

Do Cows Have Fingerprints? Using Time Series Techniques and Milk Flow Profiles to Characterise Cow Milking Performance and Detect Health Issues

Changhong Jin^{1,2}[0000-0003-2565-592X], John Upton^{2,3}[0000-0002-6142-8372], and Brian Mac Namee^{1,2}[0000-0003-2518-0274]

¹ School of Computer Science, University College Dublin, Ireland

² VistaMilk SFI Research Centre, Teagasc Moorepark, Fermoy, Co. Cork, Ireland

³ Animal and Grassland Research and Innovation Centre, Teagasc Moorepark, Fermoy, Co. Cork, Ireland

Abstract. On modern dairy farms technologies that are capable of measuring high frequency indicators (e.g. milk yield, milk flow-rates, and electrical conductivity) at every milking can play an important role in helping farmers manage animal health. The most modern dairy farms use milk meters that provide detailed, high-frequency data about the flow of milk during every milking (cows are typically milked twice daily). This forms a time series that we call a *milk flow profile*. As cows are milked twice per day, every day this data forms a series of time series collected in a relatively controlled way that offers detailed insights into cow milking performance and cow health. In this paper we show that milk flow profiles act as a finger print for cows in a herd and offer opportunities for extracting useful insights about cow health that are unexplored. We demonstrate that unsupervised time series clustering approaches, particularly those that utilize the shape of time series, can be used to characterize a herd and that supervised approaches applied to milk flow profiles can be used for automated mastitis detection. In the latter case it is interesting that approaches using standard machine learning methods applied to features extracted from milk flow profiles, out-perform approaches specifically designed for time series.

Keywords: precision agriculture · dairy farming · mastitis detection · time series classification

1 Introduction

Managing animal health protocols to ensure high-quality milk supply to dairies is a significant challenge facing dairy farms today. Modern precision dairy farming technology that is capable of measuring high frequency indicators (e.g. milk yield, milk flow-rates, and electrical conductivity) at every milking can play an important role in this. Automated detection of mastitis in dairy cows is one of the best examples. Mastitis is a health problem afflicting dairy cows that

causes inflamed and painful udders leading to reduced milk yield, and is the most economically damaging disease on modern dairy farms [34]. Automated mastitis detection systems based on the milk yield at each milking, data on the composition of the milk collected at each milking⁴ [28], genetic data, the electrical conductivity of milk collected [32, 20], and cow medical history data have all been described in the literature [6, 38, 30].

On farms that have installed more advanced sensing technology known as milk meters⁵, detailed data about the flow of milk during every milking (cows are typically milked twice per day, once in the morning and once in the afternoon) has become available. Data relating to milk flow rates have been shown to be useful for a variety of health indications [15, 16, 40]. More detailed measurement of milk flow throughout milking allows a milk flow profile that shows the cumulative amount of milk that flows during milking to be constructed. Examples of milk flow profiles for the morning and afternoon milking of four different cows on a single day are shown in Figure 1. Each curve shows the cumulative milk yield (measured in kilograms) over time (measured at intervals of 10 seconds) from each milking. Milk flow profiles offer a rich and currently unexplored time series data source for monitoring cow health, and are the subject of this paper.

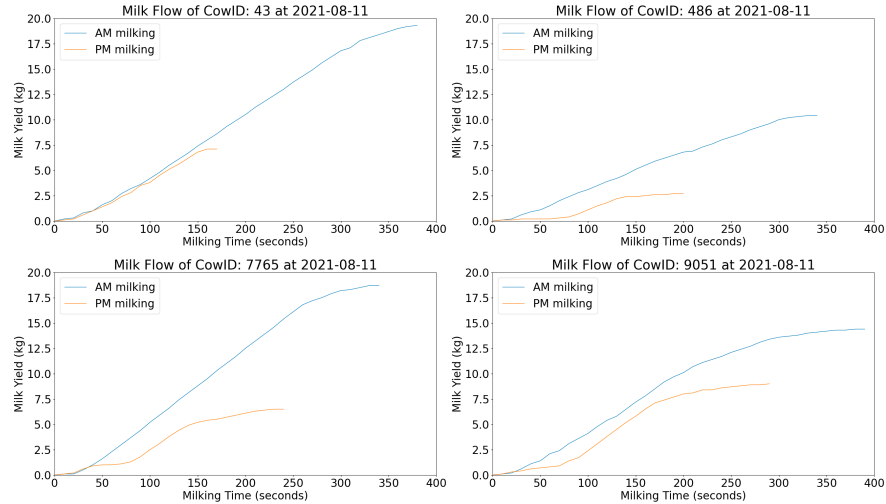


Fig. 1: Examples of milk flow profiles for the morning and afternoon milking of two different cows on a single day.

⁴ This is typically known as milk recording data and includes information such as the percentage of fat and lactose contained in the milk.

