Original Paper

Prediction of pregnancy state from milk mid-infrared (MIR) spectroscopy in dairy cows

Lisa Rienesl[1], Philipp Pfeiffer[1]*, Negar Khayatzadeh[1], Astrid Köck[2], Laura Dale[3], Andreas Werner[3], Clément Grelet[4], Nicolas Gengler[5], Franz-Josef Auer[6], Christa Egger-Danner[2], Julie Leblois[7], Johann Sölkner[1]

[1] *University of Natural Resources and Life Sciences, Vienna (BOKU), Division of Livestock Sciences, Department of Sustainable Agricultural Systems, Vienna, Austria*

[2] *ZuchtData EDV-Dienstleistungen GmbH, Vienna, Austria*

[3] *Regional association for performance testing in livestock breeding of Baden-Wuerttemberg (LKV - Baden-Wuerttemberg), Stuttgart, Germany*

[4] *Walloon Agricultural Research Center (CRA-W), Gembloux, Belgium*

[5] *Université de Liège (ULg), Gembloux Agro-Bio Tech, Gembloux, Belgium*

[6] *LKV Austria Gemeinnützige GmbH, Wien, Austria*

[7] *Elevéo (awé groupe), Ciney, Belgium*

Pregnancy assessment is a very important tool for the reproductive management in efficient and profitable dairy farms. Nowadays, mid-infrared (MIR) spectroscopy is the method of choice in the routine milk recording system for quality control and to determine standard milk components. Since it is well known that there are changes in milk yield and composition during pregnancy, the aim of this study was to develop a discriminant model to predict the pregnancy state from routinely recorded MIR spectral data. The data for this study was from the Austrian milk recording system. Test day records of Fleckvieh, Brown Swiss and Holstein Friesian cows between 3 and 305 days of lactation were included in the study. As predictor variables, the first derivative of 212 selected MIR spectral wavenumbers were used. The data set contained roughly 400,000 records from around 40,000 cows and was randomly split into calibration and validation set by farm. Prediction was done with Partial Least Square Discriminant Analysis. Indicators of model fit were sensitivity, specificity, balanced accuracy and Area Under Receiver Operating Characteristic Curve (AUC). In a first approach, one discriminant model for all cows across the whole lactation and gestation lengths was applied. The sensitivity and specificity of this model in validation were 0.856 and 0.836, respectively. Splitting up the results for different lactation stages showed that the model was not able to predict pregnant cases before the third month of lactation and vice versa not able to predict non-pregnancy after the third month of lactation. Consequently, in the second approach a prediction model for each different (expected) pregnancy stage and lactation stage was developed. Balanced accuracies ranged from 0.523 to 0.918. Whether prediction accuracies from this study are sufficient to provide farmers with an additional tool for fertility management, it needs to be explored in discussions with farmers and breeding organizations.

Keywords: MIR spectroscopy, pregnancy prediction, dairy cow, PLS

1   Introduction

Pregnancy assessment is an essential tool for the reproductive management in cattle farms (e.g. Balhara et al., 2013, Pohler et al., 2016, Hirpa et al., 2018). Ideally, a cow should calve every year and therefore the identification of pregnant and non-pregnant animals at an early stage is crucial (Hirpa et al., 2018). Early detection of pregnancy status also enables early detection and treatment of problems

---

* **Corresponding Author**: Philipp Pfeiffer. University of Natural Resources and Life Sciences, Vienna (BOKU), Division of Livestock Sciences, Department of Sustainable Agricultural Systems, Gregor-Mendel-Strasse 33, A-1180 Vienna. E-mail: pfeiffer.philipp1@gmx.at.

(Bekele et al., 2016). Fertility is the most frequent reason for the culling of cows in dairy farms, accounting 24.2% in Austrian dairy farms in 2018 (Egger-Danner et al., 2018). Tools for pregnancy detection should be inexpensive and simple to apply under field conditions (Pohler et al., 2016). Basically, there are two types of diagnosing pregnancy state: direct methods such as estrus detection, transrectal palpation and transrectal ultrasonography, and indirect methods like analysis of progesterone and pregnancy-associated glycoproteins in milk or blood (Balhara et al., 2013, Pohler et al., 2016). Another important fact to be considered, are pregnancy losses. According to a review paper of Santos et al. (2004), the average embryonic mortality rate in dairy cows was 12.8% based on 14 studies. Another study by Humblot (2001) averaged the early and the late embryonic pregnancy losses in Holstein cows in 44 French herds after first insemination to 31.6% and 14.7%, respectively. For late embryonic and fetal losses, Santos et al. (2004) reported a value of 10.7% on average, based on 17 dairy farms. Pregnancy losses reduce the benefit of the early pregnancy diagnosis. Hence, repeated information about the pregnancy status of a dairy cow (e.g. at every test day) would be a truly useful tool for dairy farmers.

