can be derived. As a result, the real and simulated IMU data can only be compared along the four limbs of the horse.

IMU data is simulated from the OMC data using methods described by Bragança et al [3]. The coordinate systems of both the IMU and OMC system has to be alligned and the derivative of the position and rotation matrices is taken to obtain acceleration and angular rate estimates. The exact procedure is proprietary to the authors of named papers and will not be elaborated on any further. The distribution of the resulting simulated IMU data is visualised in figure 4. There is considerably less simulated data than real data available (Table I).

IV. Method

In section 4A, a framework for video-based IMU data generation for animal activity recognition will be proposed. This has been provided to serve as a baseline for further research. Before developing such a framework it should first be assessed

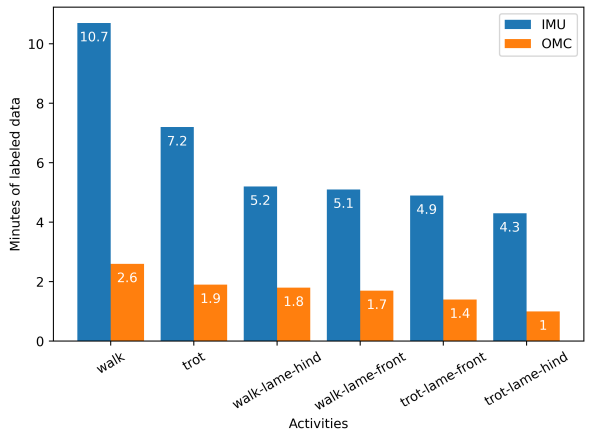


Fig. 4: Comparison of activity distribution between real and simulated IMU data

Horse	Minutes of labelled data	
	Real	simulated
Brianna	4.6	2.6
Colinda	5.7	0.4
Cuzdine	3.9	0.2
Hertogin	5.8	None
Iceman	5.6	1.2
Vlotte	6.9	3.0
Willarda	5.0	3.0

TABLE I: Data distribution between subjects

whether virtual IMU data can be beneficial to animal activity recognition in the first place. This will be evaluated in section 4B which provides a comparison of real and simulated IMU data in terms of signal characteristics as well as its effect on recognition accuracy

A. The Framework

The following theoretical framework serves as a proposal for a more convenient data collection procedure for animal activity recognition applications. This framework is based on a popular computer vision problem: pose estimation.

Some research concerns itself with 2D pose estimation while others attempt to predict poses in 3D space. The latter has proven to be a more complex problem to solve due to the obvious lack of depth information in common RGB images, more intensive data collection procedures as well as additional computational resources needed to account for this extra dimension.

The 2-dimensional heatmaps derived from RGB images do not provide enough information to infer the depth information needed to predict poses in 3D space. To tackle this, researchers have either resorted to using GAN's [5] or volumetric (3D) heatmaps. [22, 27].

Volumetric heatmaps have the disadvantage of being computationally heavy. This issue was partly resolved by Pavlakos et al.[27] by using lower-resolution heatmaps throughout the network. In 2018, Nibali et al. [26] introduced the idea of using a 2D heatmap along each axis instead of a single volumetric heatmap. With this novel approach, they achieved state of the art performance with less computational resources being required.

In response to the previously discussed arguments, an adaptation of the research by Nibali. et al. [26] is proposed. A challenging requirement for this method is the need for video footage for which ground-truth 3D annotations are available. The dataset [3] as described in Section 3 would be suitable as OMC data is available for the horses. Once 3D pose estimates are generated they can be converted to virtual IMU streams using methods described by Kwon et al. [19].

B. Signal-level evaluation

The real- and simulated IMU signals will be compared through signal analysis. Filtering will be applied to reduce signal noise and to assess the effect on signal correspondence.

Signal characteristics of the different gaits will be compared through distribution measures such as the mean, standard deviation and range. Similarly, the degree of correlation between the real and simulated data will be investigated in both the time- and frequency domain. To determine the correlation in the frequency domain, Welch method [37] is applied to obtain an estimate of the power spectral density (PSD) function of the signal. The correlation between the resulting PSD functions is then calculated.

C. Activity recognition evaluation

A series of experiments will be conducted to assess the viability of using simulated IMU data in the context of activity recognition. Four machine-learning models are used to classify horse gaits. These are Random Forest, Support Vector Machine, k-Nearest Neighbours and Naive Bayes respectively.

Pre-processing of IMU data

While the IMU sensors are securely attached to the body of the horse, animals living in the wild often are given a collar housing the sensors. These collars are subject to moving around due to rough conditions. As a result, these shifts cause changes in sensor orientation over time which can potentially lead to classification error [15, 17]. To avoid this, the signals are made orientation independent by using the magnitude, also known as the norm of the signal (equation 1).

$$\|a\| = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{1}$$

The resulting norms are filtered based on the outcomes of the signal evaluation. The data is segmented into windows with 50% overlap. A window contains $(2 \cdot L_w) \cdot n_s$ where L_w is the length of the window in samples and n_s the number of IMU sensors.

Two features are extracted from each signal to reduce the dimensions of the input vectors considerably to $4 \cdot n_s$. These are *energy* and *consistency*. The energy is defined as the mean of the magnitude within a window and the consistency is defined as $1 - \alpha$ where α is the standard deviation of the

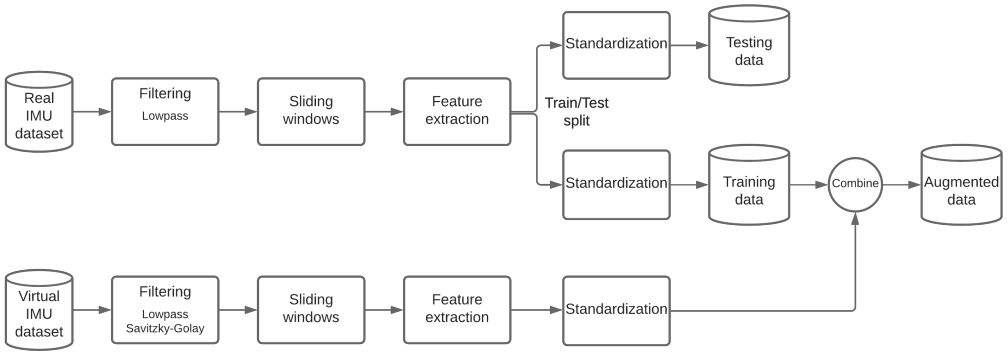


Fig. 5: Pre-processing chain used during experiments

magnitude [25].

Activity sets

Several activity sets were made for evaluation (table II). The complexity of the problem, as well as the number of classes, is varied with each subset. In the first sets, we only differentiate between sound and lame gait whereas later on, the position of said lameness is also taken into account.

Set #	Activity classes
1	walk, trot
2	lame, not-lame
3	walk, walk-lame
4	trot, trot-lame
5	walk, walk-lame, trot, trot-lame
6	walk, walk-lame-front, walk-lame-hind
7	trot, trot-lame-front, trot-lame-hind
8	walk, walk-lame-front, walk-lame-hind, trot, trot-lame-front, trot-lame-hind

TABLE II: Horse activity subsets

Experiment 1: Augmentation of real IMU data

A Varying amount of real IMU data is augmented using simulated IMU data. This experiment is carried out to determine two factors: 1) The effect on model accuracy when adding more and more simulated data for each fixed size of real data. 2) The amount of real data at which adding simulated data might become redundant.

Both real and simulated IMU data undergo pre-processing (figure 5). A subset of fixed size is then randomly chosen from both the real and simulated training data based on fractions of the total available data. For the real data, 10, 20, 40, 60, 80 and 100 percent of the available data is used during the experiments. For the simulated data similar percentages are used. The ratio of real to simulated training examples is varied with each experiment by iteratively increasing the amount of simulated training data for each fixed size of the real training set

The combined dataset is used for hyperparameter optimization, each models are tuned using a Bayesian Optimization scheme before each training session. Each model is evaluated using the same real test set.

Experiment 2: Evaluation of simulated IMU data

The ability of classifiers trained using solely simulated data will be determined and compared to that of real data. Furthermore, the effect of filtering during the pre-processing stage will be assessed. Each model is trained 3 times: with filtered simulated data, with unfiltered simulated data and finally with the corresponding real data (same size and class distribution). The same pre-processing and hype-parameter tuning procedures as described in experiment 1 are used throughout the experiment. The difference is that no data augmentation takes place.

V. Results

A. Signal-level evaluation

The measurements of horse Vlotte were used during signal analysis as most simulated IMU data was available for this particular horse.

Filtering procedure:

Figure 6 depicts a short sample of the raw, unfiltered acceleration signal along the X-axis (left) and the corresponding PSD function (right). The simulated signal contains more high-frequency noise than its real counterpart. Nevertheless, a very strong correlation already exists between both PSD functions.

In both signals, the vast majority of power lies between the range of 0 to 30 Hz. A low-pass filter with a cutoff frequency of 40Hz is applied to both signals. The result is visualised in figure 7. Signal correspondence has increased considerably in the time-domain whereas the correlation in the frequency domain has slightly decreased.

The cutoff frequency of the real signal was fixed at 40Hz as reducing it any further resulted in losing some prominent components. The simulated data was processed using an additional 3rd order Savitzky-Golay filter with a window length of 11 samples (figure 8). This was preferred over using a lower cutoff frequency as the width and position of signal peaks were better preserved. The correlation coefficient between the real and simulated signal is increased by almost 30% in the time domain and by 0.1% in the frequency domain.

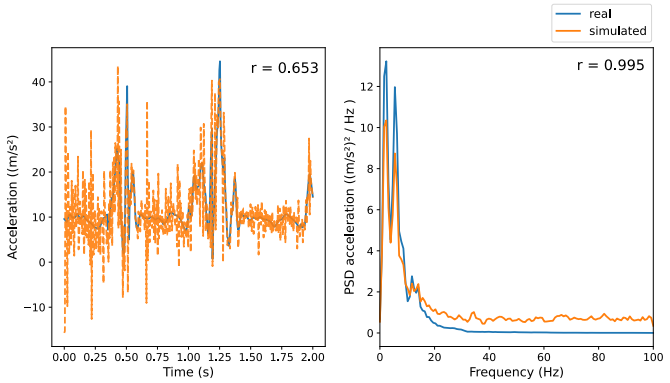


Fig. 6: Raw, unfiltered acceleration signal

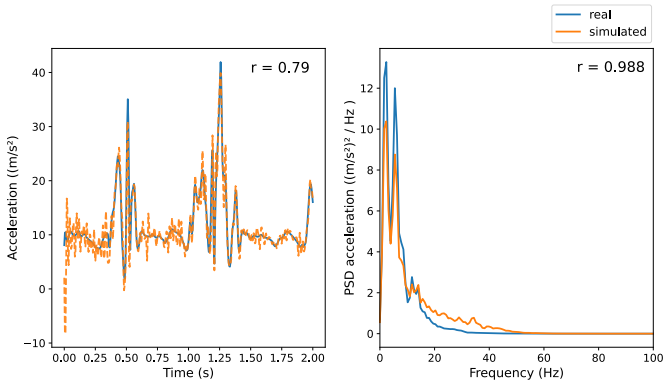


Fig. 7: Acceleration signal (40Hz low-pass)

This filtering procedure is applied to all of the signals.

Comparison of gait types: Figure 9 depicts two samples of the real and simulated angular rate along the Z-axis. A clear distinction between the two gaits can be made through visual comparison, as a higher angular rate and a faster cadence can be observed during trotting.

The statistical measures substantiate these observations. Trotting has a higher standard deviation and range compared to walking. Notably, the range of the simulated data is narrow compared to that of the real signal. This is a common observation across all sensor readings and gaits. Due to filtering, some of the peaks of the simulated signal were attenuated. A trade-off was made between reducing the noise in the signal and preserving said peaks.

A strong positive correlation exists between real and simulated signals as The OMC system was able to accurately capture movements along the sagittal plane of the horse.

Comparison of lame and sound gaits

Figure 10 and 11 illustrate similar samples for each gait type. Lameness was induced on either the fore- or hindlimbs respectively. Both the sound and lame trot have a similar distribution of values and a strong correlation between real and simulated signals. A large discrepancy between real and

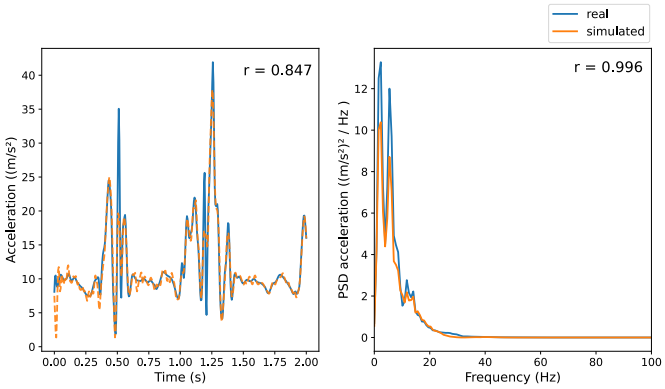
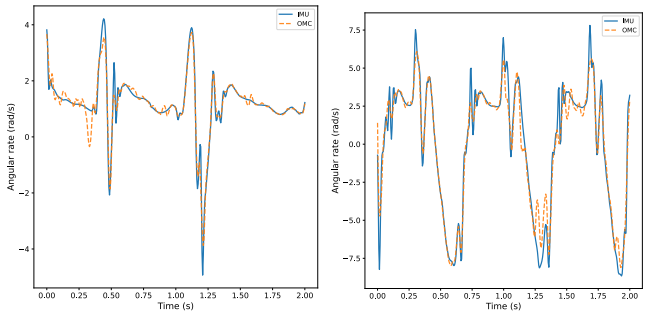


Fig. 8: Acceleration signal (40Hz low-pass + Savitzky-Golay smoothing)



(a) Walking (sound)

metric	IMU	OMC
mean	0.13	0.12
std-dev	2.46	2.31
range	(-7.5, 4.5)	(-6.8, 3.88)
t-corr	0.98	
f-corr	0.99	

(b) Trotting (sound)

metric	IMU	OMC
mean	-0.05	-0.04
std-dev	4.2	3.8
range	(-9.8, 9.8)	(-8.6, 6.5)
t-corr	0.97	
f-corr	1.0	

Fig. 9: Comparison of real and simulated angular rate measured along the Z-axis for sound gait

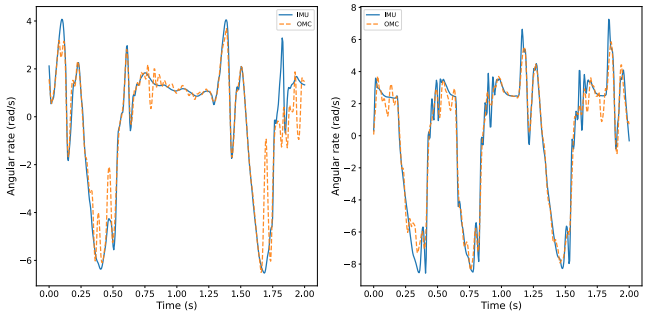
simulated data is seen when lameness is induced at the backlimbs during walking. A very weak correlation exists between both signals in the time domain whereas the PSD functions correlate well. Reasons for this could be inaccurate tracking of the reflective markers during this experiment or flaws in the conversion of OMC annotations to angular velocity and acceleration values.

Additional statistical summaries of each activity class can be found in appendix A.

B. Activity recognition evaluation

Experiment 1: Augmentation of real IMU data

Only the acceleration and angular rate signals measured at the limbs of the horse were used during the experiments. Consequently leading to an input vector containing 16 features. The length of each sample window was fixed at 3 seconds (600 samples). Figure 13 depicts the feature space of acceleration measured at the left forelimb. A clear distinction between gait types can be made whereas there is a lot of overlap between sound and lame gait.



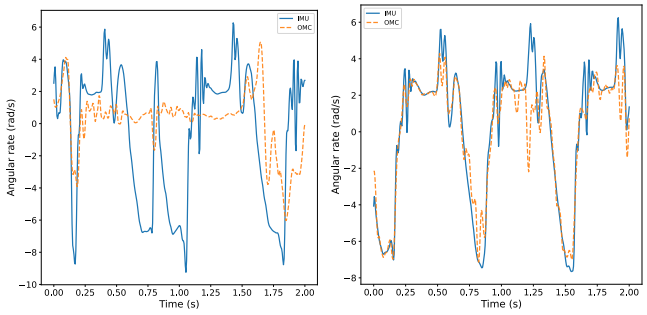
(a) Walking (lameness at forelimb)

(b) Trotting (lameness at forelimb)

metric	IMU	OMC
mean	0.12	0.07
std-dev	2.56	2.39
range	(-7.4, 3.3)	(-7.0, 4.3)
t-corr	0.95	
f-corr	1.0	

metric	IMU	OMC
mean	-0.18	-0.16
std-dev	4.38	3.98
range	(-10.0, 9.4)	(-9.3, 6.1)
t-corr	0.97	
f-corr	1.0	

Fig. 10: Comparison of real and simulated angular rate measured along the Z-axis for lame gait (front)



(a) Walking (lameness at hindlimbs)

(b) Trotting (lameness at hindlimbs)

metric	IMU	OMC
mean	-0.05	0.16
std-dev	2.62	1.93
range	(-9.7, 7.7)	(-6.5, 6.32)
t-corr	0.05	
f-corr	0.99	

metric	IMU	OMC
mean	0.00	0.00
std-dev	4.09	3.69
range	(-8.7, 8.5)	(-8.4, 5.9)
t-corr	0.97	
f-corr	1.0	

Fig. 11: Comparison of real and simulated angular rate measured along the Z-axis for lame gait (back)

Figure 14 and 15 shows the experimental results over all activity sets when augmented with a small and a larger amount of simulated data respectively. In both cases the addition of simulated data to the training set has caused both increases and decreases in F1-score during evaluation. There are also many instances in which the effect of data augmentation can be considered negligible as the F1 score of the model trained with the augmented set was either equal or within very close proximity to that of the model trained with non-augmented data.