⁵ A milk meter is a device inserted into the milk pipeline that records the individual animal milk yield during milking.

In this paper we introduce milk flow profiles as a rich source of information about cow health and explore how they can be used to characterise cows in a herd—essentially providing a fingerprint for each animal. We also explore the use of milk flow profiles for the task of mastitis detection—this can be framed as a time series classification problem. Interestingly we find that distance based approaches—specifically 1-nearest neighbour classifiers using dynamic time warping—do not work well for this, while feature-based approaches give good results. The paper is organised as follows: Section 2 describes related work; Section 3 introduces milk flow profiles and illustrates how they can be used to profile a herd; Section 4 describes how milk flow profiles can be used for detecting mastitis (framed as a time series classification problem); and Section 5 concludes the paper and suggests directions for future work.

2 Related Work

Precision agriculture [39], or the application of data science to decision making on farms, has brought significant changes to how modern farms are managed. Farms now generate masses of data about crops, animals, inputs, outputs, and activities that can be stored and analysed to help farmers make better decisions. One of the most promising uses of data on farms is managing animal health [27]. Data describing animal bio-metrics and behaviours can aid farmers in better identifying and managing health problems, especially on larger farms. On modern dairy farms mastitis is the most significant udder health issue on modern dairy farms, and there are good examples of data-driven approaches to detecting and managing mastitis in the literature.

The mastitis detection problem can be framed in different ways. Some studies, for example [7] and [29], build models that detect both clinical mastitis and sub-clinical mastitis; while others, for example [19] and [9], focus only on one of these. Other studies, like [37] and [2], do not address mastitis directly but instead attempt to predict values for cows that are expensive to measure, in particular Somatic Cell Count (SCC) which requires sending samples to a lab for analysis, for cows based other more easily available data. Other studies, for instance [33] and [10], do not attempt to build prediction models but instead investigate factors that are good indicators of mastitis and which may be useful in building prediction models.

Analysis of time series in machine learning is a well studied area with applications in agriculture widely studied [12, 1, 18, 21]. While milk flow profiles themselves have not been studied widely before, there are other time series that are similar—for example undersea valves in oil and gas mining [3]. The remainder of this paper explores how milk flow profiles can be used to understand cow behaviours and to predict cow health issues.

3 Milk Flow Profiles

A milk flow profile captures detailed data about the cumulative milk yield harvested during milking. A milk meter installed in the milking machine measures the yield (kilograms) at regular intervals (typically every 10 seconds)—the average cow milking takes between 5 and 7 minutes. Figure 2a shows examples of milk flow profiles from the same cow on her morning milkings on multiple days, referred to as the *internal milk flow distance* (IMFD). Figure 2b shows flow profiles from morning milkings for all cows in the herd for a single milking day, referred to as *external milk flow distance* (EMFD).

These images clearly show the way that the milk flow profile can act as a finger print for a cow. All of the milk flow profiles for the single cow in Figure 2a show a similar pattern with a very tight distribution—the milk flow profile for each milking is essentially the same. On the other hand the milk flow profiles in Figure 2b have much higher variation with quite different patterns evident.

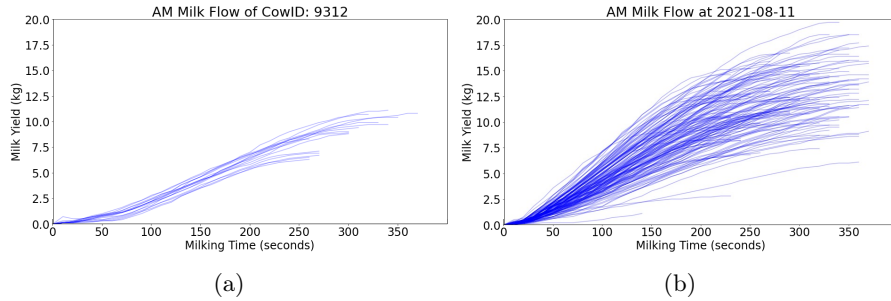


Fig. 2: A comparison of milk flow profiles for (a) all morning milking for one cow over the 4 month study period (IMFD) and (b) morning milkings for all cows in the herd for one single day (EMFD).

To formally measure this difference we calculate the distance, using dynamic time warping [26], between the milk flow profiles for all cows on a single day, EMFD, (shown in Figure 2a) and all milkings of a single cow in the dataset (shown in Figure 2b). Figure 3 compares the distributions of these distances. A permutation test [14] yields a p-value of 0.0007 indicating a statistically significant difference. It is clear that the distances between milk flow profiles for the single cow are much smaller than those calculated across the herd. Moreover, the permutation test provides evidence supporting the notion that the two groups exhibit statistically distinct distributions.