Mid-infrared (MIR) spectroscopy is the method of choice in the standard milk recording systems for quality control and to determine milk contents including fat, protein, lactose and urea (Grelet et al., 2015, 2016). MIR spectra data could also be used to predict fine components of milk such as minerals (Toffanin et al., 2015) or fatty acids (Soyeurt et al., 2011). Moreover, there are studies to predict various other traits and variables such as blood metabolites (Benedet et al., 2019) and methane emissions (Vanlierde et al., 2018). As it is well known that there are changes in milk yield and also milk composition during the pregnancy in dairy cows (Olori et al., 1997, Lainé et al., 2017), MIR spectra data could be potentially useful to predict pregnancy state of dairy cows. There are few relevant studies on this subject, which were exploring quite different approaches. Lainé et al. (2014) used residual spectra to detect pregnancy status and observed only the first 50 days after insemination. The reported prediction accuracies were very promising (sensitivity >0.99; specificity >0.84) but could not be reproduced on an independent data set (N. Gengler, 2020, University of Liege, Gembloux, Belgium, personal communication). Another study of Delhez et al. (2020) explored different modelling approaches for diagnosing pregnancy status from MIR spectra. In one strategy they only used single spectral records after insemination, from where records after a successful insemination were considered as 'pregnant' and records after an unsuccessful insemination were considered as 'open'. For the described strategy sensitivity was 0.65 and specificity 0.56. In another strategy, seven different models based on stages after insemination were developed; sensitivities ranging from 0.57 to 0.75 and specificities from 0.52 to 0.74 were obtained.

The aim of this study was to develop a discriminant model to predict the pregnancy status from routinely recorded MIR spectral data, and to further provide probabilities of pregnancy for each test day. Pregnancy probabilities obtained in this way could provide extra information for farmers in the framework of routine milk recording. Two different approaches were evaluated. The novelty of the second approach was the exploration of separate prediction models for different lactation stages.

2 Material and methods

2.1 Data and data preparation

The data for this study was from the Austrian milk recording system for the period July 2014 to February 2019 and was kindly provided by Zuchtdata GmbH. Test day milk data contained information on breed, herd, parity, days in milk, milk components (fat, protein, urea, lactose), somatic cell count (SCC) and standardized MIR spectral data for the respective test days. Additionally, information on the exact insemination and calving dates was available. Test day records of Fleckvieh, Brown Swiss and Holstein Friesian cows between 3 and 305 days of lactation were included. On average cows in the data set were pregnant at lactation day 93. Merging of the data sets and primary data preparation were done with the software SAS (SAS Institute Inc., 2017). Table 1 shows the number of records of the complete data set.

To define test day records as 'pregnant' and 'open', the pregnancy state of each cow was connected to the associated test day by the following procedure: gestation length was calculated as the date of re-calving minus the date of latest insemination, also defined as successful insemination. Only records of cows with ranges of gestation length as implemented in the joint genetic evaluation of Austria, Germany and Czech Republic were included: Fleckvieh 275 to 305 days, Holstein 268 to 298 days, Brown Swiss 276 to 306 days (C. Fuerst, 2019, Zuchtdata GmbH, Vienna, personal communication). Test days without a confirmed date of next calving were excluded. Test day records before

a successful insemination date were coded as 'open' and all test day records between the date of successful insemination and date of re-calving were coded as 'pregnant'. This procedure is visualized in Figure 1. The distribution of all test day records by class of pregnancy status (open or pregnant) along the stage of lactation is displayed in Figure 2.