A trend can be seen in which the positive effect of augmentation gradually diminishes as the amount of real data is increased. The same trend is observed in the variance of the F1-scores. A possible explanation for these observations is that the difference between the real and simulated data is

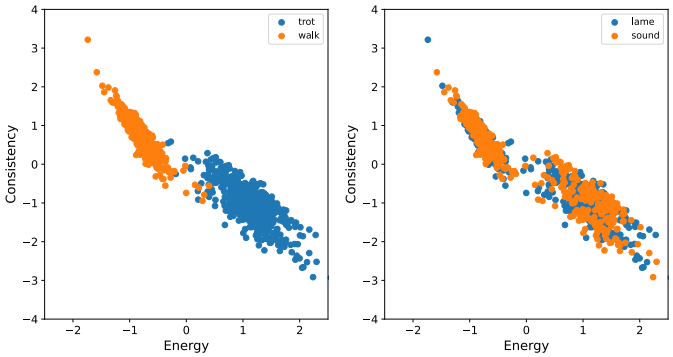


Fig. 12: 2D feature space of real IMU data (standardized)

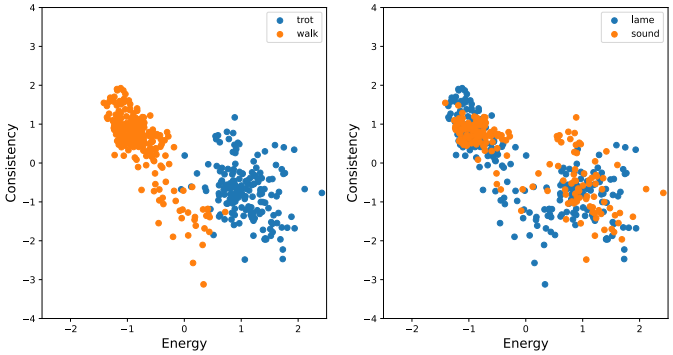
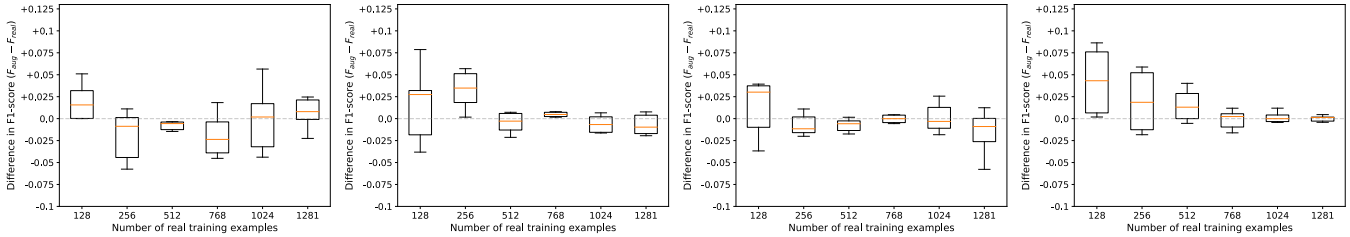


Fig. 13: 2D feature space of simulated IMU data (standardized)

too small to further improve the learned parameters of each classifier once sufficient data is available.

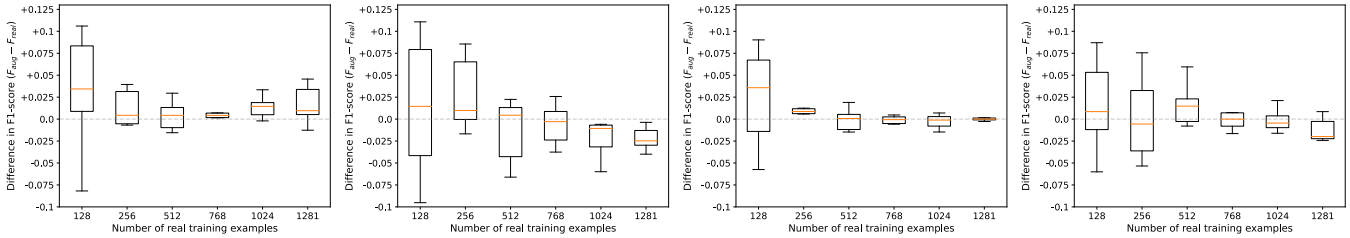
There is some evidence to suggest that the amount of simulated data used for data augmentation affects recognition accuracy. More extreme increases as well as decreases in F1-score are observed when a greater amount of simulated data is used.

The effect of data augmentation varies a lot depending on the presented classification problem. For this reason, Each activity set will be discussed individually using a series of visualisations (figure **??**-**??**). Each bar indicates the obtained F1-score during evaluation. The absence of an orange bar represents that the model trained using the augmented set did not outperform the model trained on the non-augmented set. Cases in which the increase in F1-score was smaller than *0.005* were excluded from the visualisations. More extensive figures can be found in appendix B.



(a) Support Vector Machine (b) Random Forest (c) k-Nearest Neighbours (d) Naive Bayes

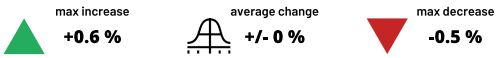
Fig. 14: How data augmentation (sample size = 82) affects F1-score over all activity sets



(a) Support Vector Machine (b) Random Forest (c) k-Nearest Neighbours (d) Naive Bayes

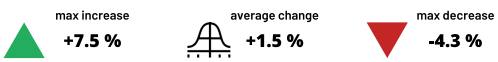
Fig. 15: How data augmentation (sample size = 411) affects F1-score overall

Activity Set 1: Recognizing gait type regardless of lameness.



classes: walk, trot

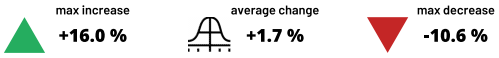
(Figure 17) - Despite using different models and varying the amount of real and simulated data in the training set, the F1-score consistently lies between 0.97 and 1.0 when training with both the augmented and non-augmented set. The results of this experiment are visualised in figure 17. As explained during signal analysis, walking and trotting are generally easy to differentiate. Due to their apparent differences, very little data is needed to get high predictive performance. Data augmentation has a minimal effect on this classification problem. *Activity Set 2: Recognizing lameness regardless of gait type.*



classes: lame, not-lame

(Figure 18) - Depending on the amount of data, an average F1-score between .5 and .8 can be achieved on the test sets. The highest improvements occur when a small amount of real data is available, yet increases in F1-score are also observed when larger amounts of real data are used.

Activity Set 3: Recognizing lameness during walking.

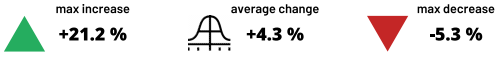


classes: walk, walk-lame

(Figure 19) - When isolating lameness in walking, data augmentation can lead to a small improvement in recognition accuracy. Naive Bayes trained using the augmented set outperformed the non-augmented version after every training session. Still, its predictive performance remains lower than that of the other models.

An interesting observation is that the baseline accuracy of the other models temporarily decreases when increasing the number of real training samples from 128 to 256. A possible explanation is yet to be found for this phenomenon. Most improvement through data augmentation occurs when lower sample sizes are used for training, which is in line with results from other activity sets.

Set 4: Recognizing lameness during trotting.



classes: trot, trot-lame

(Figure 20) - By isolating trotting, considerable improvements in recognition can be achieved through data augmentation. Yet again the greatest increase can be found when real data is scarce.

Noticeably, obtained F1-scores are higher in general compared to classifying lameness during walking. There are several possible explanations for this. First of all, lameness becomes more noticeable when the forces exerted on the limbs of the horse become greater. A horse will feel more discomfort while

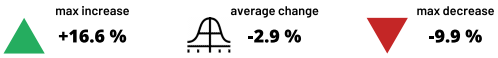
trotting, causing a more pronounced change in its movement pattern. Secondly, there is a large error between the real and simulated signals during walking. The addition of simulated data is thus less beneficial as it does not resemble the real data as much.

Set 5: Recognizing lameness for both gaits.

classes: walk, walk-lame, trot, trot-lame, not-lame

(Figure 21) - As seen before, classifiers mostly benefit from data augmentation when using less real training examples. However, some slight improvements in F1-score is also witnessed at greater sample sizes. Data augmentation causes the most inferior performance when using Naive Bayes, for which augmentation resulted in an average decrease in F1-score of 13%. The model with the highest accuracy is the Support Vector Machine, which had an average increase in F1-score of 1.8% after augmentation.

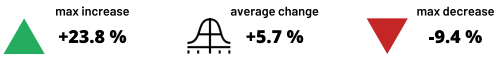
Set 6: Recognizing the position of lameness during walking.



classes: walk, walk-lame-front, walk-lame-hind

(Figure 22) - Data augmentation is the least beneficial to classification given this problem. An explanation could be the dissimilarity between real and simulated data seen during walking. Furthermore, the detection of lameness in a walking gait is more complex than in a trotting gait. The best performing model (SVM) saw an average increase of only .3% after augmentation. Regardless, improvements in F1-score were observed at low sample sizes. With increases of 11.3%, 5.9% and 16.6% for SVM, KNN and Naive Bayes respectively.

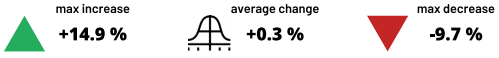
Set 7: Recognizing the position of lameness during trotting.



classes: trot, trot-lame-front, trot-lame-hind

(Figure 23) - Considerable increases in classifier performance are seen consistently at lower sample sizes. The performance of each model strongly resembles that of the classification problem in which the position of lameness was not taken into account (figure X). Thus there must be a clear distinction between movement patterns when lameness is induced on the back-limbs versus the fore-limbs of the horse. An average increase of 5.7% is realized across all models after data augmentation.

Set 8: Recognizing the position of lameness for both gaits.



classes: all

(Figure 24) - The inclusion of walking gait results in a slight drop in model performance. Nevertheless, F1-scores with and without data augmentation can reach up to 0.75 when sufficient data is available. Improvement through data augmentation is consistently achieved for the Random Forest although marginal with an average increase of only .3% through augmentation

Experiment 2: Evaluation of simulated IMU data

The experimental results are depicted in figure 16. Two activity sets were evaluated to compare predictive performance for a simple and more complex classification problem. A classifier trained solely on simulated data, whether filtered or raw, can differentiate walking and trotting at approximately the same level of accuracy as a classifier trained on real data.

Models trained on real data generally perform better when presented with the task of recognizing lameness in a horse's gait. This is especially the case for the Support Vector Machine and k-Nearest Neighbours. Nevertheless, classifiers trained on simulated data can achieve classification accuracy which comes close to that of a classifier trained with real data.

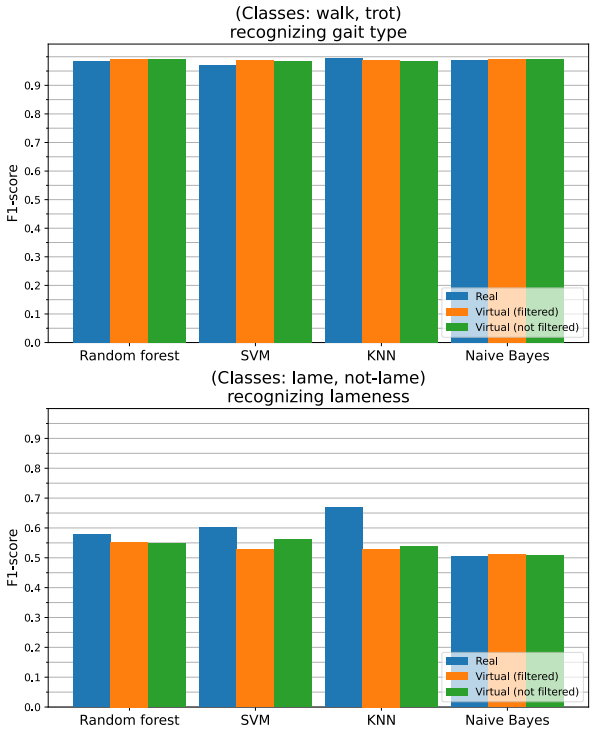
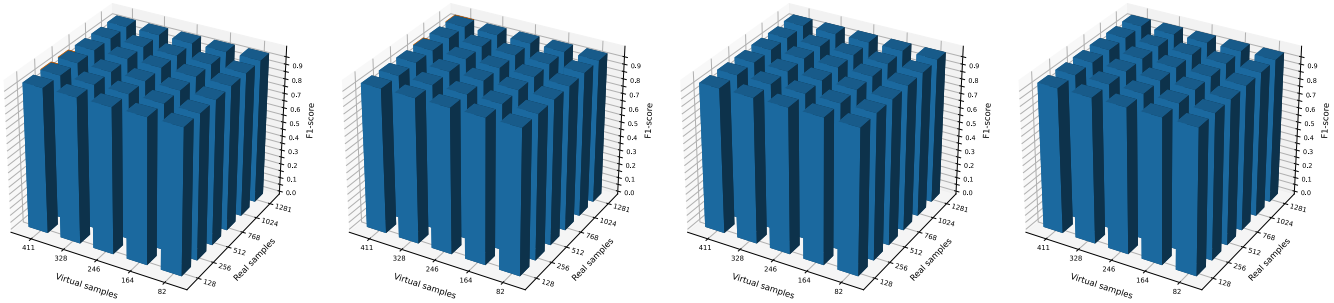


Fig. 16: Comparison of real, raw simulated and filtered simulated data for a simple and complex classification task.

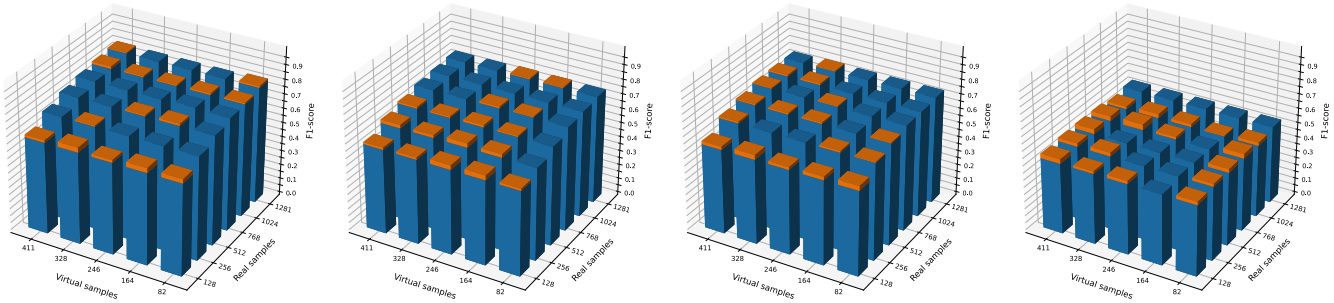
VI. Discussion

The goal of this research was to investigate the potential merit of using simulated IMU data in the context of AAR.