Although normal milk flow profiles are characterised by a common upward sloping curve, as milk yield increases with milking time, different cows can have dramatically different milk flow profiles. One cow may complete the entire milking process in a short time, resulting in a milk curve with a steep gradient, while another cow may take a long time to complete its milking and the correspond-

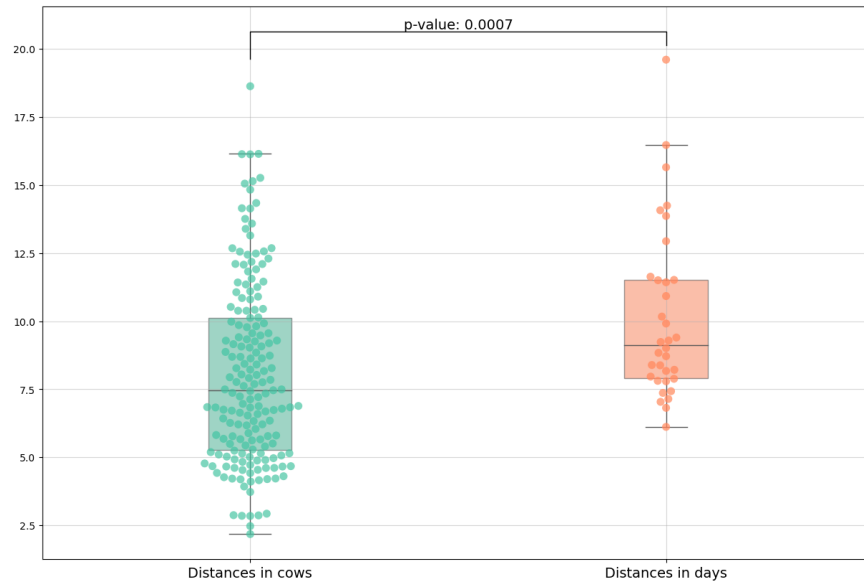


Fig. 3: Box plots showing the distribution of distances between all milk flow profiles from a single cow across the full dataset (IMFD) and distances between milk flow profiles of all cows in a herd on a single day (EMFD).

ing milk curve would be relatively flat. It could even be the case that in both instances the peak flow rates are the same. Not only does milk flow profile data provide greater detail about individual milking events of each cow, but changes in a cow's milk flow profiles could also help in the diagnosis of health problems such as mastitis.

To further explore the use of milk flow profiles as a finger print for different cows we use a dataset collected from milking equipment and laboratory analysis at Teagasc Dairy Research Centre at Moorepark, Ireland⁶. The milking parlour software was modified to record the milk flow-rate from each cow at every milking at 10 second intervals. Cows were managed in a pasture-based system and were milked twice per day, once in the morning and once in the afternoon.

The dataset covered 9 months from March 2021 to November 2021 and included 18,662 milk flow profiles at SCC measurement days from 293 dairy cows⁷. The cows were electronically identified on entering the milking parlor and data

⁶ A midi-line 30 unit Dairymaster herringbone, swing-over milking system (Dairymaster, Ireland) was used to milk the cows twice per day. The milking system utilised simultaneous pulsation and was fitted with automatic cluster removers and weigh-all milk meters (Dairymaster, Ireland). The standard farm milk flow rate switch-point of the automatic cluster removers was 0.2 kg/min.

⁷ The majority of cows were from the Holstein-Friesian breed. 92 of the cows were at parity of 1 (meaning they had given birth to just one calf, or were primiparous) and

from each milking session was linked to a specific cow. Cumulative milk yield, and conductivity were collected simultaneously as milking began, and then recorded at 10 second intervals until the end of the milking. A sample of milk was also taken from each cow on one occasion per week for composition and somatic cell count (SCC) analysis⁸.

To explore the different groups that exist within the dataset based on the milk flow profiles we performed a series of clustering exercises. Figure 4 shows the clusters found using k -means clustering with $k = 8$ (we experimented with different values and found that 8 clusters gave a good mix between achieving a high silhouette score and usefulness determined by domain experts) compared to other and dynamic time warping as the distance measure. In each case the blue lines represent the milk flow profile for a single milking and the red lines show the average milk flow profile for a cluster.

Examining the different clusters we see interesting sub-groups within the herd emerge. Clusters 0, 4, and 6 illustrate an almost linear milk profile where the rate of milking remains largely constant throughout milking indicating a relatively consistent flow of milk from the udder during the period of milking unit attachment. Clusters 1 and 3 show a pronounced curve illustrating rapid let down of milk followed by a relatively high and sustained peak milk flow period. These curves then tail off to an elongated low flow period, this typically indicates that one quarter of the udder is still producing milk at a low level until the milking unit is removed. Clusters 2 and 5 sit somewhere between these two, for example they have an identifiable peak milk flow period followed by an elongated low flow tail. Cluster 7 is a typical *other* cluster, and contains a broad spread of milk flow curve types that did not fit well with clusters 0 to 6.