Table 1 Characteristics of the final data set

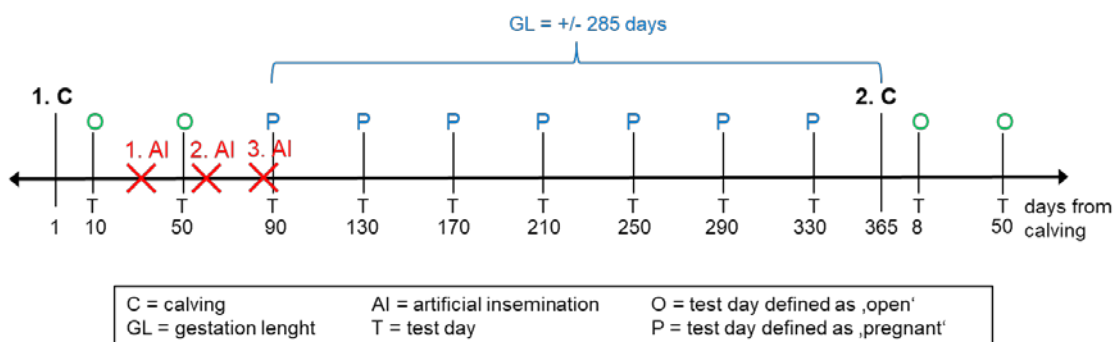| Item	*n*
Farms	6,899
Cows	40,106
Fleckvieh	30,589
Brown Swiss	3,854
Holstein Friesian	5,663
Test day records	403,863
open	124,163
pregnant	279,700

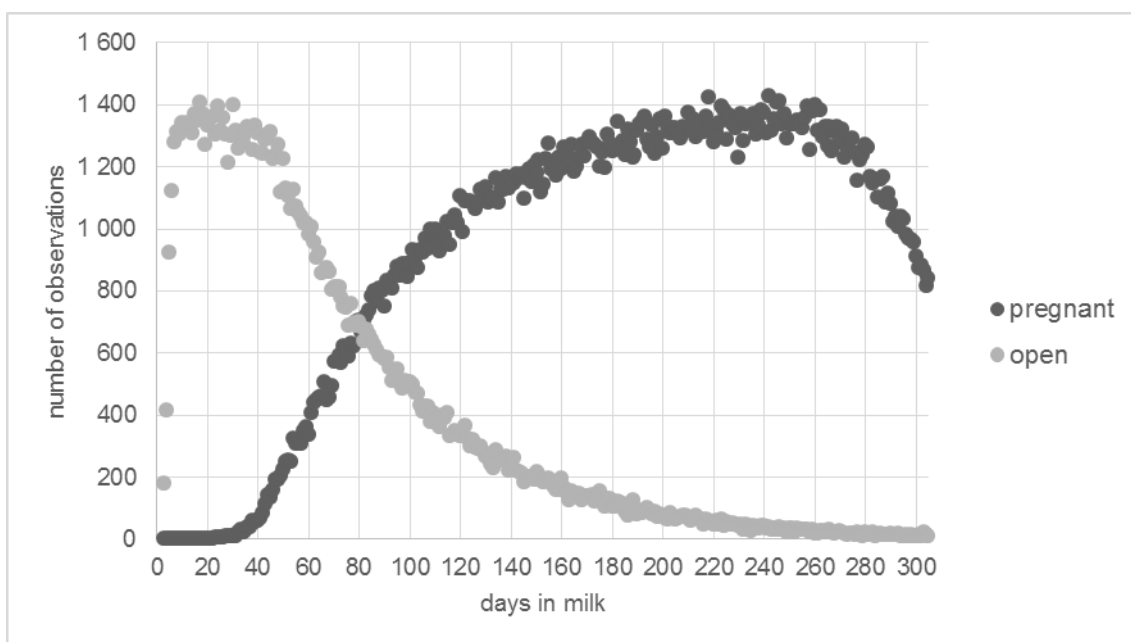**Figure 1** Defining test day records as 'open' or 'pregnant'

Figure 2 Distribution of test day records by pregnancy status (pregnant and open) along days in milk