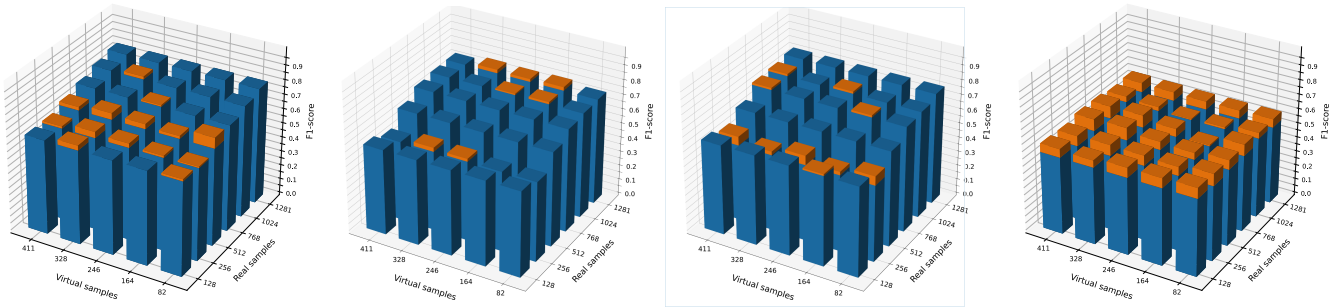
(a) Support Vector Machine (b) Random Forest (c) k-Nearest Neighbours (d) Naive Bayes

Fig. 17: Recognizing walking and trotting regardless of lameness without augmentation (blue) and with augmentation (orange)



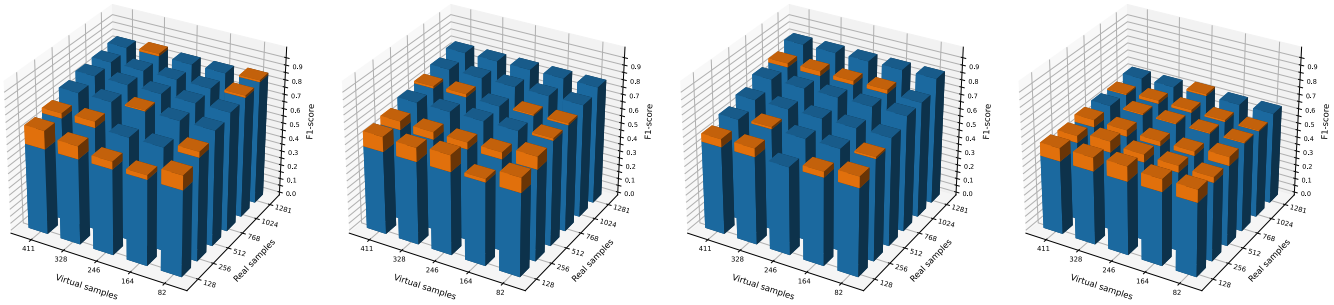
(a) Support Vector Machine (b) Random Forest (c) k-Nearest Neighbours (d) Naive Bayes

Fig. 18: Recognizing lameness regardless of gait type without augmentation (blue) and with augmentation (orange)



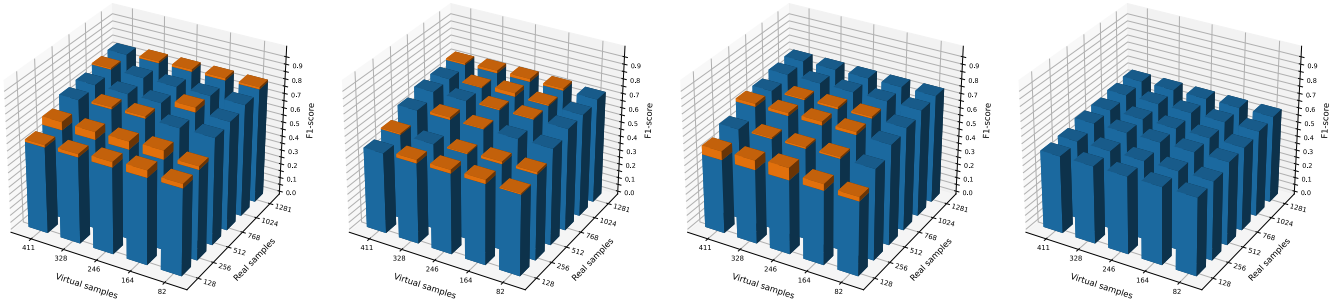
(a) Support Vector Machine (b) Random Forest (c) k-Nearest Neighbours (d) Naive Bayes

Fig. 19: Recognizing lameness during walking without augmentation (blue) and with augmentation (orange)



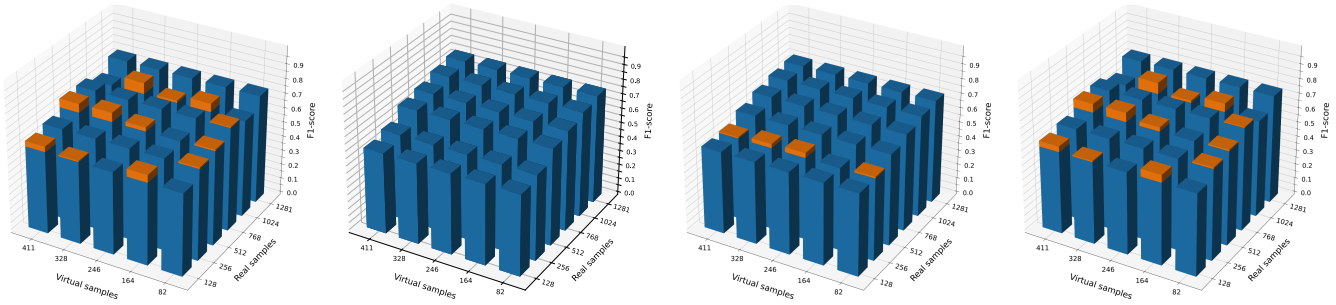
(a) Support Vector Machine (b) Random Forest (c) k-Nearest Neighbours (d) Naive Bayes

Fig. 20: Recognizing lameness during trotting without augmentation (blue) and with augmentation (orange)



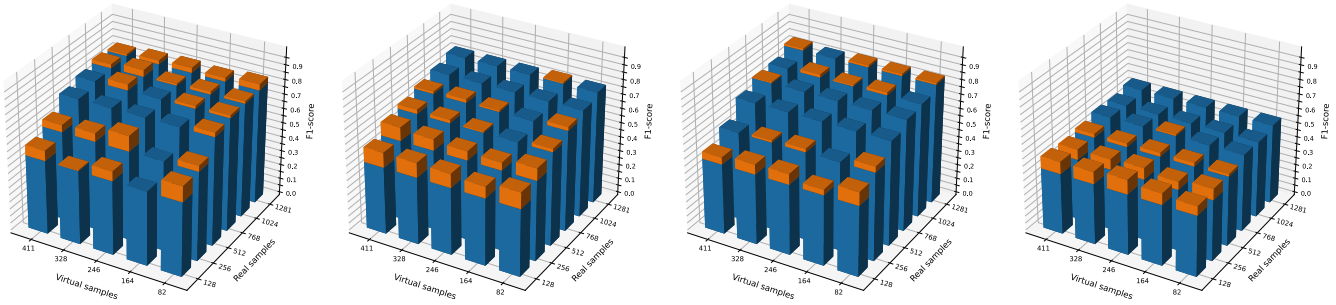
(a) Support Vector Machine (b) Random Forest (c) k-Nearest Neighbours (d) Naive Bayes

Fig. 21: Recognizing lameness in both gaits without augmentation (blue) and with augmentation (orange)



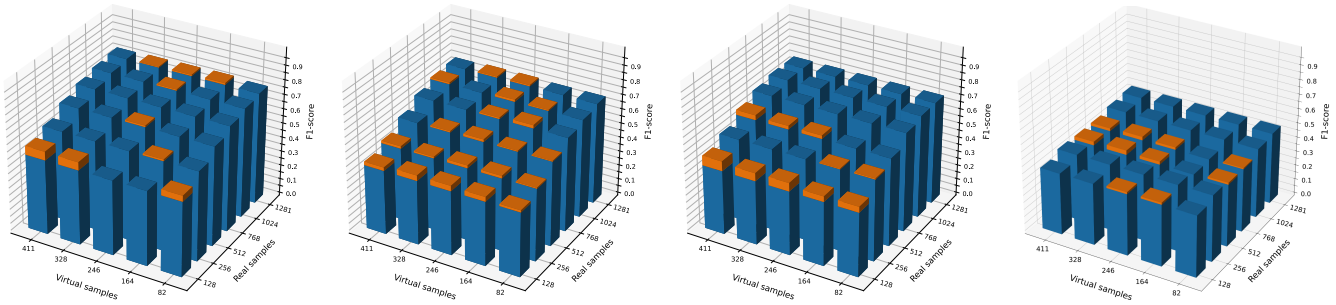
(a) Support Vector Machine (b) Random Forest (c) k-Nearest Neighbours (d) Naive Bayes

Fig. 22: Recognizing the position of lameness during walking without augmentation (blue) and with augmentation (orange)



(a) Support Vector Machine (b) Random Forest (c) k-Nearest Neighbours (d) Naive Bayes

Fig. 23: Recognizing the position of lameness during trotting without exp1 (blue) and with augmentation (orange)



(a) Support Vector Machine (b) Random Forest (c) k-Nearest Neighbours (d) Naive Bayes

Fig. 24: Recognizing the position of lameness in both gaits without augmentation (blue) and with augmentation (orange)

The varying results that were achieved during the evaluation leave room for debate whether this simulated data can actually be beneficial in general. Large improvements in the predictive performance of several classifiers is observed when a small amount of real data is augmented with said simulated data, yet there are also observations in which a decrease in accuracy on our evaluation data is seen. Furthermore classifiers trained using solely real data tend to outperform those which are trained using only simulated data. Potential causes for this are discrepancies between the real and simulated data and a higher noise-to-signal ratio in our simulated data. The increases in predictive performance were generally greater than the observed decreases. Thus, to provide an answer to the previously posed research question: there are certain cases in which using simulated data for activity recognition problems is beneficial, especially when little real data is available for training.

Interestingly, correlation between real and simulated data does not seem to have a significant effect on model accuracy. The hypothesis that greater signal resemblance between real and simulated data will yield better predictions does therefore not hold as there is no evidence to support this. In fact, classifiers trained using unfiltered data sometimes make slightly better predictions than their filtered equivalent. One theory behind these observations is that the used features are robust against signal noise. An experiment that could yield more conclusive results would be to train classifiers using the raw signals directly rather than its extracted features.

F1 scores obtained during the second experiment in section 5 are noticeably lower when comparing them with scores obtained with a similar or smaller training set during the first experiment. The uneven distribution of the simulated dataset could be a possible explanation for this. As indicated in table I, some subjects are not or barely represented in the simulated data whereas the set used for evaluation contains data of each subject. A very challenging aspect of activity recognition problems is that movement patterns can be very subject specific, meaning that the classifiers fail to generalize to other subjects if there is a large enough difference in movement patterns. This is also one of the main reasons why a lot of data is needed to train robust predictive models for activity recognition.

VII. Conclusion

In this paper, the viability of using simulated IMU data for the recognition of animal activity has been validated. Acceleration and angular rate streams were derived from optical motion capture data of horses. The resulting simulated IMU data was compared with real IMU data in terms of signal characteristics as well as the resulting recognition accuracy when used to train animal activity classifiers. Support Vector Machine, Random Forest, k-Nearest Neighbours and Naive Bayes were trained using real, simulated and a combination of both types of data. Classifiers trained using only simulated data generally show inferior predictive performance compared to models trained with real data. Nonetheless, competitive recognition accuracy was achieved during the evaluation of Random Forest, Support Vector Machine and Naive Bayes.

Augmenting real data with simulated data yields promising results as average increases in F1-score are observed throughout various ratios of real to simulated data. Classifiers trained using augmented data can increase F1-score by more than 20% when a small amount of real data is available for training. Decreases in model performance are also observed after augmentation, which is likely caused by a degree of error between real and simulated data.

References

[1] C. Arcidiacono, S. M. Porto, M. Mancino, and G. Cascone. Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data. *Computers and Electronics in Agriculture*, 134:124–134, 3 2017. ISSN 01681699. doi: 10.1016/j.compag.2017.01.021.

[2] J. Banzi. A sensor based anti-poaching system in tanzania national parks. *International Journal of Scientific & Technology Research*, Volume 4, 04 2014.

[3] S. Bosch et al. Equimoves: A wireless networked inertial measurement system for objective examination of horse gait. *Sensors*, 18(3), 2018. ISSN 1424-8220. doi: 10.3390/s18030850. URL https://www.mdpi.com/1424-8220/18/3/850.

[4] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *CoRR*, abs/1812.08008, 2018. URL http://arxiv.org/abs/1812.08008.

[5] J. Y. Chang, G. Moon, and K. M. Lee. V2v-posenet: Voxel-to-voxel prediction network for accurate 3d hand and human pose estimation from a single depth map. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5079–5088, 2018. doi: 10.1109/CVPR.2018.00533.

[6] P. Dario, C. Feichtenhofer, D. Grangier, and M. Auli. 3d human pose estimation in video with temporal convolutions and semi-supervised training. *CoRR*, abs/1811.11742, 2018. URL http://arxiv.org/abs/1811.11742.

[7] R. Dutta et al. Dynamic cattle behavioural classification using supervised ensemble classifiers. *Computers and Electronics in Agriculture*, 111:18–28, 2015. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2014.12.002. URL https://www.sciencedirect.com/science/article/pii/S0168169914003123.

[8] V. Fortes Rey, K. K. Garewal, and P. Lukowicz. Translating videos into synthetic training data for wearable sensor-based activity recognition systems using residual deep convolutional networks. volume 11, 2021. doi: 10.3390/app11073094. URL https://www.mdpi.com/2076-3417/11/7/3094.

[9] L. Gerencsér, G. Vásárhelyi, M. Nagy, T. Vicsek, and A. Miklósi. Identification of Behaviour in Freely Moving Dogs (Canis familiaris) Using Inertial Sensors. *PLoS ONE*, 8(10), 10 2013. ISSN 19326203. doi: 10.1371/journal.pone.0077814.

[10] Y. Huang et al. Deep inertial poser learning to reconstruct human pose from sparse inertial measurements in real time. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 37(6):185:1–185:15, nov 2018.

[11] C. Jobanputra, J. Bavishi, and N. Doshi. Human activity recognition: A survey. *Procedia Computer Science*, 155:698–703, 2019. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2019.08.100. The 16th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2019).

[12] A. Jukan, X. Masip-Bruin, and N. Amla. Smart computing and sensing technologies for animal welfare: A systematic review. 50(1), Apr. 2017. ISSN 0360-0300. doi: 10.1145/3041960. URL https://doi.org/10.1145/3041960.

[13] J. Kamminga. *Hiding in the Deep: Online Animal Activity Recognition using Motion Sensors and Machine Learning*. PhD thesis, University of Twente, Netherlands, Sept. 2020.

[14] J. Kamminga, M. Jones, K. Seppi, N. Meratnia, and P. Havinga. Synchronization between sensors and cameras in movement data labeling frameworks. 11 2019. ISBN 9781450369930. doi: 10.1145/3359427.3361920.

[15] J. W. Kamminga, H. C. Bisby, D. V. Le, N. Meratnia, and P. J. M. Havinga. Generic online animal activity recognition on collar tags. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, page 597–606, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450351904. doi: 10.1145/3123024.3124407. URL https://doi.org/10.1145/3123024.3124407.

[16] J. W. Kamminga, D. V. Le, J. P. Meijers, H. Bisby, N. Meratnia, and P. J. Havinga. Robust sensor-orientation-independent feature selection for animal activity recognition on collar tags. 2(1), Mar. 2018. doi: 10.1145/3191747. URL https://doi.org/10.1145/3191747.

[17] J. W. Kamminga, M. Nirvana, D. V. Le, and P. J. Havinga. Towards deep unsupervised representation learning from accelerometer time series for animal activity recognition. aug 2020.

[18] N. Kleanthous et al. Machine learning techniques for classification of livestock behavior. In *Neural Information Processing*, pages 304–315, Cham, 2018. Springer International Publishing. ISBN 978-3-030-04212-7.

[19] H. Kwon, C. Tong, H. Haresamudram, Y. Gao, G. Abowd, N. Lane, and T. Ploetz. Imutube: Automatic extraction of virtual on-body accelerometry from video for human activity recognition. 05 2020.

[20] V. Leos-Barajas et al. Analysis of animal accelerometer data using hidden markov models. *Methods in Ecology and Evolution*, 8(2):161–173, 2017. doi: https://doi.org/10.1111/2041-210X.12657.

[21] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6): 248:1–248:16, Oct. 2015.

[22] D. C. Luvizon, D. Picard, and H. Tabia. 2d/3d pose estimation and action recognition using multitask deep learning. *CoRR*, abs/1802.09232, 2018. URL http://arxiv.org/abs/1802.09232.

[23] P. Martiskainen et al. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied Animal Behaviour Science*, 119(1):32–38, 2009. ISSN 0168-1591. doi: https://doi.org/10.1016/j.applanim.2009.03.005.

[24] R. Nathan. An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, 105:19050–1, 01 2009. doi: 10.1073/pnas.0808918105.

[25] T. Nguyen, V. Nguyen, F. Salim, D. Le, and D. Phung. A simultaneous extraction of context and community from pervasive signals using nested dirichlet process. *Pervasive and Mobile Computing*, 38, 08 2016. doi: 10.1016/j.pmcj.2016.08.019.

[26] A. Nibali et al. 3d human pose estimation with 2d marginal heatmaps. *CoRR*, abs/1806.01484, 2018. URL http://arxiv.org/abs/1806.01484.

[27] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis. Coarse-to-fine volumetric prediction for single-image 3d human pose. *CoRR*, abs/1611.07828, 2016. URL http://arxiv.org/abs/1611.07828.

[28] L. Pei et al. MARS: mixed virtual and real wearable sensors for human activity recognition with multi-domain deep learning model. *CoRR*, abs/2009.09404, 2020. URL https://arxiv.org/abs/2009.09404.

[29] Y. Peng et al. Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units. *Computers and Electronics in Agriculture*, 157:247–253, 2019. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2018.12.023. URL https://www.sciencedirect.com/science/article/pii/S016816991830869X.

[30] E. Ramanujam, T. Perumal, and S. Padmavathi. Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. *IEEE Sensors Journal*, pages 1–1, 2021. doi: 10.1109/JSEN.2021.3069927.

[31] Y. Sahin. Animals as mobile biological sensors for forest fire detection. *Sensors*, 7:3084–3096, 12 2007. doi: 10.3390/s7123084.

[32] A. Schmutz, L. Chèze, J. Jacques, and P. Martin. A method to estimate horse speed per stride from one imu with a machine learning method. *Sensors*, 20(2), 2020. ISSN 1424-8220. doi: 10.3390/s20020518. URL https://www.mdpi.com/1424-8220/20/2/518.

[33] F. M. Serra Bragança et al. Improving gait classification in horses by using inertial measurement unit (IMU) generated data and machine learning. *Scientific Reports*, 10(1), 12 2020. ISSN 20452322. doi: 10.1038/s41598-020-73215-9.

[34] S. Takeda, T. Okita, P. Lago, and S. Inoue. A multi-sensor setting activity recognition simulation tool. In *Proceedings*

of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, page 1444–1448, 2018. ISBN 9781450359665. doi: 10.1145/3267305.3267509. URL https://doi.org/10.1145/3267305.3267509.