The value of using milk flow profiles for this clustering, is illustrated by the the scatter plots shown in Figure 5. Figure 5a shows a scatter plot of the dataset used for clustering where the vertical axis shows total milk yielded from the milking and the horizontal axis shows total time taken for milking. The colours of the points indicate cluster membership for clustering performed using non-standardised data and Euclidean distance. It is clear from this that clustering using the milk profile in a non-standardised format and Euclidean distance captures no more information from the milk flow profiles than clustering simply using total yield and total milking time. Essentially we can see that cluster membership is determined by milk yield and milking time by the colour bands showing that each cluster is clearly separated from the others in this scatter plot.

the remaining 201 were at parity of 2 or greater (meaning they had given birth to just one calf, or were multiparous).

⁸ Milk composition analysis, often know as milk recording, is used to analyse the content of milk. Typically the percentage of fat, protein, and lactose in the milk is measured as well as the amount of casein and urea (both important in cheese making) present. The number of somatic cells, usually white blood cells, is also measured and typically referred to as somatic cell count (SCC). SCC is the most used and studied indicator in mastitis detection research, especially for sub-clinical mastitis detection [36]. A Fossomatic machine (Foss, Denmark) was used to measure SCC and other indicators of milk composition.

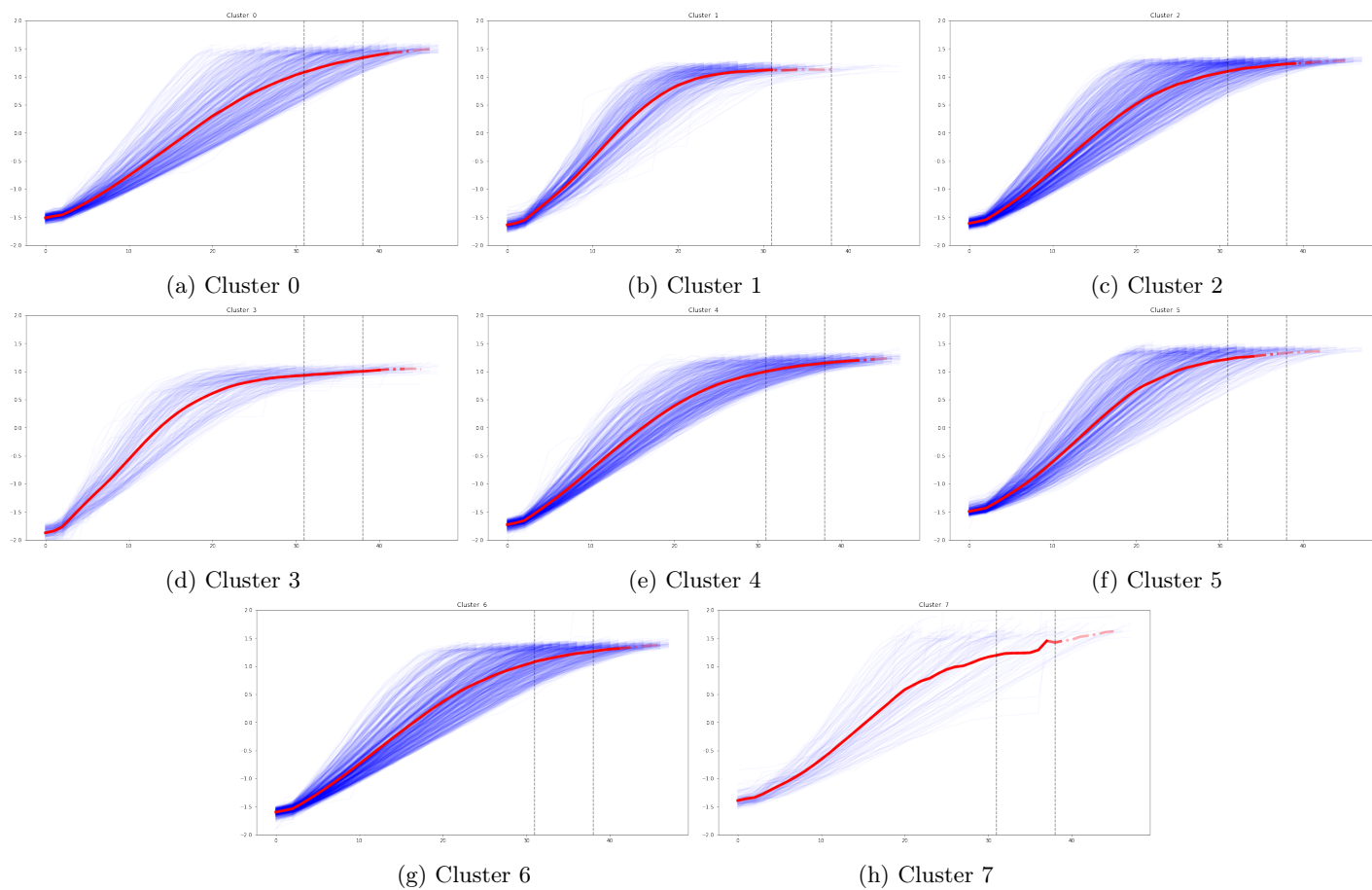


Fig. 4: The results of clustering the milk flow profiles in the dataset into 8 clusters using k -means clustering based on DTW distances. The blue lines show milk flow profiles and the red lines show the average for each cluster.