MIR spectra were collected in several Austrian milk labs with Foss instruments spectrometers. Those MIR spectra consist of 1,060 data points, which are the absorbance values of infrared light at different wavenumbers, with frequencies from 926 to 5,010 cm^{-1}. Spectral data from different machines and different periods were previously standardized into a common basis (Grelet et al., 2015). For prediction models only selected areas of the spectra were used: 968.1 to 1,577.5 cm^{-1}, 1,731.8 to 1,762.6 cm^{-1}, 1,781.9 to 1,808.9 cm^{-1} and 2,831 to 2,966 cm^{-1} (Grelet et al., 2016). The 212 selected data points contain most of the usable information after removal of areas known to be non-reproducible between instruments or non-informative due to strong water absorption. According to other relevant studies (Soyeurt et al., 2011, 2012, Grelet et al., 2016, Lainé et al., 2017, Mineur et al., 2017, Ho et al., 2019, Rienesl et al., 2019), first derivatives of selected spectra values (Savitzky-Golay-Filter) were taken. All further data preparation was done in Rstudio (R Development Core Team, 2008). The first derivative of 212 selected spectra variables were additionally corrected for days in milk (DIM), according to Vanlierde et al. (2015): each first derivative value of the selected spectra was multiplied by a constant (i.e., 1), a linear ($\sqrt{3} * x$) and a quadratic [$\sqrt{5/4} * (3x^2 - 1)$] modified Legendre polynomial (Gengler et al., 1999), where $x = -1 + 2[(DIM - 3)/(305 - 3)]$. This modification resulted in 636 (212 constant, 212 linear, 212 quadratic) spectra variables, which were finally used as predictor variables. The complete data set (403,863 test day records) was randomly split by farm and pregnancy state into half a calibration (training) set and a validation (test) set. Further, calibration set got balanced (1:1) in terms of pregnancy state by using random down sampling. The validation data set was kept unbalanced to get realistic conditions. For every test day record two additional variables were introduced: 'days pregnant' (test day date *minus* date of successful insemination) and 'days after insemination' (test day date *minus* date of latest insemination). The variable 'days after insemination' was needed to define expected pregnancy stage of a cow at a certain test day, as in validation we assumed that true pregnancy stage is unknown.

2.2   Methodology

Two different approaches of predicting pregnancy status were explored, considering the potential effects of stage of lactation and stage of pregnancy on milk composition and resulting MIR spectra patterns.

Approach 1: Single prediction model across the whole lactation and gestation

The first approach was to develop a single prediction model for all test day records, regardless lactation and gestation stage of the cow at the respective test day, similar to the study of Lainé et al. (2014).

Approach 2: Separate prediction models for each different (expected) pregnancy and lactation stage

The second approach was to produce separate models for different pregnancy and lactation stages (DIM). Test day records were clustered into 24 classes according to true or expected pregnancy stage and DIM (Table 2). In calibration clustering for 'pregnant' cases was done by true pregnancy stage ('days pregnant') and for 'open' cases by expected pregnancy stage ('days after insemination'). In validation, both 'open' and 'pregnant' cases were grouped by the variable 'days after insemination'. For each class a separate discriminant model was developed and evaluated in the respective class in validation.

Model predictions were done with Partial Least Square Discriminant Analysis (PLS-DA) by applying the function 'trainControl' of R package 'caret' (Kuhn, 2008). A 10-fold cross validation was used to fine tune the model, the number of components was set automatically (within a maximum number of 60 to avoid overfitting) for every run and discrimination was done by class probabilities. Spectra values were centered and scaled. Indicators of model fit were sensitivity (proportion of pregnant cases correctly assigned as pregnant), specificity (proportion of open cases correctly assigned as open), balanced accuracy (mean of sensitivity and specificity) and Area Under Receiver Operating Characteristic Curve (AUC). Model performance was evaluated with an external validation. The validation data set consisted only of data from farms which were not included in model building. The results presented below are means of 5 independent replicates per setting. The standard deviations of indicators of model fit were typically very low in approach 1 (0.001 to 0.003) and from 0.001 to 0.050 in approach 2, depending on sample size in the respective class.

Data sets, data processing and methodology were very similar to a study on mastitis detection from MIR spectroscopy of Rienesl et al. (2019), carried out within the framework of the same project.

Table 2 Classes used to cluster test day records according to different (true or expected) pregnancy and lactation stages (days in milk) for approach 2

	Lactation stage (days in milk)				
	3 - 60	61 - 120	121 - 180	181 - 240	241 - 305
(true or expected) pregnancy stage in days	1 - 30	1 - 30	1- 30	1 - 30	1- 30
	31 - 60	31 - 60	31 - 60	31 - 60	31 - 60
	-	61 - 90	61 - 90	61 - 90	61 - 90
	-	91 - 120	91 - 120	91 - 120	91 - 120
	-	-	151 - 180	181 - 210	181 - 210
	-	-	-	211 - 240	211 - 240
	-	-	-	-	241 - 305