[35] T. von Marcard, B. Rosenhahn, M. J. Black, and G. Pons-Moll. Sparse inertial poser: Automatic 3d human pose estimation from sparse imus. *CoRR*, abs/1703.08014, 2017. URL http://arxiv.org/abs/1703.08014.

[36] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec.2018.02.010. URL https://www.sciencedirect.com/science/article/pii/S016786551830045X. Deep Learning for Pattern Recognition.

[37] P. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967. doi: 10.1109/TAU.1967.1161901.

[38] C. C. Wilmers et al. The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. Technical Report 7, 2015.

I. Appendix A: signal evaluation tables

metric	sensor	Acceleration			Angular rate		
		X	Y	Z	X	Y	Z
Mean	IMU	10.86	-1.92	-0.7	-0.02	0.16	0.13
	OMC	11.02	-1.85	-0.55	-0.04	0.15	0.12
Std-dev	IMU	9.18	11.93	4.82	1.57	0.39	2.46
	OMC	8.32	10.79	5.31	1.39	0.45	2.31
Range	IMU	(-20.8, 59.6)	(-49.9, 39.2)	(-25.8, 40.5)	(-9.1, 4.8)	(-1.6, 2.1)	(-7.5, 4.5)
	OMC	(-16.4, 46.9)	(-40.0, 37.5)	(-27.4, 30.5)	(-6.0, 4.9)	(-1.7, 1.9)	(-6.8, 3.9)
T-correlation		0.85	0.86	0.5	0.88	0.76	0.98
F-correlation		0.99	0.99	0.86	0.99	0.96	0.99

Fig. 1: Signal evaluation results for activity class: walk

metric	sensor	Acceleration			Angular rate		
		X	Y	Z	X	Y	Z
Mean	IMU	10.93	-1.72	-0.89	-0.02	0.15	0.12
	OMC	10.86	-2.3	-0.75	-0.04	0.17	0.07
Std-dev	IMU	9.97	13.19	4.80	1.66	0.42	2.55
	OMC	9.87	13.22	8.22	1.66	0.52	2.4
Range	IMU	(-13.5, 55.1)	(-71.9, 44.3)	(-30.8, 34.3)	(-8.0, 5.14)	(-1.3, 1.7)	(-7.4, 4.4)
	OMC	(-20.9, 54.6)	(-61.5, 53.4)	(-45.1, 53.5)	(-6.9, 5.6)	(-2.2, 2.7)	(-7.0, 4.3)
T-correlation		0.79	0.76	0.32	0.87	0.66	0.95
F-correlation		0.98	0.98	0.75	0.99	0.88	0.99

Fig. 2: Signal evaluation results for activity class: walk-lame-front

metric	sensor	Acceleration			Angular rate		
		X	Y	Z	X	Y	Z
Mean	IMU	10.38	-1.79	-1.00	-0.28	0.29	-0.06
	OMC	-6.44	-0.87	-9.39	-0.01	-0.23	0.16
Std-dev	IMU	11.1	16.32	7.29	1.91	0.77	2.62
	OMC	9.5	10.11	7.64	1.38	1.31	1.93
Range	IMU	(-34.6, 83.9)	(-134.5, 81.7)	(-35.7, 89.1)	(-8.0, 8.5)	(-2.4, 3.2)	(-9.7, 7.7)
	OMC	(-6.4, 34.2)	(-61.5, 53.4)	(-36.4, 14.3)	(-3.1, 3.7)	(-3.0, 3.7)	(-6.5, 6.3)
T-correlation		0.02	-0.02	0.04	0.03	-0.07	0.05
F-correlation		0.94	0.76	0.83	0.77	0.87	0.98

Fig. 3: Signal evaluation results for activity class: walk-lame-hind

metric	sensor	Acceleration			Angular rate		
		X	Y	Z	X	Y	Z
Mean	IMU	11.46	-2.74	-1.69	0.09	0.30	-0.05
	OMC	11.38	-2.34	-1.68	0.11	0.33	-0.04
Std-dev	IMU	17.39	25.40	9.07	2.62	0.73	4.20
	OMC	15.10	20.64	9.9	2.34	0.78	3.84
Range	IMU	(-48.2, 76.9)	(-121.5, 81.6)	(-56.5, 87.8)	(-7.8, 10.7)	(-2.5, 2.5)	(-9.8, 9.8)
	OMC	(-39.9, 77.4)	(-80.1, 65.3)	(-61.0, 45.9)	(-7.1, 7.9)	(-2.6, 3.0)	(-8.6, 6.5)
T-correlation		0.81	0.80	0.39	0.88	0.75	0.98
F-correlation		0.99	0.98	0.74	0.99	0.98	0.99

Fig. 4: Signal evaluation results for activity class: trot

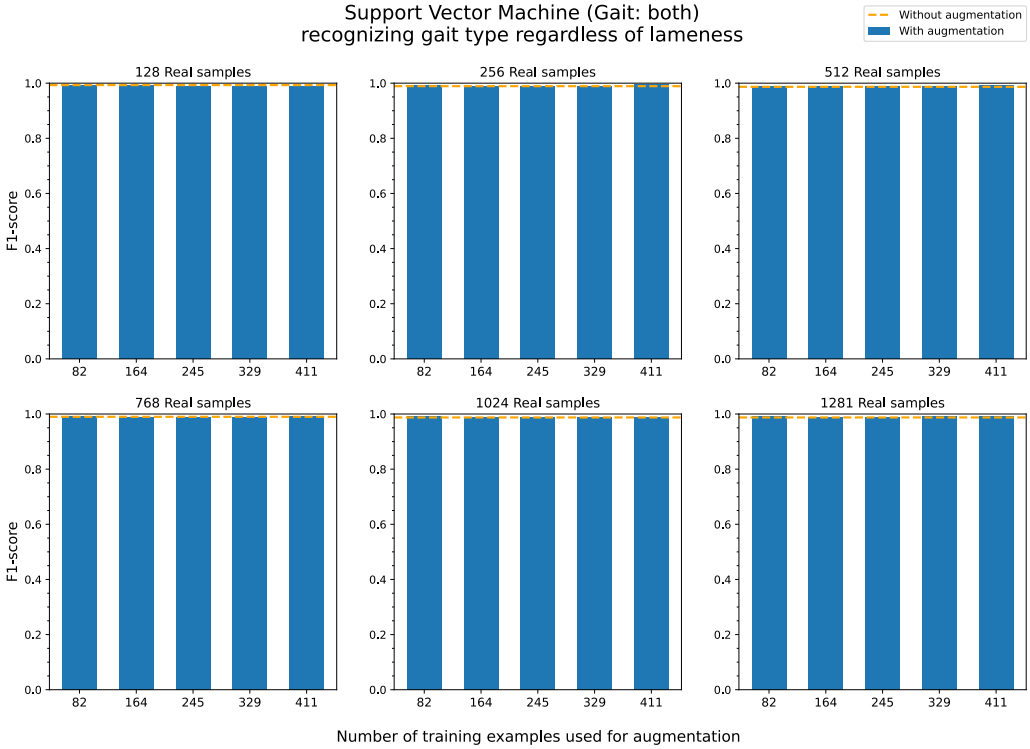
metric	sensor	Acceleration			Angular rate		
		X	Y	Z	X	Y	Z
Mean	IMU	11.68	-2.38	-2.03	0.24	0.32	-0.18
	OMC	11.83	-1.43	-2.63	0.22	0.33	-0.16
Std-dev	IMU	18.59	28.99	8.00	2.59	0.68	4.38
	OMC	16.02	22.15	8.98	2.33	0.73	3.98
Range	IMU	(-49.8, 70.5)	(-151.2, 85.4)	(-42.9, 55.6)	(-6.7, 9.6)	(-1.9, 2.4)	(-10.0, 9.3)
	OMC	(-40.5, 58.1)	(-68.6, 67.1)	(-40.4, 51.9)	(-6.3, 7.8)	(-2.7, 3.3)	(-9.3, 6.1)
T-correlation		0.80	0.80	0.35	0.88	0.73	0.97
F-correlation		0.98	0.98	0.74	0.99	0.7	1.00

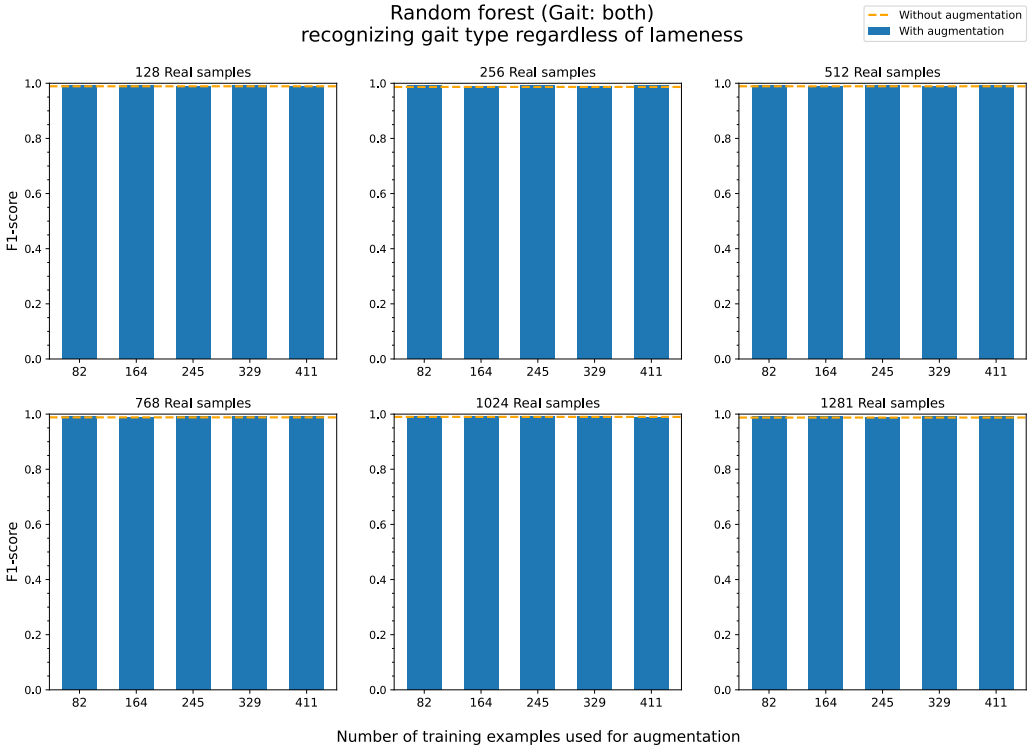
Fig. 5: Signal evaluation results for activity class: walk-lame-front

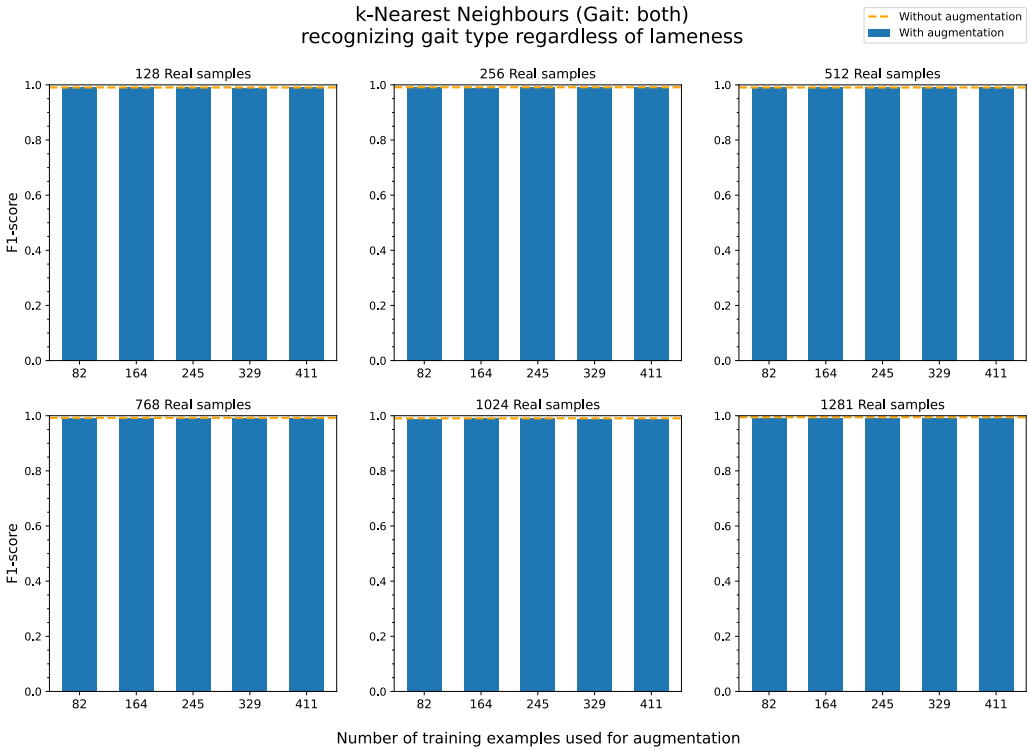
metric	sensor	Acceleration			Angular rate		
		X	Y	Z	X	Y	Z
Mean	IMU	11.78	-3.16	-1.43	0.12	0.27	0.00
	OMC	11.71	-2.87	-1.48	0.2	0.30	0.00
Std-dev	IMU	16.78	26.90	8.54	2.62	0.75	4.09
	OMC	13.26	20.92	9.50	2.32	0.83	3.69
Range	IMU	(-41.6, 62.8)	(-129.6 81.6)	(-50.4, 42.6)	(-7.9, 9.2)	(-2.1, 2.4)	(-8.7, 8.4)
	OMC	(-31.3, 47.9)	(-72.5, 59.1)	(-42.6, 49.8)	(-6.8, 7.5)	(-2.7, 3.6)	(-8.4, 5.9)
T-correlation		0.82	0.88	0.36	0.85	0.7	0.97
F-correlation		0.98	0.96	0.72	0.99	0.98	1.00

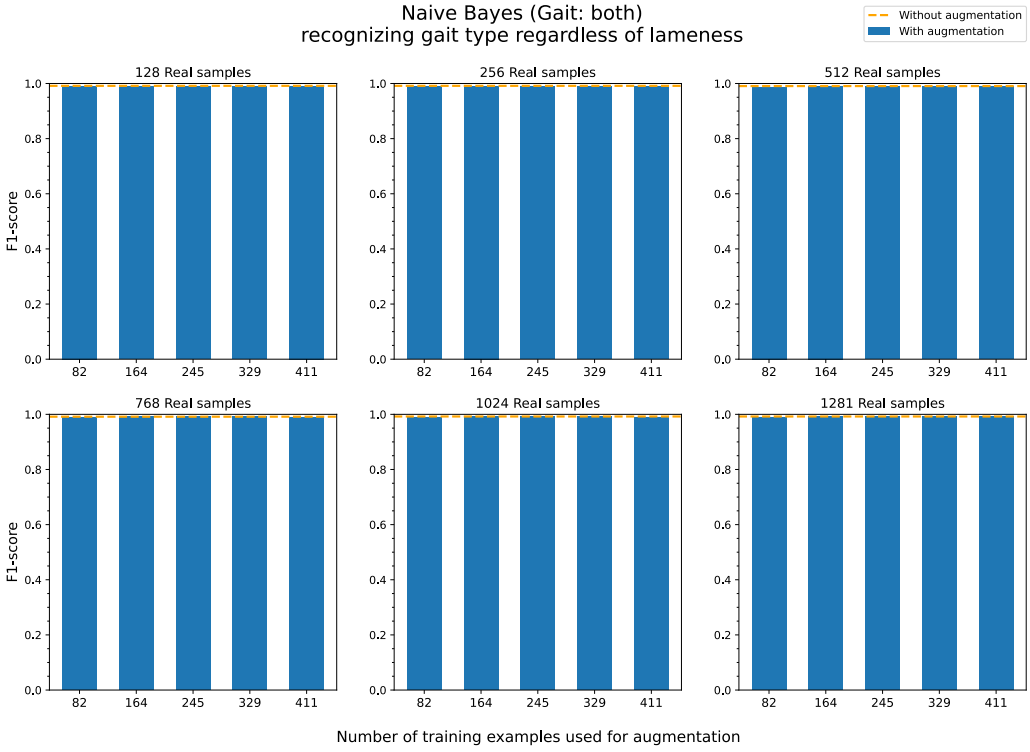
Fig. 6: Signal evaluation results for activity class: walk-lame-hind