In Figure 5b showing the same plot after clustering using standardised data and DTW distance, however, we see no distinction between cluster memberships showing that the milk flow profiles provide distinctions between cows and milkings beyond what is evident from simply looking at total yield and milking time. This is a compelling illustration of the value of milk flow profiles. In the next section we explore how these profiles can be used to predict health problems in cows.

4 Using Milk Flow Profiles for Mastitis Detection

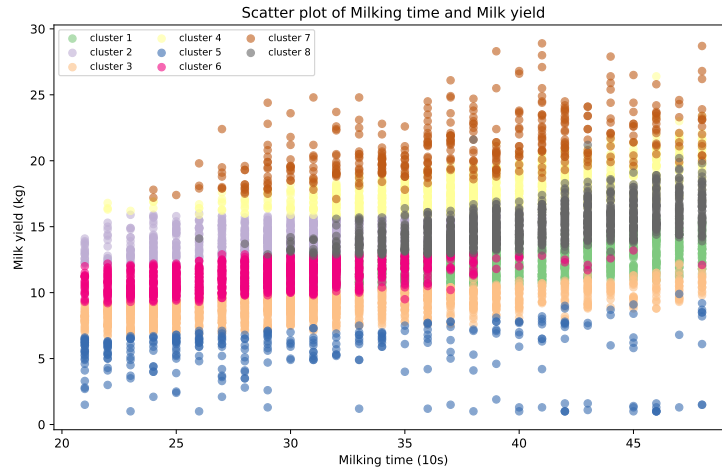
Mastitis is recognised as the most significant health problem on modern dairy farms [34]. Elevated levels of SCC are recognised as indicating sub-clinical mastitis—the early stages of mastitis, before symptoms (such as inflamed udders or clots in milk) are visible [13].

Detecting the onset of sub-clinical mastitis gives farmers the opportunity to intervene before cows progress to clinical mastitis or an infection spreads further to other cows in the herd. Although high SCC recordings are considered a reliable indicator of sub-clinical mastitis [36], recording SCC is a time-intensive process, typically only performed once per month on dairy farms. For this reason automated detection of sub-clinical mastitis using machine learning methods applied to data more routinely available from milking machines are attractive and have been widely studied (e.g. [28, 2]). However, milk flow profiles have not been widely used for this before. In this section we describe an experiment to use time series classification techniques to classify milk flow profiles as belonging to days on which a cow had a high SCC value, and so likely sub-clinical mastitis, versus days when they did not. This illustrates the value of the milk flow profile time series as an indicator of cow health.

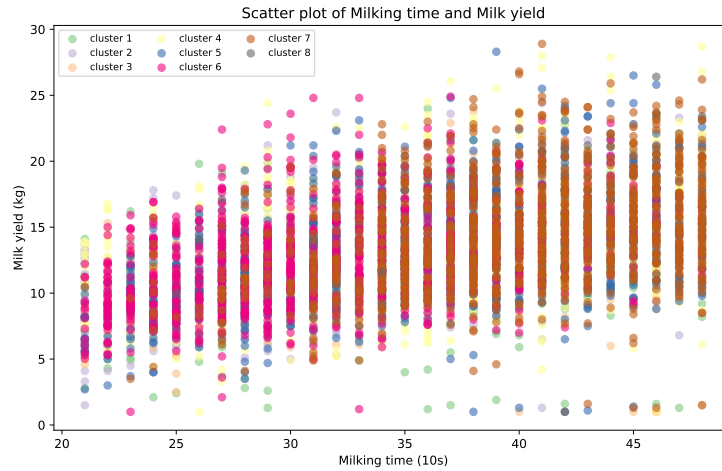
4.1 Dataset

The dataset described in Section 3 was used in this study. This includes milk flow profiles for morning and afternoon milkings for each cow in the herd, and details of weekly laboratory testing of each cows’ milk characteristics (including the percentage of fat, protein, and lactose in the milk, as well as the amount of casein and urea present, and the SCC). Milk characteristic sampling was performed every 7 days, during the morning milking.