3   Results and discussion

In this study, we examined 2 approaches to predict the pregnancy status of dairy cows from routinely recorded MIR spectra. The results of the first approach, a single prediction model for all cows and test day records, regardless lactation and gestation stage, are displayed in Table 3. Both sensitivity (0.86) and specificity (0.84) were almost identical in calibration and validation. The value of AUC was 0.928 which indicates an outstanding performance of the prediction model according to Lantz (2015). In the study of Lainé et al. (2014) sensitivity was higher (>0.99) and specificity (>0.84) almost the same as in the present study. However, the results are difficult to compare because of differences in methodology. In the reported approach of Delhez et al. (2020), who used single spectra after insemination, prediction accuracies were lower than our results. Even if differences are expected given the studied populations and countries (Australia vs. Austria), the definition of 'open' was also different, which may have a considerable effect.

Table 3 Indicators of model fit in calibration and validation using a single prediction equation across the whole lactation and gestation length (approach 1)

	Sensitivity	Specificity	Balanced accuracy	AUC
Calibration	0.857	0.837	0.847	-
Validation	0.856	0.836	0.846	0.928

AUC – Area Under Receiver Operating Characteristic Curve

To get closer insight on how prediction worked, the results of approach 1 were split up for different lactation stages (Table 4).

Table 4 Indicators of model fit (sensitivity, specificity and balanced accuracy) in validation split up for lactation stage for the model with only one prediction model (approach 1)

Lactation stage (days in milk)	Records, *n*		Sensitivity	Specificity	Balanced accuracy
	open	pregnant			
1 (3 - 60)	35,140	2,515	0.000	1.000	0.500
2 (61 - 120)	17,756	23,111	0.285	0.895	0.590
3 (121 -180)	6,003	35,325	0.974	0.086	0.530
4 (181 - 240)	1,998	39,549	0.999	0.002	0.501
5 (241 - 305)	684	38,915	0.999	0.000	0.500

This procedure showed an immense imbalance in sensitivity and specificity for the 5 different lactation stages. In lactation stage 1 most of the cows (2,515 out of 37,655) were open. This was expected as in our data cows got pregnant on average after 93 DIM. Sensitivity in the first lactation stage was 0.000 and specificity was 1.000. Hence, the model classified all open cases correctly as open, but was not able to classify a pregnant cow correctly as pregnant. Precisely, the model predicted only a single

cow out of 2,515 actually pregnant cows in lactation stage one to be pregnant. In lactation stage 2 the number of test day records was much more balanced in terms of open (17,756) and pregnant cases (23,111). Though, the majority of cow's test day records (32,428) were predicted to be open and only 8,439 were predicted to be pregnant, which resulted in a sensitivity of 0.285 and specificity of 0.895. In lactation stage 3 the proportion of open cases decreased strongly (<15%) and also sensitivity (0.975) and specificity (0.086) changed dramatically compared to previous lactation stages. Similar results were found in lactation stage 4. In lactation stage 5, where the number of open cases was only 1.7%, sensitivity was 0.999 and specificity 0.000. According to those results it can be concluded that the model was not able to predict pregnant cases before the third month of lactation and vice versa not able to predict non-pregnancy after the third month of lactation. Moreover, we assume that the model was predicting the lactation stage to a quite high degree, which is strongly linked to pregnancy state. The sensitivity of approach 1 was additionally split up for different months of pregnancy (Table 5). We observed a very low sensitivity (0.380) in the first months of pregnancy and a moderate sensitivity (0.695) in the second month. In the third month of pregnancy sensitivity increased strongly to 0.95 and was above 0.99 from month 4 onward.

Table 5 Sensitivity in validation split up for pregnancy months for the model with only one prediction model (approach 1)

Pregnancy month	Records, *n* pregnant	Sensitivity	Pregnancy month	Records, *n* pregnant	Sensitivity
1	20,271	0.380	6	17,059	0.999
2	20,406	0.695	7	14,350	1.000
3	20,036	0.945	8	8,371	0.999
4	19,646	0.997	9	525	1.000
5	18,750	0.999	10	1	1.000