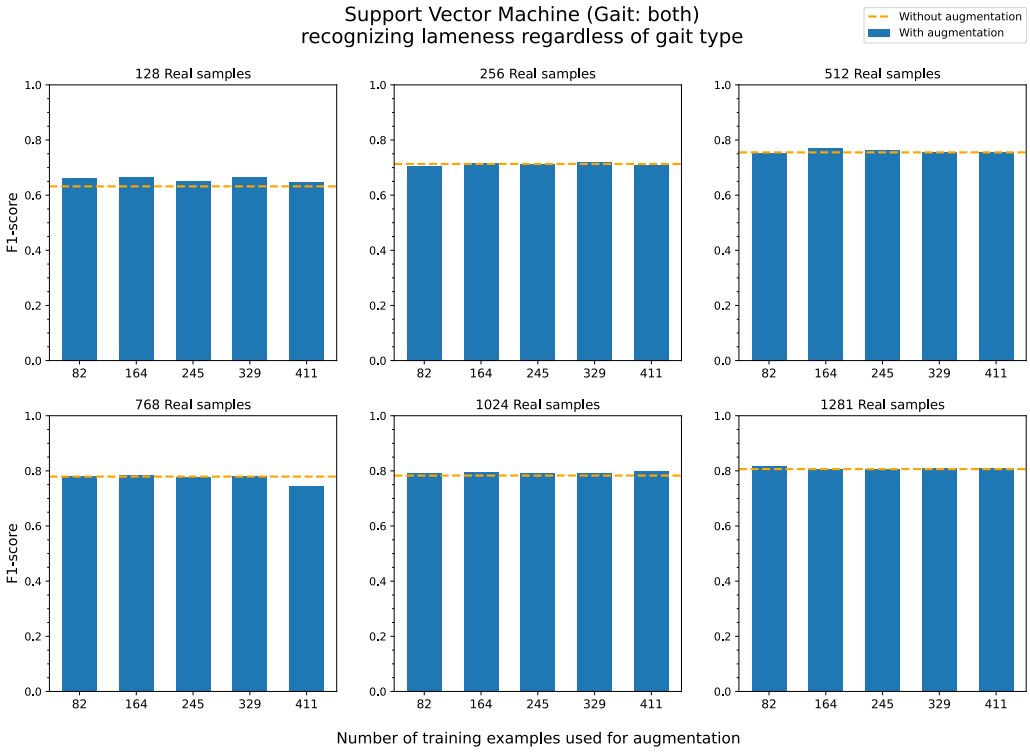
I. Appendix B: visualisations for experiment 1

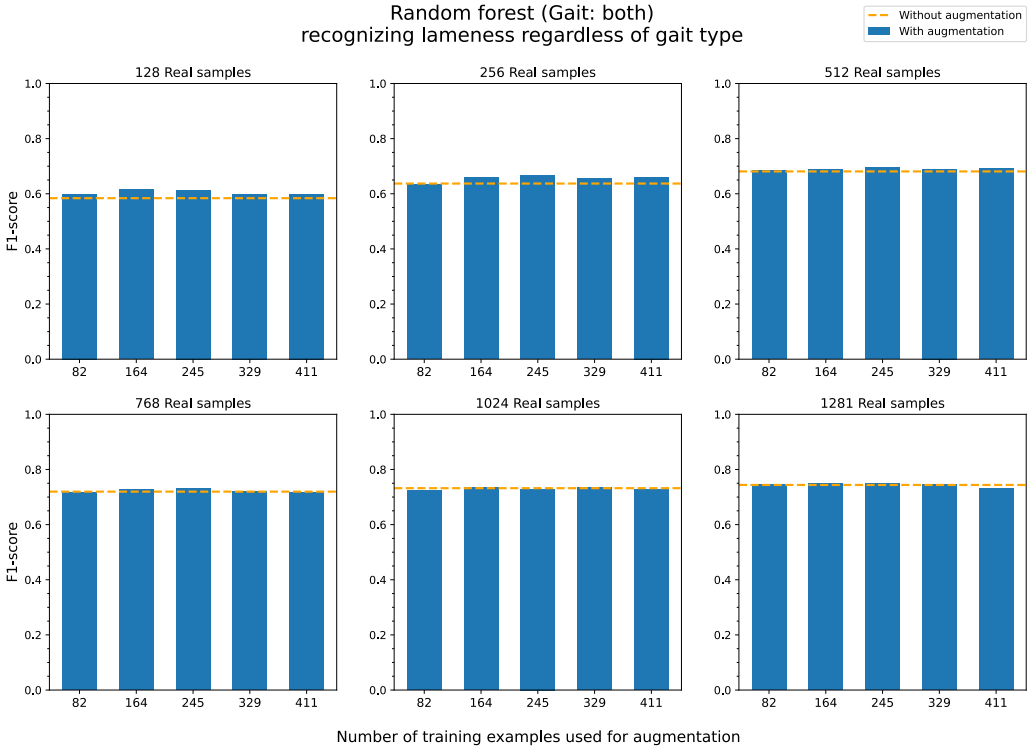


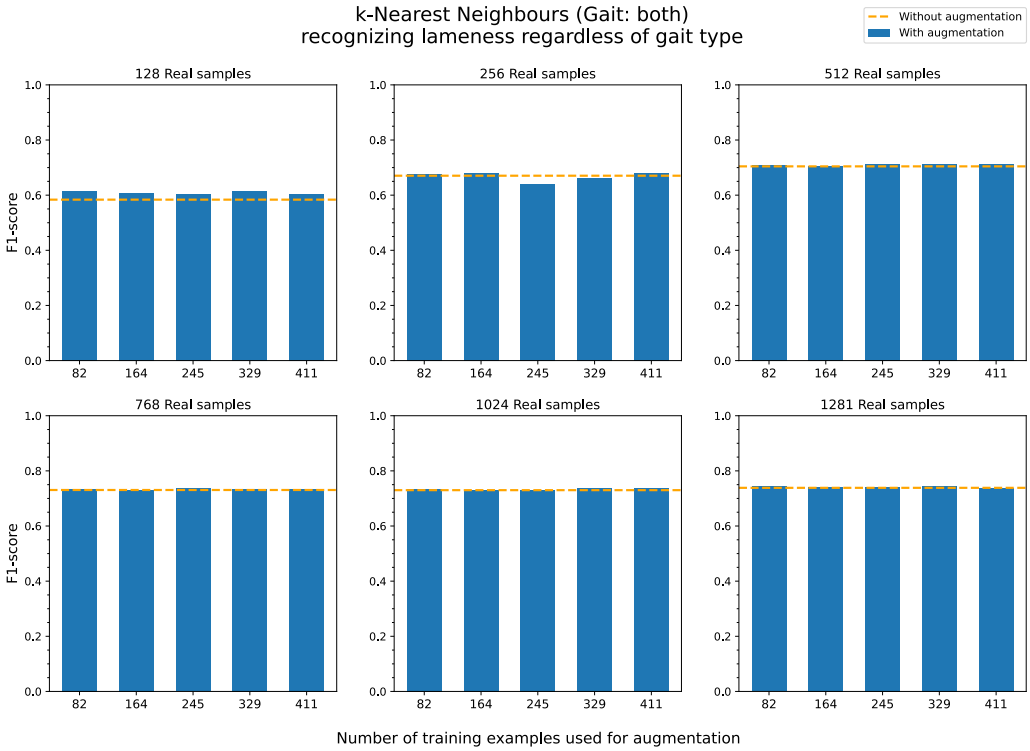


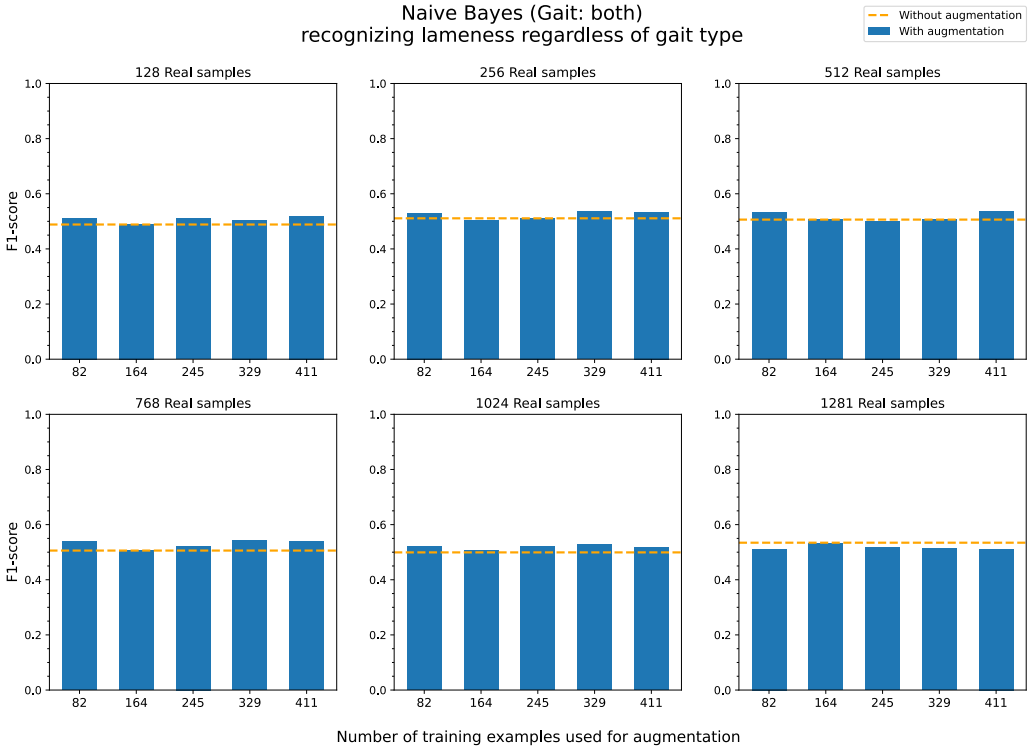

k-Nearest Neighbours (Gait: both)
recognizing gait type regardless of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

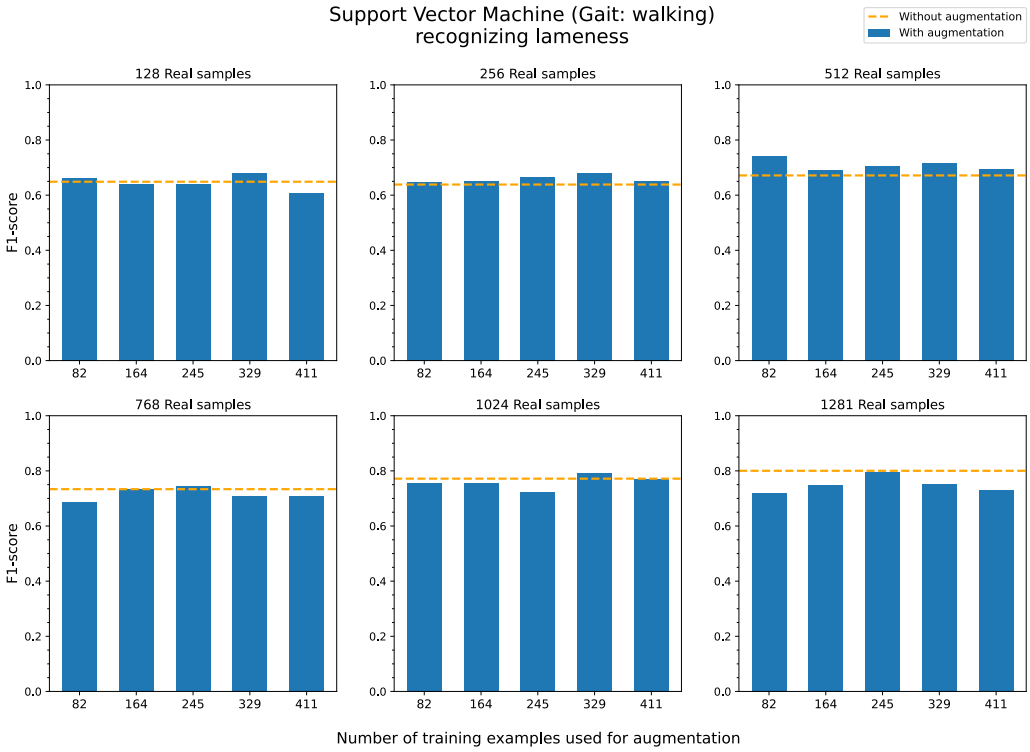

Naive Bayes (Gait: both)
recognizing gait type regardless of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

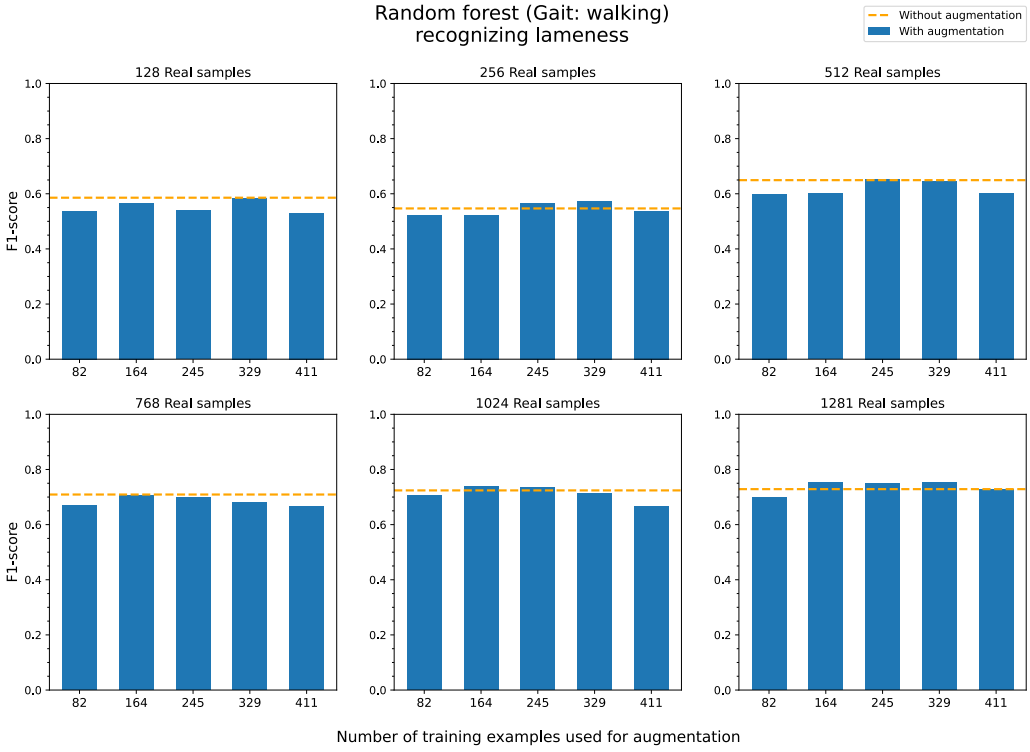

Support Vector Machine (Gait: both)
recognizing lameness regardless of gait type
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

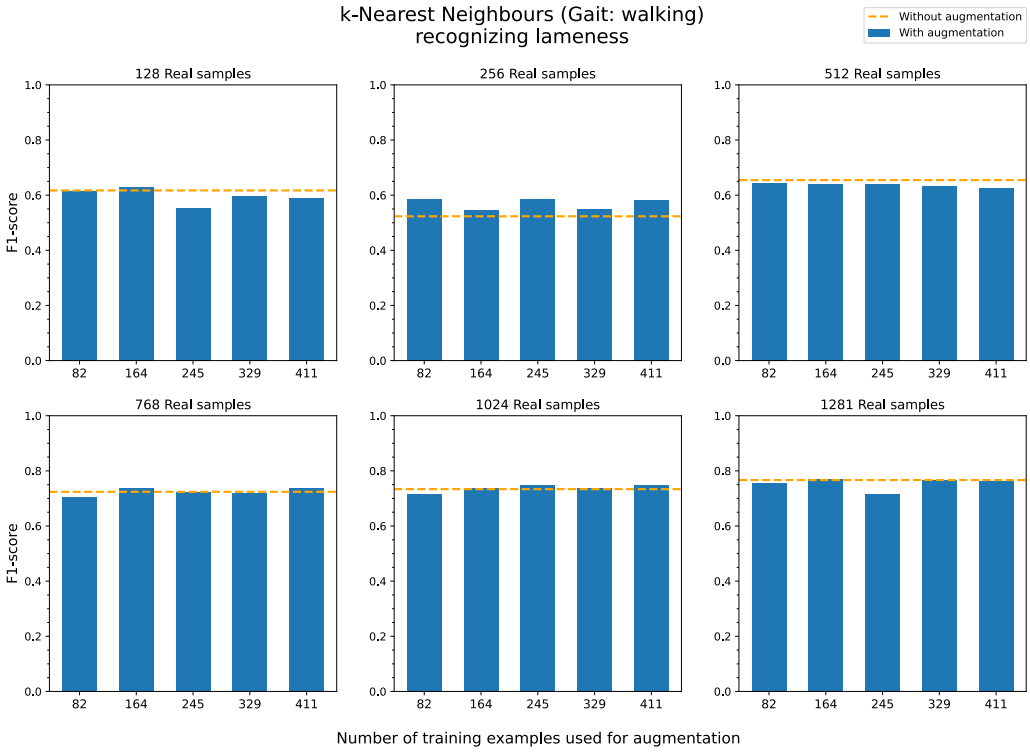

Random forest (Gait: both)
recognizing lameness regardless of gait type
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