The SCC recordings captured every week were used to define the target feature used in these experiments. Based on the recommendation of [17], for cows having their first calf (referred to as primiparous) sub-clinical mastitis was indicated when the recorded SCC value exceeded 150,000 cells/mL. For cows having their second or later calves (referred to as multiparous) sub-clinical mastitis was indicated when the SCC value exceeded 250,000 cells/mL. The prediction problem then was to replicate the information in the SCC measurements using the much cheaper and easier to collect milk flow profile data.



(a) Non-standardised data Euclidean distances



(b) Standardised data DTW distances

Fig. 5: Scatter plots of milk yield versus milking time for the cow dataset with points coloured based on cluster membership for (a) clusters found using non-standardised data and Euclidean distance, and (b) standardised data and DTW distance.

Keeping only the morning milkings for which an SCC recording existed results in a dataset with 8,703 rows each containing a milk flow profile and an SCC reading for the corresponding day. To remove outliers, records containing values

exceeding +/- three standard deviations from the mean were removed. Similarly, any records containing values outside the technical parameters of the data collection devices were removed (e.g. SCC values above 10 million cells/ml). Also, records with milk flow profiles not starting from 0 and negative values for milk characteristics were treated as outliers. These steps removed 5.2% of the original dataset. The dataset contained significant class-imbalance, with many more instances where the sub-clinical mastitis target feature was negative than when it was positive (7225 versus 1478). Under-sampling was applied in this study to reduce the proportion of majority classes by randomly removing observations from the majority class.

We use two representations of this data: the raw milk flow profiles which are used as inputs to the best ranked classifiers for univariate time series classification problems, and representations based on features extracted from the flow profiles which are used as input to other models. The most common method of extracting features from time series data of the type of milk flow profiles is to calculate summary statistics, such as mean, standard deviation, maximum, and minimum from the values in the time series. The features extracted from the milk flow profiles in this work can be grouped into three categories⁹. First, statistical time domain features such as maximum and minimum values, and lengths. Secondly, univariate temporal features that provide the dynamics of the milk flow profile, for example, auto-correlation, entropy, and total energy. Finally, spectral time-frequency features, representing changes in the frequency domain over time. In our experiments we also use the Catch-22 feature set [23], a small set of derived features shown to be effective for time series classification problems.

4.2 Experimental Setup

In this section we describe the setup of experiments to evaluate the effectiveness of using milk flow profiles to predict the presence of sub-clinical mastitis. First we describe the models used, the experimental methodology, and finally the performance measures employed.

To demonstrate the ability of machine learning methods based on raw time series representations and representations based on features extracted from time series data, we use four machine learning algorithms in our experiments: 1-*nearest neighbor* using *dynamic time warping* as a distance measure (1NN-DTW) [4, 26, 35, 41], *random convolutional kernel transform* (ROCKET) [8], *the hierarchical vote collective of transformation-based ensembles* (HIVE-COTE 2.0) [24] and *gradient boosting machines* (GBMs) [11] that use a combination of many weak learning models to create a strong predictive model.

To evaluate the performance of each model type for the mastitis detection problem we perform a 5-fold cross-validation. For each machine learning algorithm we evaluate performance on the original dataset and on an under sampled

⁹ All features were extracted from flow profile data using the tsfresh package in Python [5]

version (under-sampling is only applied to the training partition for each experiment). Accuracy fails to effectively represent model performance for imbalanced data sets. For this reason, the area under the receiver operating characteristic curve (AUC) [22] is more appropriate.

4.3 Results & Discussion

Table 1 shows the results of experiments completed. The AUC scores show that it is possible to train effective mastitis prediction models using milk flow profiles as input. The AUC scores shown are in line with similar results in the literature based on other data type inputs. It is particularly interesting to note that the feature based approaches were more effective than the approaches based on raw time series representations. This is contradictory to other benchmark results which typically indicate the success of ROCKET and HIVE-COTE 2.0 [31, 25]. It is also interesting to note that the models trained using the larger feature set are more effective than those trained on the set of Catch-22 features.

5 Conclusions & Future Work

This paper has introduced milk flow profiles. These are enabled by the use of milk meters in modern dairy parlours and allow the use of time series techniques for analysis. We have shown how the milk profiles themselves can be used as a cow finger-print offering insights into cow milking performance and cow health. Clustering techniques show how a herd can be partitioned according to these profiles. We also demonstrate that milk flow profiles can be used for predicting health problems faced by cows, in this case sub-clinical mastitis (one of the most significant health issues faced on dairy farms).

This work introduces this new data representation and in the future we intend developing this work to build better health prediction models by integrating milk flow profiles with other data sources. We also intend to take advantage of the interpretable nature of milk flow profiles to build explainable prediction models.

Table 1: Experimental results for model comparison.