Consequently, a single prediction model for the full lactation and gestation lengths was not sufficient. This led us to approach 2, where separate prediction models for different lactation stages and (expected) pregnancy stages were developed. The results are displayed in Table 6. The sensitivities were in the range from 0.494 (lactation stage 3, pregnancy day 1 to 30) and 0.995 (lactation stage 4, pregnancy day 211 to 240). Thus, differences were very big. The range of specificities was also very wide, from 0.512 (lactation stage 2, pregnancy day 1 to 30) to 0.884 (lactation stage 3, pregnancy day 121 to 180). Within lactation stage, the indicators of model fit increased in later (expected) gestation stages. For example, in the first lactation stage, sensitivity increased from 0.890 for the first (expected) pregnancy month to 0.946 for the second (expected) pregnancy month and specificity from 0.690 to 0.750. In lactation stage 4, sensitivity increased from 0.553 to 0.995 and specificity from 0.527 to 0.820. This finding partially coincides with the results of Delhez et al. (2020), who reported sensitivity and specificity greater than 0.74 from the 180th day of pregnancy and lower (sensitivities 0.57 to 0.71; specificities 0.52 to 0.66) in earlier pregnancy stages.

Within the classes of (expected) pregnancy days, indicators of model fit mostly decreased with a later lactation stage. Sensitivities for the first 'expected' pregnancy months were 0.890 in the first lactation stage, 0.620 in the second lactation stage, and 0.504 in the last lactation stage. These results indicate that prediction accuracy decreases for cows which become pregnant (much) later than the average cow in the data set.

In general, results of approach 2 are very hard to compare with other mentioned studies. In the study of Delhez et al. (2020), the authors explored a strategy with 7 different modelling groups based on stages (days) after insemination, but regardless lactation stage, which was novel in our study.

Faculty of Agrobiology and Food Resources

Table 6 Records (open/pregnant) and indicators of model fit in validation for the different (expected) pregnancy and lactation classes (approach 2)

Lactation stage (days in milk)	(expected) pregnancy days	Records, *n*		Sensitivity	Specificity	Balanced accuracy	AUC
		open	pregnant				
1 (3 - 60)	1 - 30	35,158	2,497	0.890	0.690	0.790	0.853
	31 - 60	32,479	37	0.946	0.750	0.848	0.919
2 (61 - 120)	1 - 30	15,023	11,421	0.620	0.512	0.566	0.592
	31 - 60	10,992	9,575	0.827	0.623	0.725	0.792
	61 - 90	8,780	2,438	0.922	0.840	0.881	0.937
	91 - 120	8,388	43	0.969	0.857	0.913	0.974
3 (121 - 180)	1 - 30	3,647	4,303	0.494	0.553	0.523	0.534
	31 - 60	2,819	7,791	0.565	0.590	0.577	0.606
	61 - 90	2,037	11,555	0.597	0.625	0.611	0.658
	91 - 120	1,318	9,074	0.772	0.676	0.724	0.798
	121 - 150	1,324	2,367	0.897	0.834	0.866	0.929
	151 - 180	1,157	45	0.951	0.884	0.918	0.964
4 (181 - 240)	1 - 30	990	1,582	0.553	0.527	0.540	0.559
	31 - 60	771	2,729	0.592	0.597	0.594	0.636
	61 - 90	573	4,606	0.607	0.622	0.614	0.660
	91 - 120	459	7,918	0.587	0.652	0.620	0.666
	121 - 150	380	11,407	0.623	0.623	0.623	0.674
	151 - 180	288	9,506	0.766	0.718	0.742	0.821
	181 - 210	289	2,477	0.896	0.846	0.871	0.939
	211 - 240	279	48	0.995	0.820	0.908	0.958
5 (241 - 305)	1 - 30	307	559	0.504	0.584	0.545	0.568
	31 - 60	256	938	0.598	0.537	0.568	0.612
	61 - 90	155	1,643	0.623	0.633	0.628	0.675
	91 - 120	155	2,759	0.610	0.706	0.658	0.713
	121 - 150	107	4,557	0.612	0.676	0.644	0.691
	151 - 180	93	8,153	0.638	0.725	0.681	0.742
	181 - 210	88	11,741	0.684	0.768	0.726	0.792
	211 - 240	83	8,563	0.776	0.849	0.813	0.885
	241 - 305	79	506	0.852	0.855	0.854	0.933

AUC – Area Under Receiver Operating Characteristic Curve

4 Conclusions

This work explored the use of routinely recorded MIR spectral data to predict pregnancy status from dairy cows by developing and evaluating a discriminant model. Results indicate that prediction of pregnancy state is difficult because of the strong effect of lactation stage on the MIR spectrum and the fact that cows are typically open in early lactation and pregnant in late lactation. If a cow did not match this "common pattern", prediction was not satisfactory. Hence, a single prediction model for the full lactation and all periods of gestation was not sufficient. Developing separate prediction equations for stages of lactation and periods of expected pregnancy improved predictability of pregnancy status to some degree. Whether the prediction accuracies found in this study will be sufficient to provide farmers with an additional tool for fertility management needs to be explored in discussion with farmers and breeding organizations.