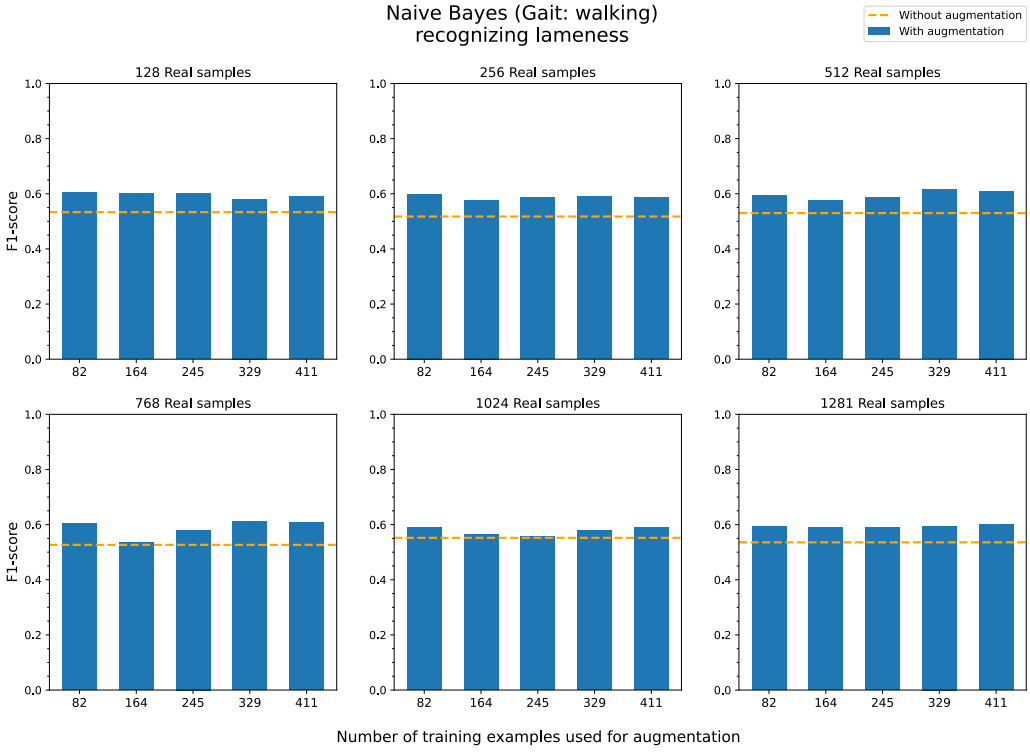

k-Nearest Neighbours (Gait: both)
recognizing lameness regardless of gait type
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

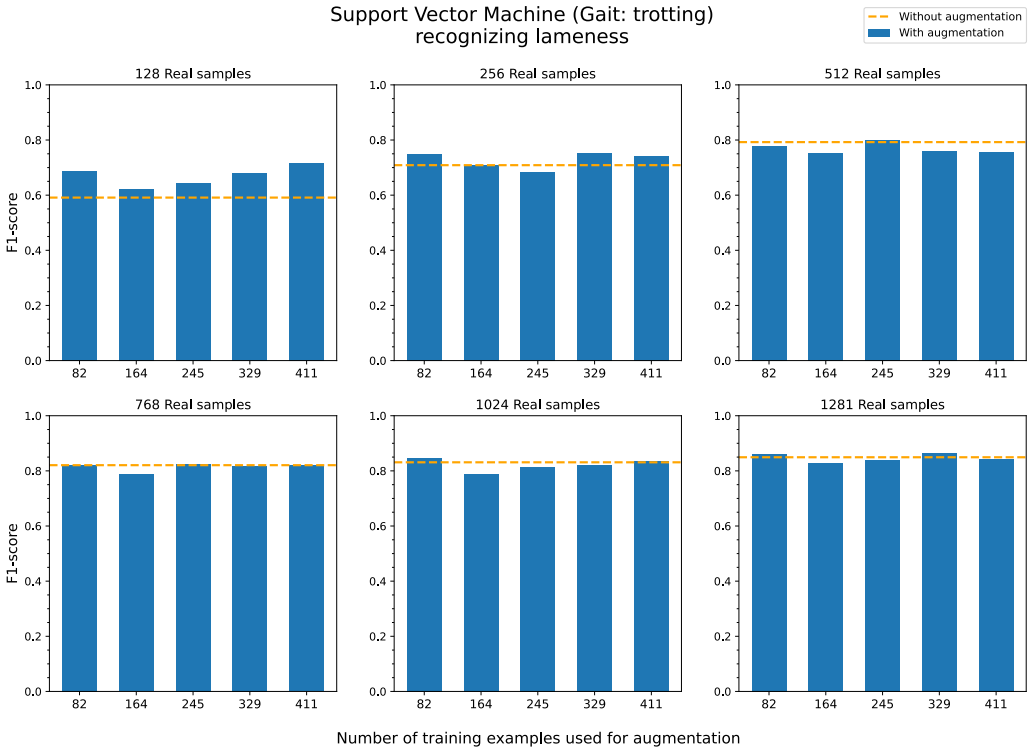

Naive Bayes (Gait: both)
recognizing lameness regardless of gait type
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

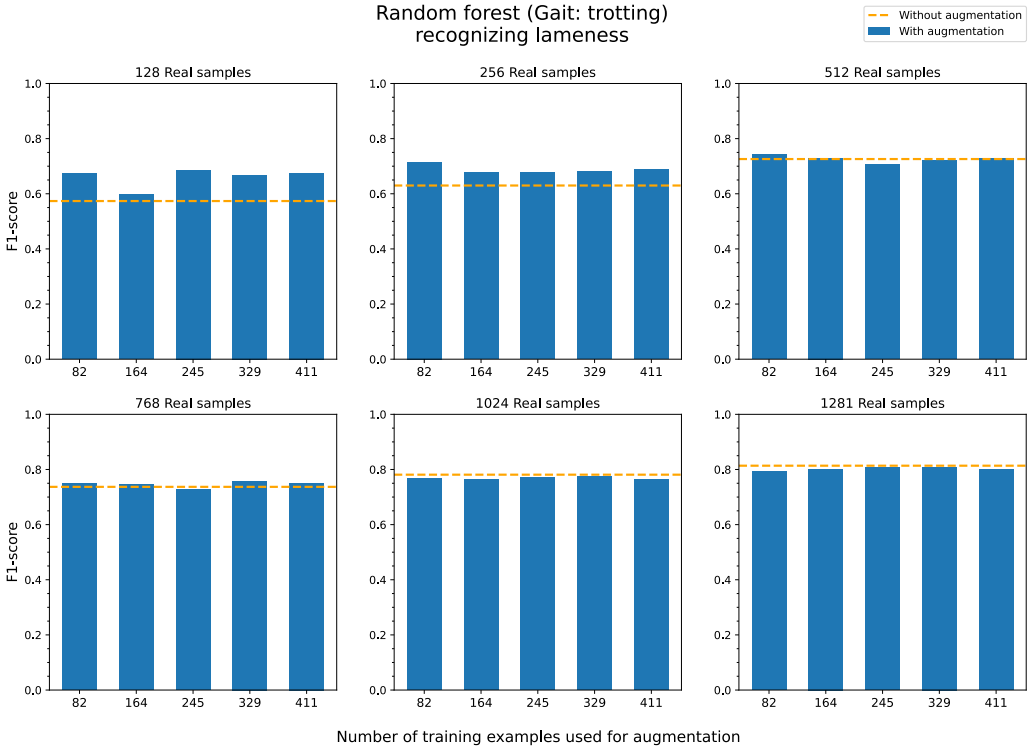

Support Vector Machine (Gait: walking)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

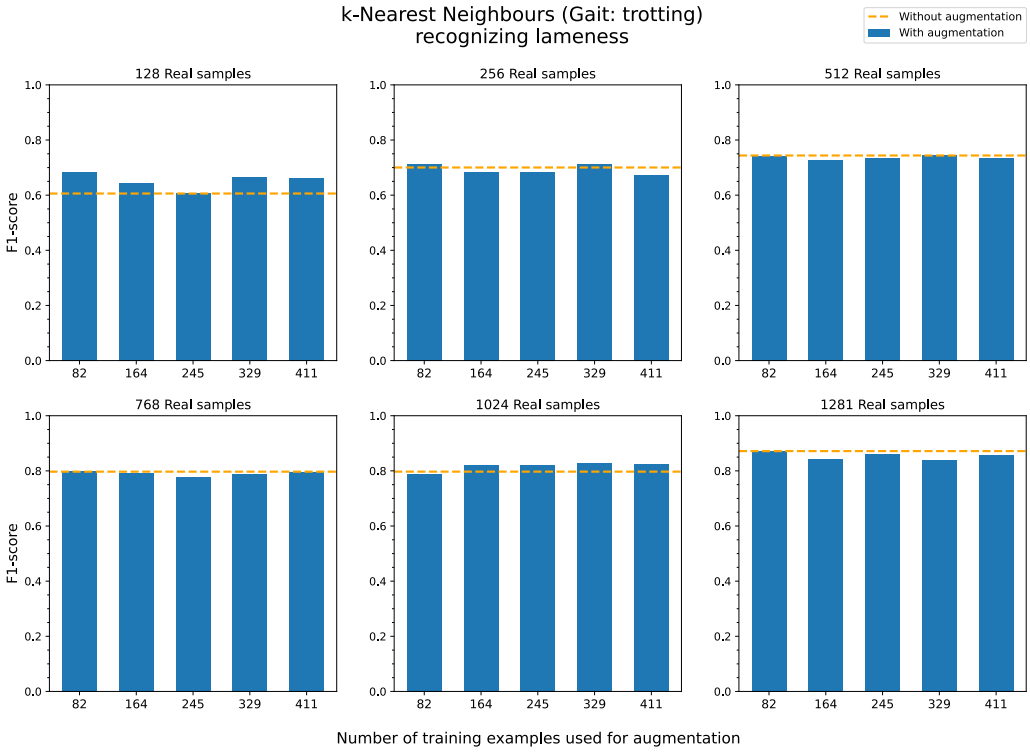

Random forest (Gait: walking)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

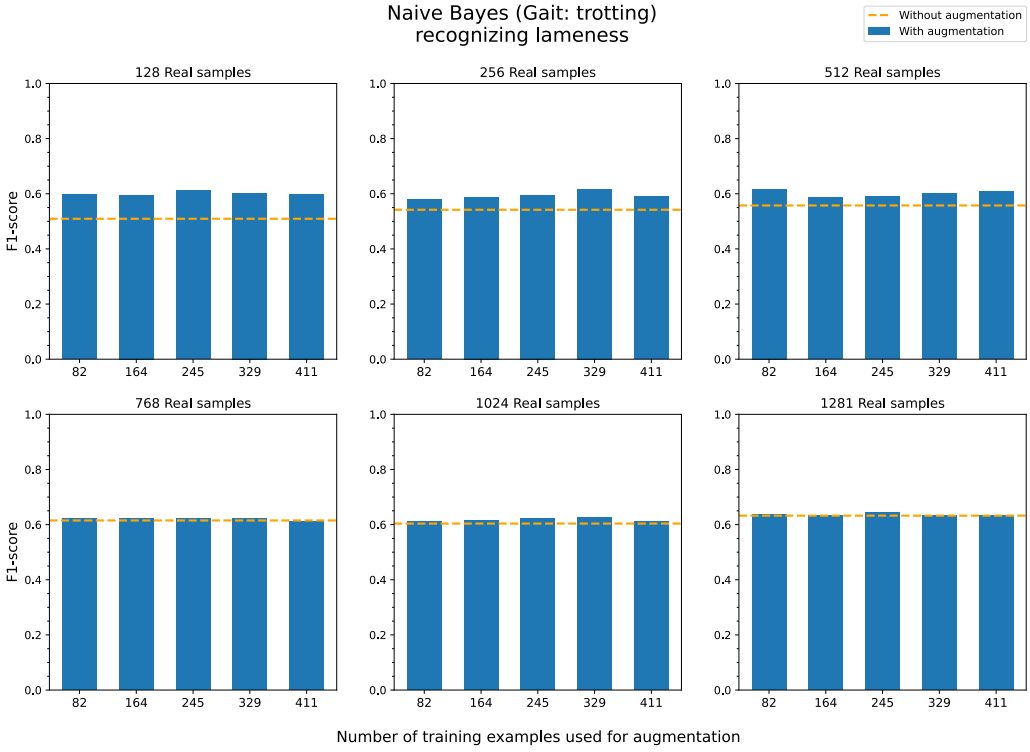

k-Nearest Neighbours (Gait: walking)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

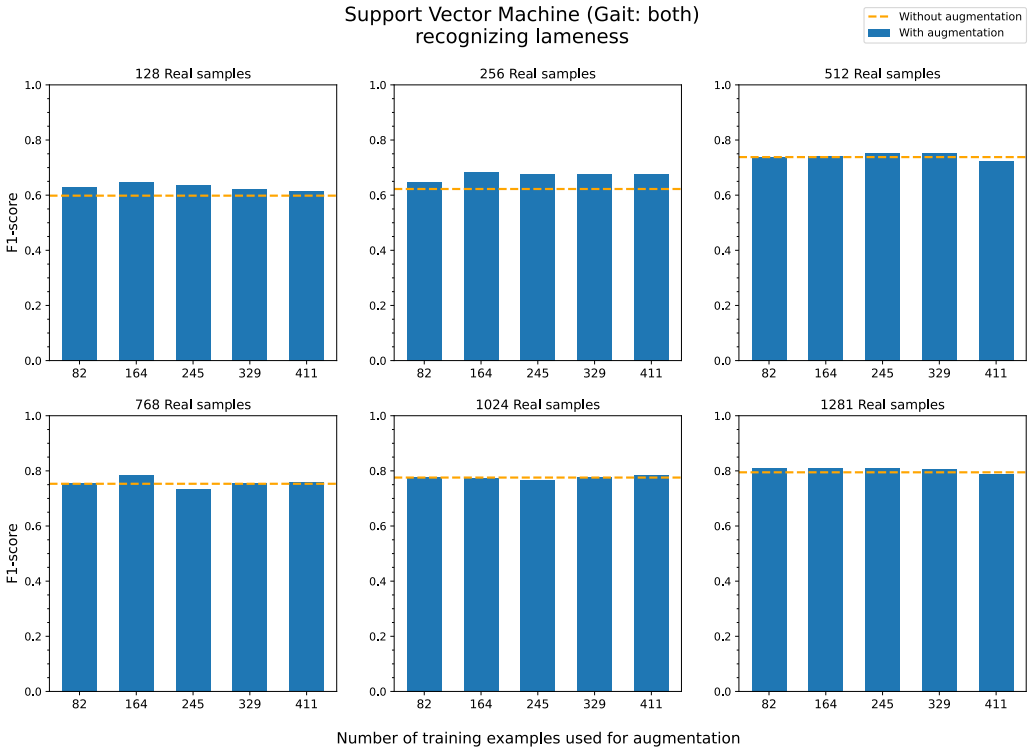

Naive Bayes (Gait: walking)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

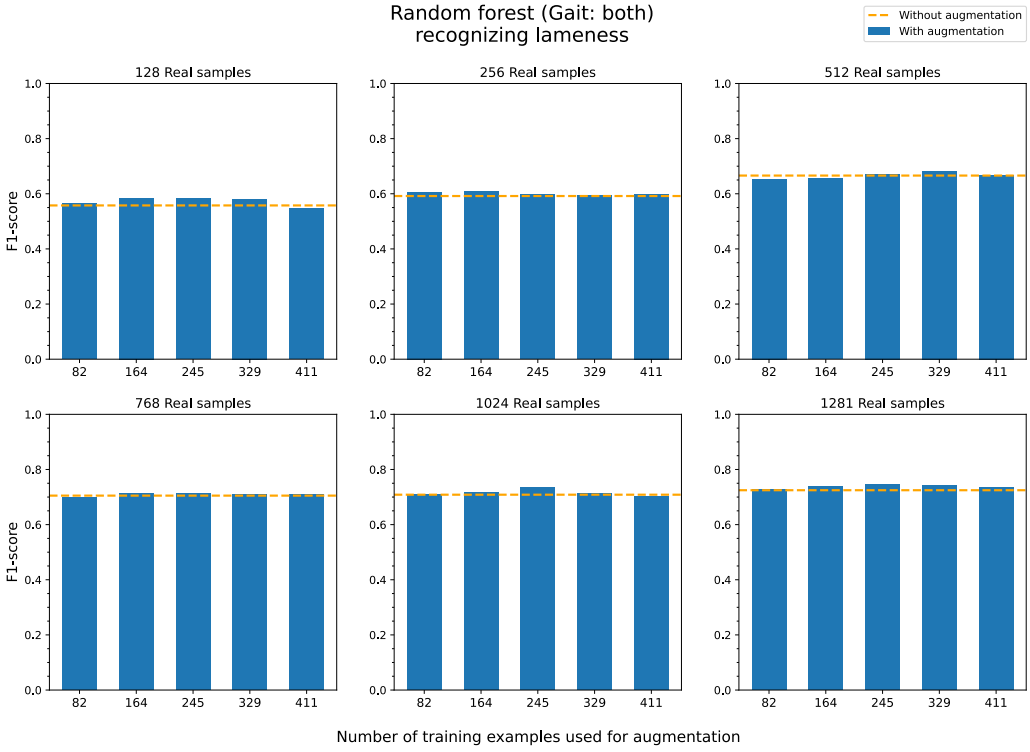

Support Vector Machine (Gait: trotting)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