Modelling	AUC Score
1NN-DTW	0.513
ROCKET	0.607
HIVE-COTE 2.0	0.622
GBM with Catch-22	0.659
GBM with tsfresh	0.677

Acknowledgements This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) and the Department of Agriculture, Food and Marine on behalf of the Government of Ireland under Grant Number [16/RC/3835]. Further financial and technical support was provided by Dairymaster.

References

1. Abdoli, A., Murillo, A.C., Yeh, C.C.M., Gerry, A.C., Keogh, E.J.: Time series classification to improve poultry welfare. In: 2018 17TH IEEE International conference on machine learning and applications (ICMLA). pp. 635–642. IEEE (2018)
2. Anglart, D., Hallén-Sandgren, C., Emanuelson, U., Rönnegård, L.: Comparison of methods for predicting cow composite somatic cell counts. *Journal of dairy science* **103**(9), 8433–8442 (2020)
3. Atif Qureshi, M., Miralles-Pechuán, L., Payne, J., O’Malley, R., Namee, B.M.: Valve health identification using sensors and machine learning methods. In: IoT Streams for Data-Driven Predictive Maintenance and IoT, Edge, and Mobile for Embedded Machine Learning: Second International Workshop, IoT Streams 2020, and First International Workshop, ITEM 2020, Co-located with ECML/PKDD 2020, Ghent, Belgium, September 14–18, 2020, Revised Selected Papers 2. pp. 45–60. Springer (2020)
4. Berndt, D.J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: KDD workshop. vol. 10, pp. 359–370. Seattle, WA, USA: (1994)
5. Christ, M., Braun, N., Neuffer, J., Kempa-Liehr, A.W.: Time series feature extraction on basis of scalable hypothesis tests (tsfresh—a python package). *Neurocomputing* **307**, 72–77 (2018)
6. Cogato, A., Bršćić, M., Guo, H., Marinello, F., Pezzuolo, A.: Challenges and tendencies of automatic milking systems (ams): A 20-years systematic review of literature and patents. *Animals* **11**(2), 356 (2021)
7. De Mol, R., Kroeze, G., Achten, J., Maatje, K., Rossing, W.: Results of a multivariate approach to automated oestrus and mastitis detection. *Livestock Production Science* **48**(3), 219–227 (1997)
8. Dempster, A., Petitjean, F., Webb, G.I.: Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery* **34**(5), 1454–1495 (2020)
9. Ebrahimi, M., Mohammadi-Dehcheshmeh, M., Ebrahimie, E., Petrovski, K.R.: Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep learning and gradient-boosted trees outperform other models. *Computers in biology and medicine* **114**, 103456 (2019)
10. Ebrahimie, E., Ebrahimi, F., Ebrahimi, M., Tomlinson, S., Petrovski, K.R.: A large-scale study of indicators of sub-clinical mastitis in dairy cattle by attribute weighting analysis of milk composition features: highlighting the predictive power of lactose and electrical conductivity. *Journal of dairy research* **85**(2), 193–200 (2018)
11. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. *Annals of statistics* pp. 1189–1232 (2001)
12. Frizzarin, M., Visentin, G., Ferragina, A., Hayes, E., Bevilacqua, A., Dhariyal, B., Domijan, K., Khan, H., Ifrim, G., Le Nguyen, T., et al.: Classification of cow diet