Acknowledgments

References

Balhara, A. K., Gupta, M., Singh, S., Mohanty, A. K., & Singh, I. (2013). Early pregnancy diagnosis in bovines: Current status and future directions. *The Scientific World Journal*, *2013*. https://doi.org/10.1155/2013/958540

Bekele, N., Addis, M., Abdela, N., & Ahmed, W. M. (2016). Pregnancy Diagnosis in Cattle for Fertility Management: A Review. *Global Veterinaria*, *16*(4), 355–364. https://doi.org/10.5829/idosi.gv.2016.16.04.103136

Benedet, A., Franzoi, M., Penasa, M., Pellattiero, E., & De Marchi, M. (2019). Prediction of blood metabolites from milk mid-infrared spectra in early-lactation cows. *Journal of Dairy Science*, *102*(12), 11298–11307. https://doi.org/10.3168/jds.2019-16937

Delhez, P., Ho, P. N., Gengler, N., Soyeurt, H., & Pryce, J. E. (2020). Diagnosing the pregnancy status of dairy cows: How useful is milk mid-infrared spectroscopy? *Journal of Dairy Science*, *103*(4), 3264–3274. https://doi.org/10.3168/jds.2019-17473

Egger-Danner, C., Fürst, C., Mayerhofer, M., Rain, C., & Rehling, C. (2018). ZuchtData Jahresbericht 2018. Vienna. [Online]. Available at: https://zar.at/Downloads/Jahresberichte/ZuchtData-Jahresberichte.html. [Accessed: 2020, May 15].

Gengler, N., Tijani, A., Wiggans, G. R., & Misztal, I. (1999). Estimation of (Co)variance function coefficients for test day yield with a expectation-maximization restricted maximum likelihood algorithm. *Journal of Dairy Science*, *82*(8), 1849.e1-1849.e23. https://doi.org/10.3168/jds.S0022-0302(99)75417-2

Grelet, C., Fernández Pierna, J. A., Dardenne, P., Baeten, V., & Dehareng, F. (2015). Standardization of milk mid-infrared spectra from a European dairy network. *Journal of Dairy Science*, *98*(4), 2150–2160. https://doi.org/10.3168/jds.2014-8764

Grelet, C., Bastin, C., Gelé, M., Davière, J. B., Johan, M., Werner, A., Reding, R., Fernandes Pierna, J. A., Colinet, F. G., Dardenne, P., Gendler, N., Soyeurt, H. & Dehareng, F. (2016). Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. *Journal of Dairy Science*, *99*(6), 4816–4825. https://doi.org/10.3168/jds.2015-10477

Hirpa, A., Yehualaw, B., Wube, A., Asnake, A., Jemberu, A., Medicine, V., & Box, P. O. (2018). Review on Pregnancy Diagnosis in Dairy Cows, *9*(2), 45–55. https://doi.org/10.5829/idosi.jri.2018.45.55

Ho, P. N., Bonfatti, V., Luke, T. D. W., & Pryce, J. E. (2019). Classifying the fertility of dairy cows using milk mid-infrared spectroscopy. *Journal of Dairy Science*. https://doi.org/10.3168/jds.2019-16412

Humblot, P. (2001). Monitor Pregnancy and Determine the Timing , Frequencies and Sources of Embryonic Mortality in Ruminants. *Theriogenology*, *56*(01), 1417–1433.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5), 1–26.

Lainé, A., Bel Mabrouk, H., Dale, L. M., Bastin, C., & Gengler, N. (2014). How to use mid-infrared spectral information from milk recording system to detect the pregnancy status of dairy cows. *Communications in Agricultural and Applied Biological Sciences*, *79*(1), 33–38.