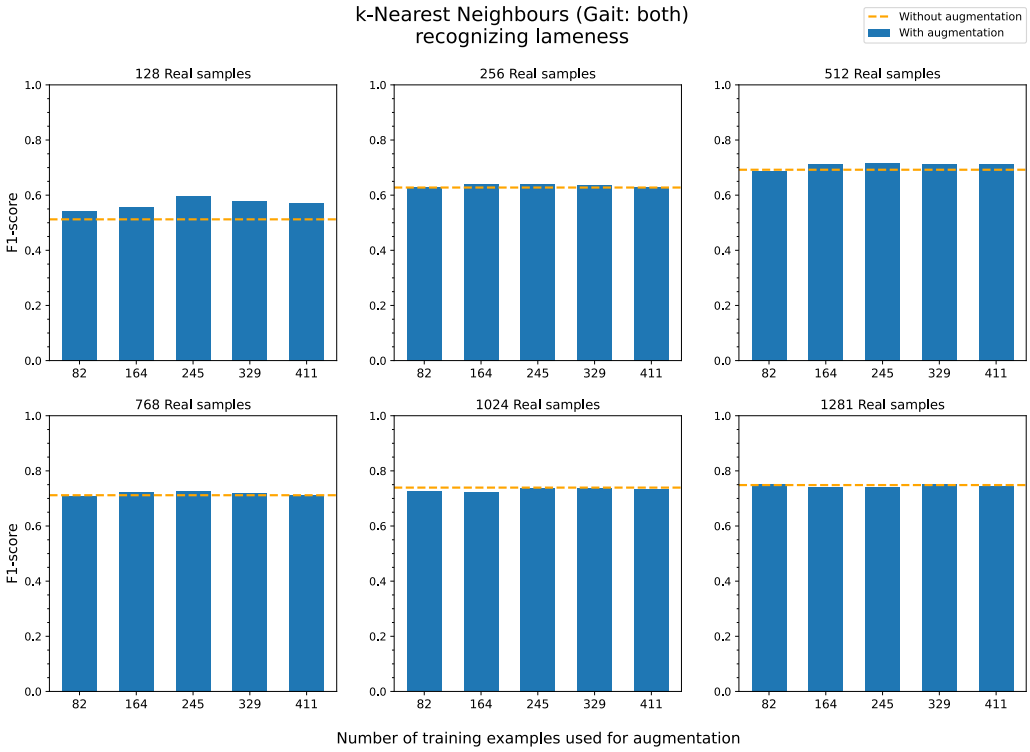

Random forest (Gait: trotting)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

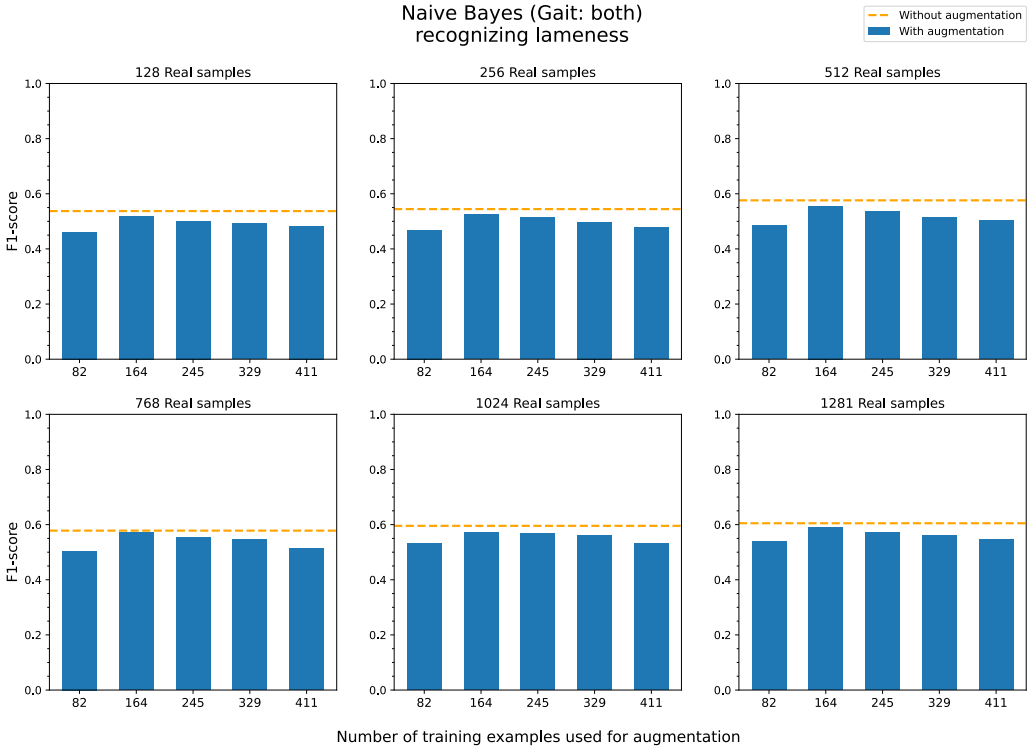

k-Nearest Neighbours (Gait: trotting)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

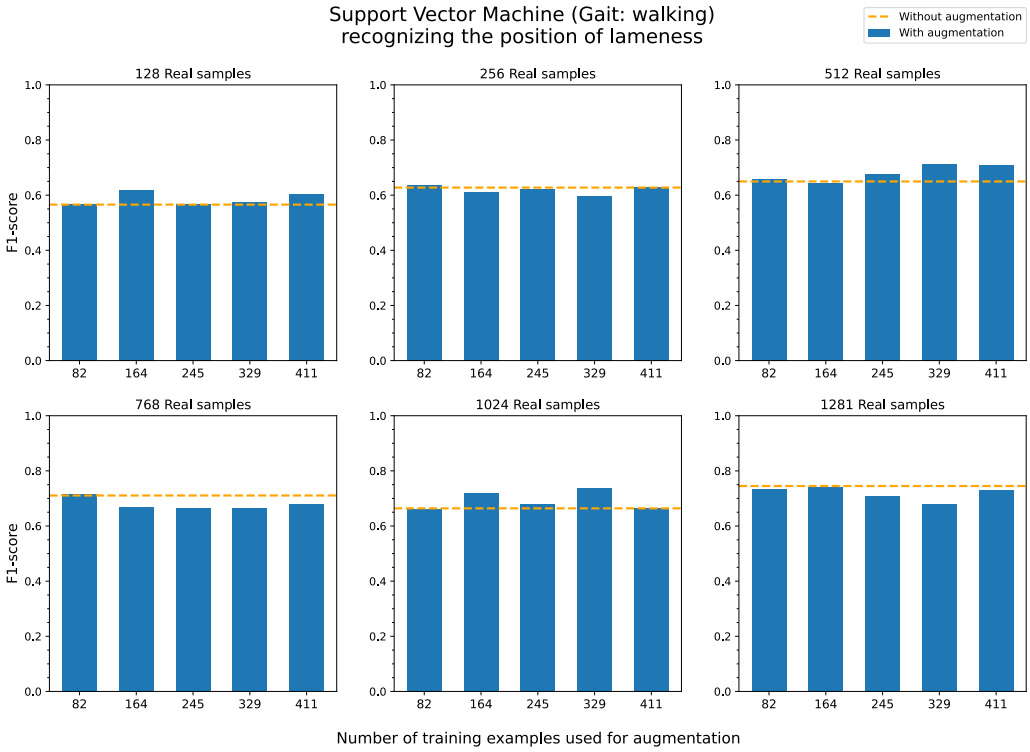

Naive Bayes (Gait: trotting)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

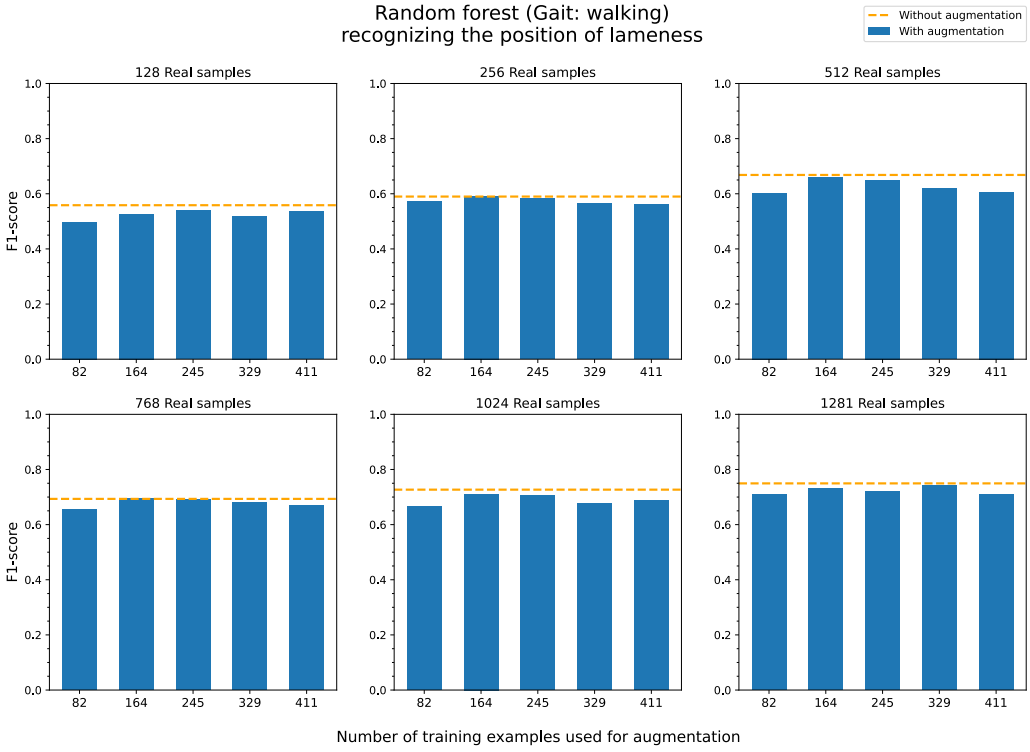

Support Vector Machine (Gait: both)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

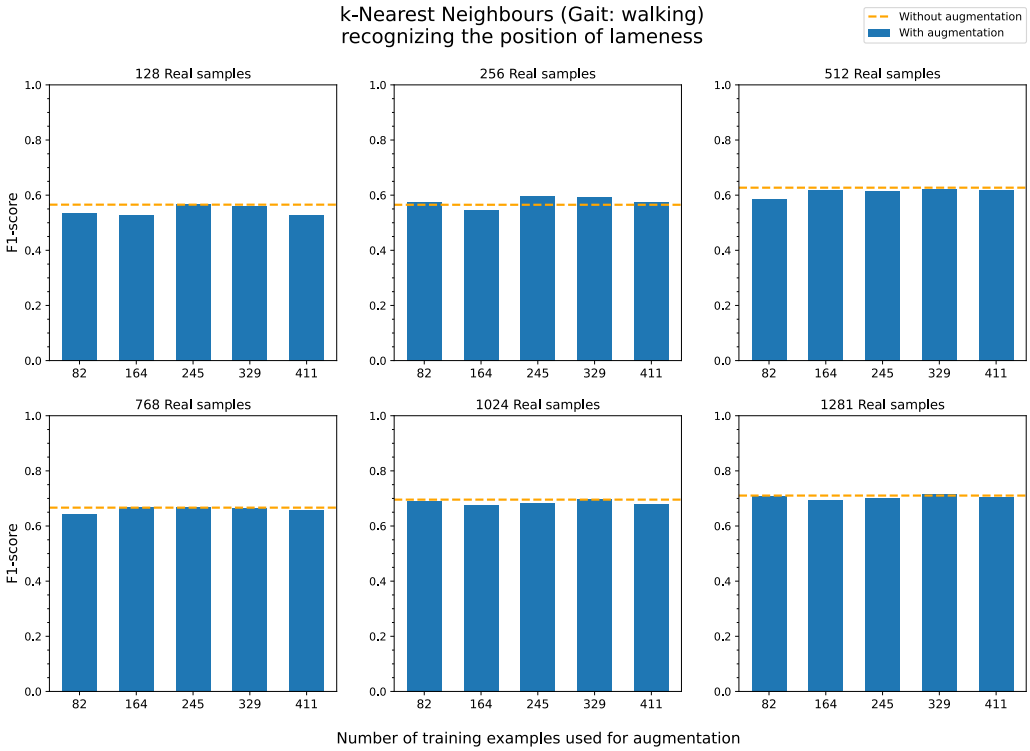

Random forest (Gait: both)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

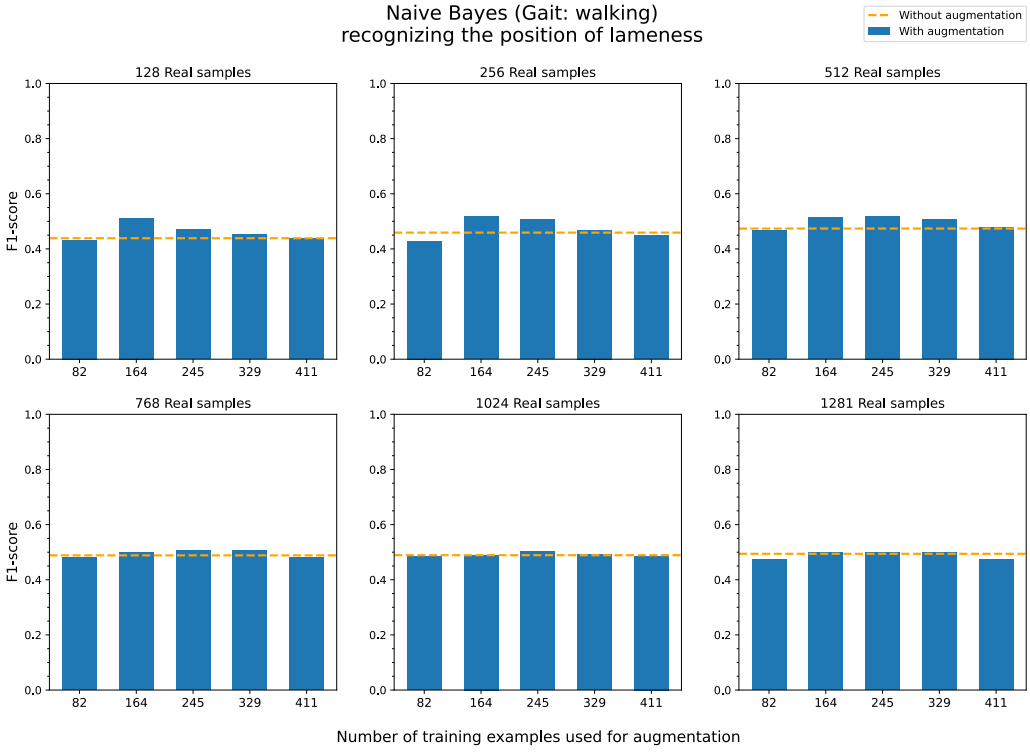

k-Nearest Neighbours (Gait: both)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

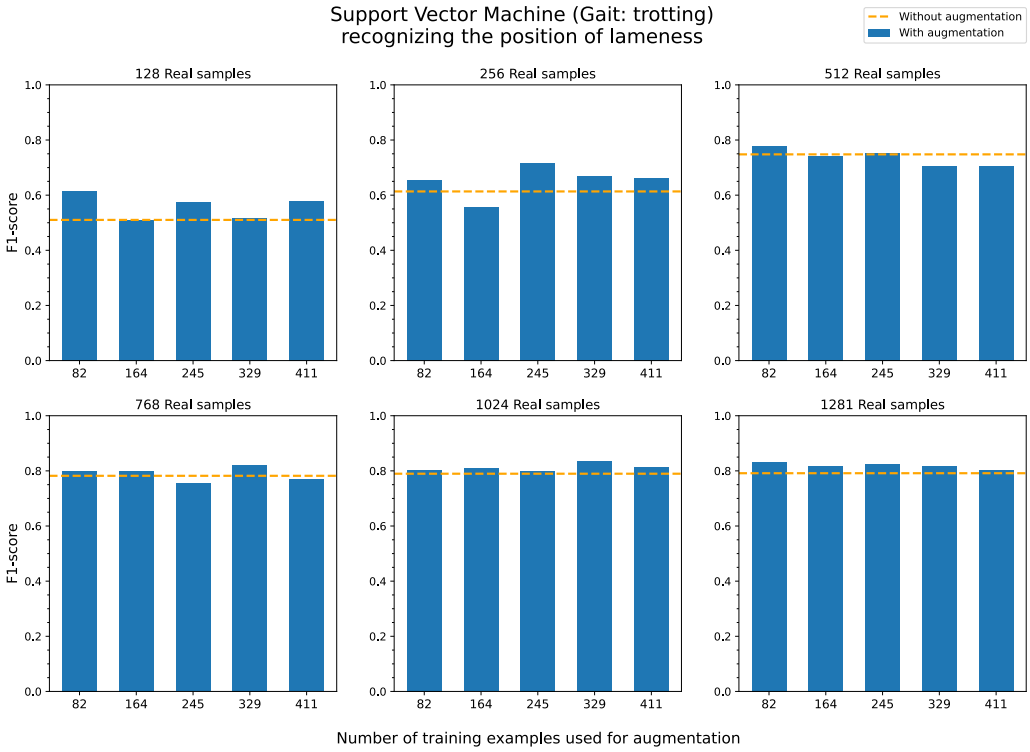

Naive Bayes (Gait: both)
recognizing lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