- based on milk mid infrared spectra: A data analysis competition at the “international workshop on spectroscopy and chemometrics 2022”. *Chemometrics and Intelligent Laboratory Systems* p. 104755 (2023)
13. Frössling, J., Ohlson, A., Hallén-Sandgren, C.: Incidence and duration of increased somatic cell count in swedish dairy cows and associations with milking system type. *Journal of Dairy Science* **100**(9), 7368–7378 (2017)
 14. Good, P.: *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media (2013)
 15. Grindal, R.J., Hillerton, J.E.: Influence of milk flow rate on new intramammary infection in dairy cows. *Journal of Dairy Research* **58**(3), 263–268 (1991)
 16. Japertiene, R., Juozaitiene, V., Kriauciene, J., Rudejeviene, J., Japertas, S.: The interrelationships between milkability traits and subclinical mastitis in cows. *Polish journal of veterinary sciences* **10**(4), 255–261 (2007)
 17. Jensen, D.B., van der Voort, M., Hogeveen, H.: Dynamic forecasting of individual cow milk yield in automatic milking systems. *Journal of dairy science* **101**(11), 10428–10439 (2018)
 18. Kamilaris, A., Prenafeta-Boldú, F.X.: Deep learning in agriculture: A survey. *Computers and electronics in agriculture* **147**, 70–90 (2018)
 19. Khatun, M., Thomson, P., Kerrisk, K., Lyons, N., Clark, C., Molfino, J., García, S.: Development of a new clinical mastitis detection method for automatic milking systems. *Journal of dairy science* **101**(10), 9385–9395 (2018)
 20. Khatun, M., Clark, C.E., Lyons, N.A., Thomson, P.C., Kerrisk, K.L., García, S.C.: Early detection of clinical mastitis from electrical conductivity data in an automatic milking system. *Animal Production Science* **57**(7), 1226–1232 (2017)
 21. Liu, G., Zhong, K., Li, H., Chen, T., Wang, Y.: A state of art review on time series forecasting with machine learning for environmental parameters in agricultural greenhouses. *Information Processing in Agriculture* (2022)
 22. Lobo, J.M., Jiménez-Valverde, A., Real, R.: Auc: a misleading measure of the performance of predictive distribution models. *Global ecology and Biogeography* **17**(2), 145–151 (2008)
 23. Lubba, C.H., Sethi, S.S., Knaute, P., Schultz, S.R., Fulcher, B.D., Jones, N.S.: catch22: Canonical time-series characteristics: Selected through highly comparative time-series analysis. *Data Mining and Knowledge Discovery* **33**(6), 1821–1852 (2019)
 24. Middlehurst, M., Large, J., Flynn, M., Lines, J., Bostrom, A., Bagnall, A.: Hive-cote 2.0: a new meta ensemble for time series classification. *Machine Learning* **110**(11-12), 3211–3243 (2021)
 25. Middlehurst, M., Schäfer, P., Bagnall, A.: Bake off redux: a review and experimental evaluation of recent time series classification algorithms. *arXiv preprint arXiv:2304.13029* (2023)
 26. Müller, M.: Dynamic time warping. *Information retrieval for music and motion* pp. 69–84 (2007)
 27. Neethirajan, S.: The role of sensors, big data and machine learning in modern animal farming. *Sensing and Bio-Sensing Research* **29**, 100367 (2020)
 28. Pakrashi, A., Ryan, C., Guéret, C., Berry, D., Corcoran, M., Keane, M., Mac Namee, B.: Early detection of subclinical mastitis in lactating dairy cows using cow-level features. *Journal of Dairy Science* **106**(7), 4978–4990 (2023). <https://doi.org/https://doi.org/10.3168/jds.2022-22803>, <https://www.sciencedirect.com/science/article/pii/S0022030223002977>

29. Panchal, I., Sawhney, I., Sharma, A., Dang, A.: Classification of healthy and mastitis murrah buffaloes by application of neural network models using yield and milk quality parameters. *Computers and Electronics in Agriculture* **127**, 242–248 (2016)
30. Pyörälä, S.: Indicators of inflammation in the diagnosis of mastitis. *Veterinary research* **34**(5), 565–578 (2003)
31. Ruiz, A.P., Flynn, M., Large, J., Middlehurst, M., Bagnall, A.: The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery* **35**(2), 401–449 (2021)
32. Rutten, C.J., Velthuis, A., Steeneveld, W., Hogeveen, H.: Invited review: Sensors to support health management on dairy farms. *Journal of dairy science* **96**(4), 1928–1952 (2013)
33. Santman-Berends, I., Riekerink, R.O., Sampimon, O., Van Schaik, G., Lam, T.: Incidence of subclinical mastitis in dutch dairy heifers in the first 100 days in lactation and associated risk factors. *Journal of dairy science* **95**(5), 2476–2484 (2012)
34. Seegers, H., Fourichon, C., Beaudeau, F.: Production effects related to mastitis and mastitis economics in dairy cattle herds. *Veterinary research* **34**(5), 475–491 (2003)
35. Senin, P.: Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA* **855**(1-23), 40 (2008)
36. Sharma, N., Singh, N., Bhadwal, M.: Relationship of somatic cell count and mastitis: An overview. *Asian-Australasian Journal of Animal Sciences* **24**(3), 429–438 (2011)
37. Sitkowska, B., Piwczynski, D., Aerts, J., Kolenda, M., ÖZKAYA, S.: Detection of high levels of somatic cells in milk on farms equipped with an automatic milking system by decision trees technique. *Turkish journal of veterinary and animal sciences* **41**(4), 532–540 (2017)
38. Slob, N., Catal, C., Kassahun, A.: Application of machine learning to improve dairy farm management: A systematic literature review. *Preventive Veterinary Medicine* **187**, 105237 (2021)
39. Stafford, J.V.: Implementing precision agriculture in the 21st century. *Journal of agricultural engineering research* **76**(3), 267–275 (2000)
40. Upton, J., Penry, J., Rasmussen, M., Thompson, P., Reinemann, D.: Effect of pulsation rest phase duration on teat end congestion. *Journal of Dairy Science* **99**(5), 3958–3965 (2016)
41. Xi, X., Keogh, E., Shelton, C., Wei, L., Ratanamahatana, C.A.: Fast time series classification using numerosity reduction. In: *Proceedings of the 23rd international conference on Machine learning*. pp. 1033–1040 (2006)