Lainé, A., Bastin, C., Grelet, C., Hammami, H., Colinet, F. G., Dale, L. M., Gillon, A., Vandenplas, J., Deharend, F. & Gengler, N. (2017). Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. *Journal of Dairy Science*, *100*(4), 2863–2876. https://doi.org/10.3168/jds.2016-11736

Lantz, B. (2015). *Machine Learning with R. Machine Learning* (Second Edi). Packt Publishing Ltd. https://doi.org/10.1002/9781119642183.ch14

Mineur, A., Köck, A., Grelet, C., Gengler, N., Egger-Danner, C., & Sölkner, J. (2017). First Results in the Use of Milk Mid-infrared Spectra in the Detection of Lameness in Austrian Dairy Cows Genomic evaluation View project MACSUR View project. *Agriculturae Conspectus Scientifi Cus*, *Vol. 82*(No. 2 (163-166)), (163-166). Retrieved from https://www.researchgate.net/publication/325450513

Olori, V. E., Brotherstone, S., Hill, W. G., & McGuirk, B. J. (1997). Effect of gestation stage on milk yield and composition in Holstein Friesian dairy cattle. *Livestock Production Science*, *52*(2), 167–176. https://doi.org/10.1016/S0301-6226(97)00126-7

Pohler, K. G., Franco, G. A., Reese, S. T., Dantas, F. G., Ellis, M. D., & Payton, R. R. (2016). Past, present

and future of pregnancy detection methods. *Applied Reproductive Strategies in Beef Cattle 7-8 September 2016*, 251–259.

Rienesl, L., Khayatzadeh, N., Köck, A., Dale, L., Werner, A., Grelet, C., Gengler, N., Auer, F-J., Egger-Danner, C., Massart, X. & Sölkner, J. (2019). Mastitis detection from milk mid-infrared (MIR) spectroscopy in dairy cows. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *67*(5), 1221–1226. https://doi.org/10.11118/actaun201967051221

Santos, J. E. P., Thatcher, W. W., Chebel, R. C., Cerri, R. L. A., & Galvão, K. N. (2004). The effect of embryonic death rates in cattle on the efficacy of estrus synchronization programs. *Animal Reproduction Science*, *82–83*, 513–535. https://doi.org/10.1016/j.anireprosci.2004.04.015

SAS Institute Inc. (2017). SAS software 9.4. *SAS Institute Inc., Cary, NC, USA.*

Soyeurt, H., Dehareng, F., Gengler, N., McParland, S., Wall, E., Berry, D. P., Coffey, P. & Dardenne, P. (2011). Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *Journal of Dairy Science*, *94*(4), 1657–1667. https://doi.org/10.3168/jds.2010-3408

Soyeurt, H., Bastin, C., Colinet, F. G., Arnould, V. M.-R., Berry, D. P., Wall, E., Dehareng, F., Nguyen, H. N., Pardenne, P., Schefers, J., Vandenplas, J., Weigel, K., Coffey, M., Théron, L., Detilleux, J., Reding, E., Gengler, N. & McParland, S. (2012). Mid-infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis. *Animal*, *6*(11), 1830–1838. https://doi.org/10.1017/s1751731112000791

Toffanin, V., De Marchi, M., Lopez-Villalobos, N., & Cassandro, M. (2015). Effectiveness of mid-infrared spectroscopy for prediction of the contents of calcium and phosphorus, and titratable acidity of milk and their relationship with milk quality and coagulation properties. *International Dairy Journal*, *41*, 68–73. https://doi.org/10.1016/j.idairyj.2014.10.002

Vanlierde, A., Vanrobays, M.-L., Dehareng, F., Froidmont, E., Soyeurt, H., McParland, S., S., Lewis, E., Deighton, M. H., Grandl, F., Kreuzer, M., Gredler, B., Dardenne, P. & Gengler, N. (2015). Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra. *Journal of Dairy Science*, *98*(8), 5740–5747. https://doi.org/10.3168/jds.2014-8436

Vanlierde, A., Soyeurt, H., Gengler, N., Colinet, F. G., Froidmont, E., Kreuzer, M., Grandl, F., Bell, M., Lund, P., Olijhoek, D. W., Eugéne M., Martin, C., Kuhla, B. & Dehareng, F. (2018). Short communication: Development of an equation for estimating methane emissions of dairy cows from milk Fourier transform mid-infrared spectra by using reference data obtained exclusively from respiration chambers. *Journal of Dairy Science*, *101*(8). https://doi.org/10.3168/jds.2018-14472

Faculty of Agrobiology and Food Resources