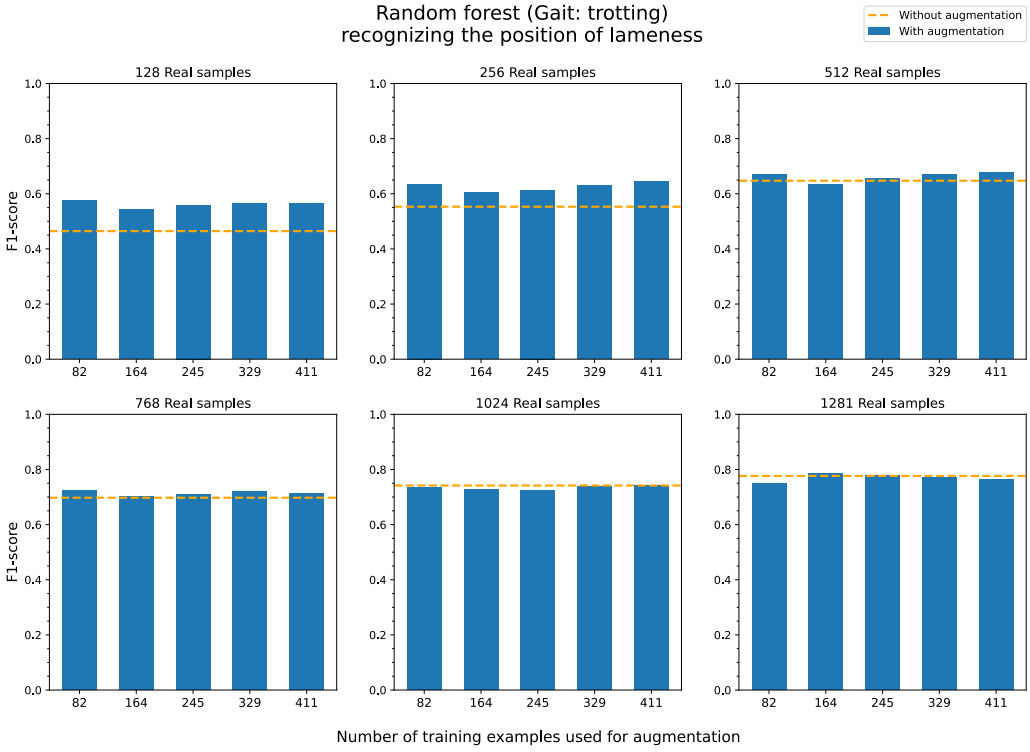

Support Vector Machine (Gait: walking)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

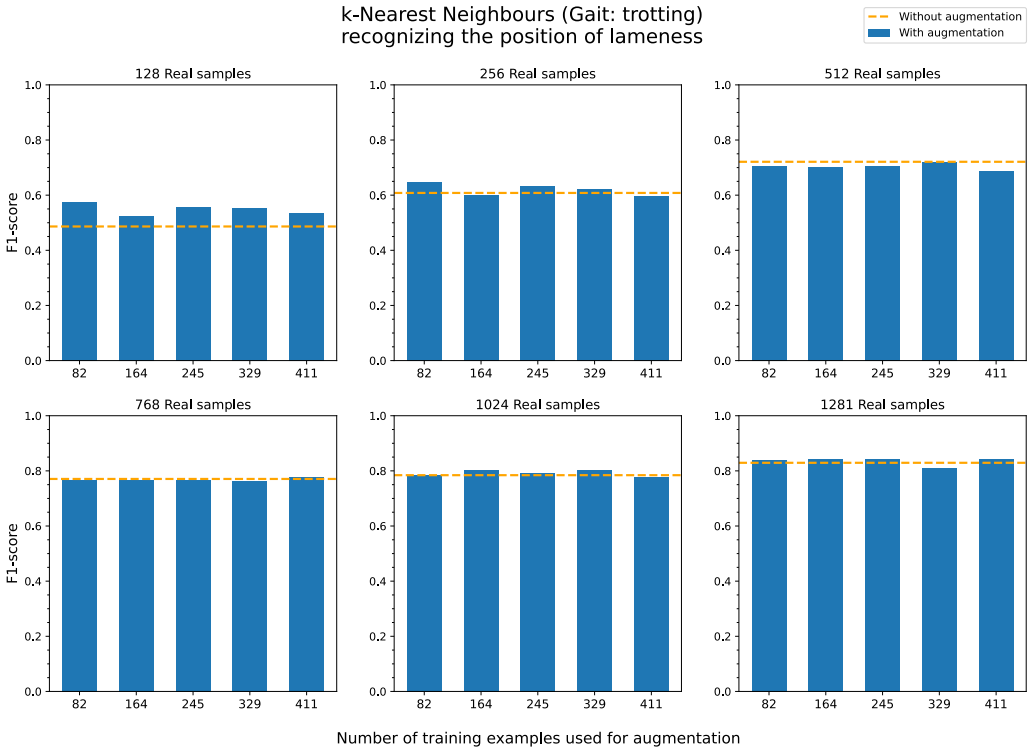

Random forest (Gait: walking)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

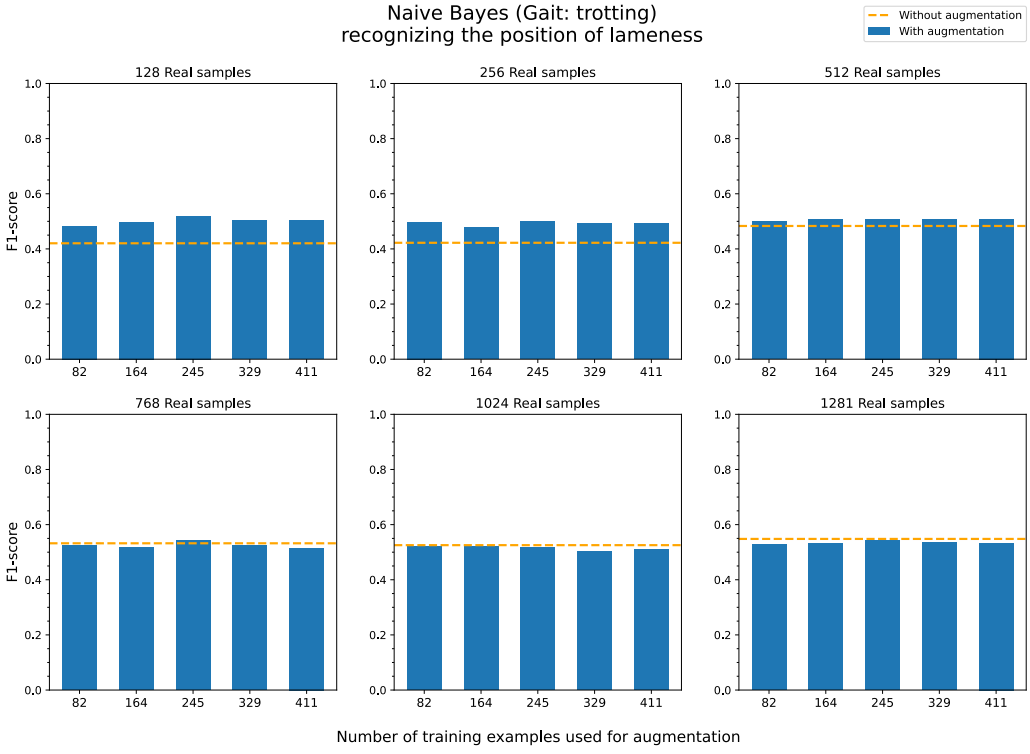

k-Nearest Neighbours (Gait: walking)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

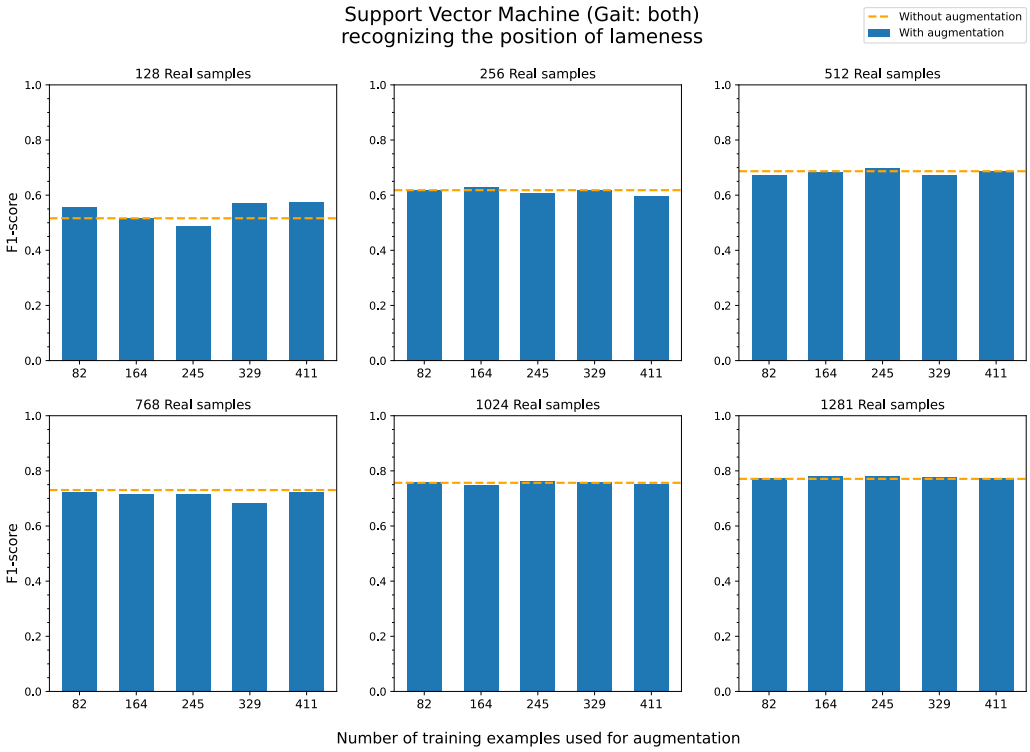

Naive Bayes (Gait: walking)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

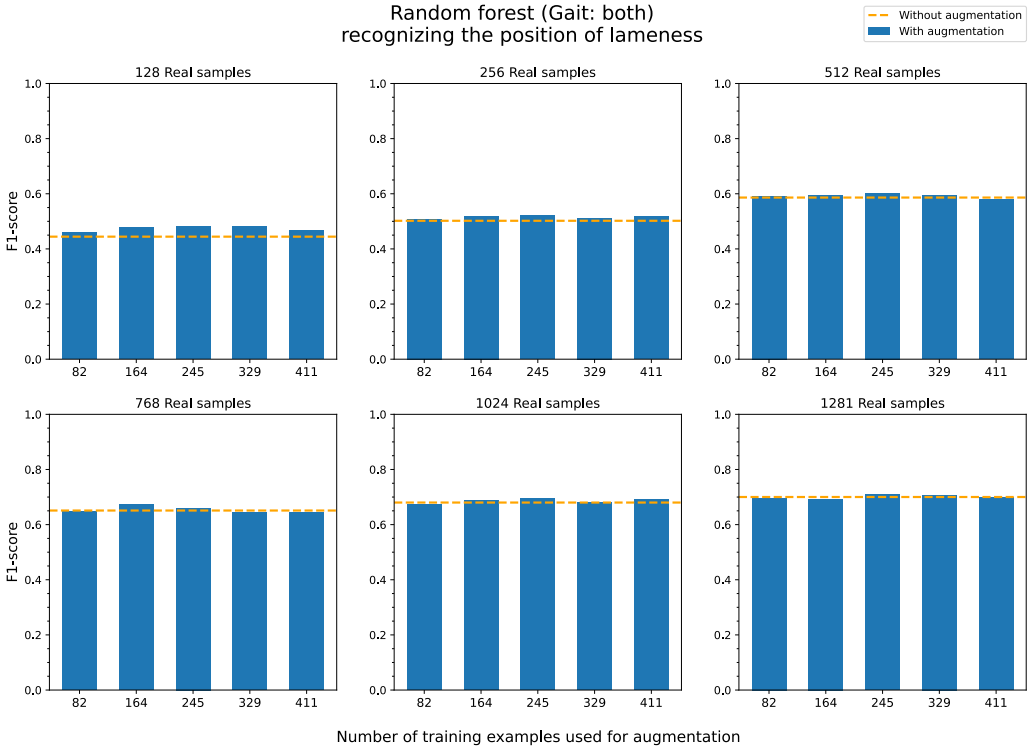

Support Vector Machine (Gait: trotting)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

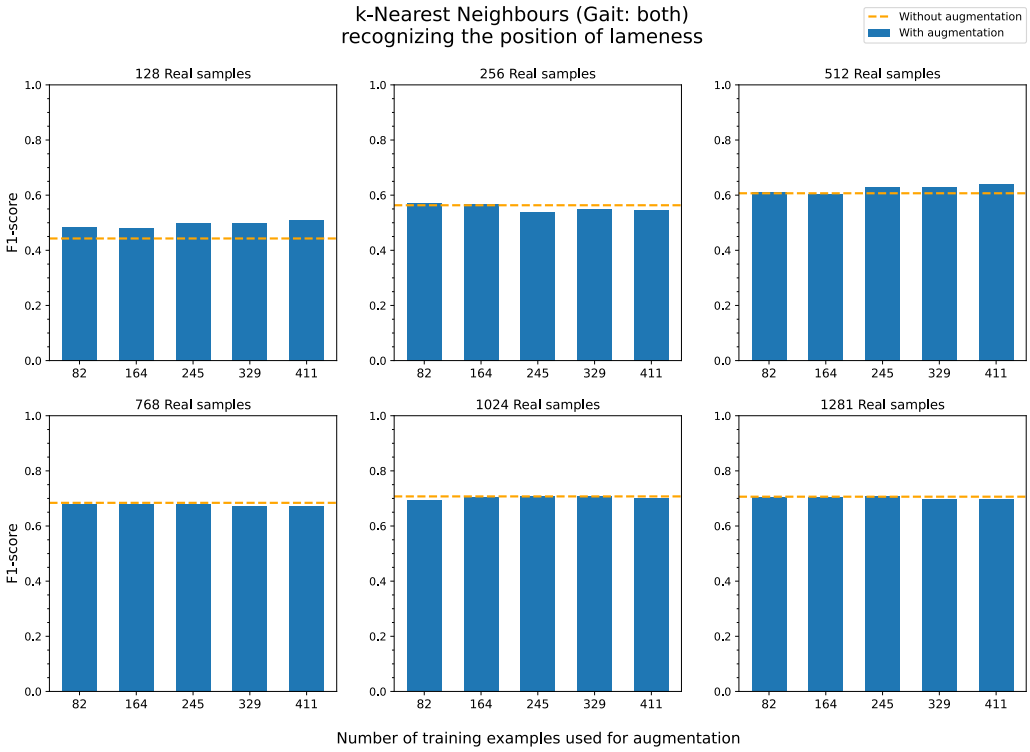

Random forest (Gait: trotting)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation

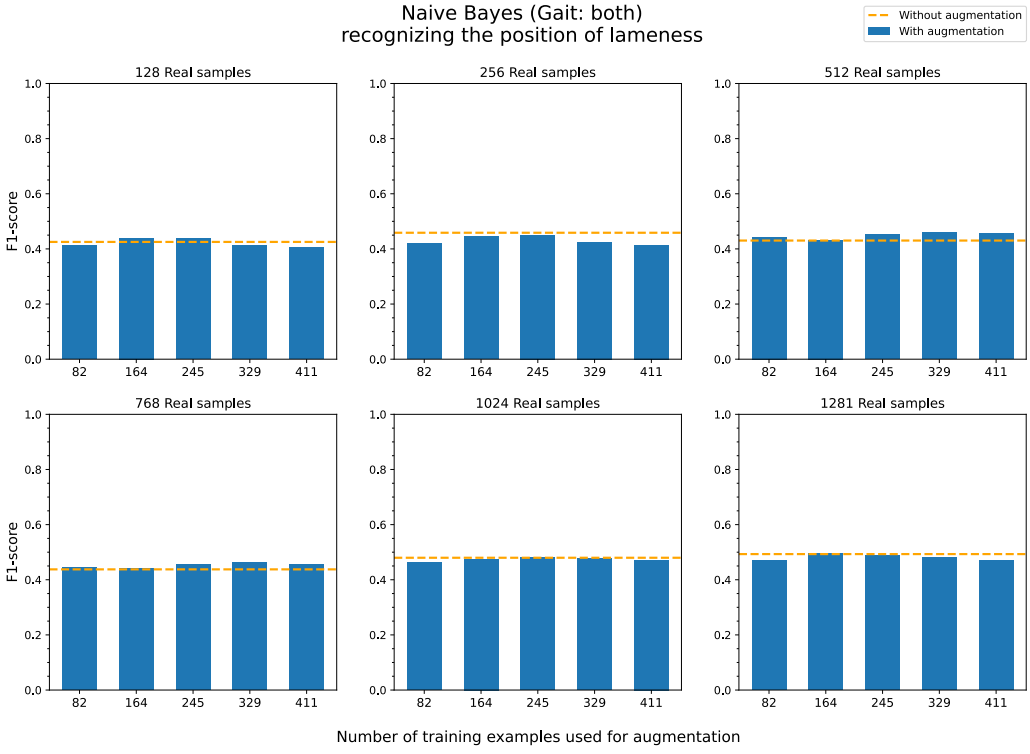

k-Nearest Neighbours (Gait: trotting)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
Number of training examples used for augmentation


Naive Bayes (Gait: trotting)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
Number of training examples used for augmentation


Support Vector Machine (Gait: both)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation


Random forest (Gait: both)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation


k-Nearest Neighbours (Gait: both)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation


Naive Bayes (Gait: both)
recognizing the position of lameness
Without augmentation
With augmentation
128 Real samples
256 Real samples
512 Real samples
768 Real samples
1024 Real samples
1281 Real samples
F1-score
0.0
0.2
0.4
0.6
0.8
1.0
82
164
245
329
411
Number of training examples used for augmentation