

## Article

# Mid-Infrared Spectroscopy for Predicting Goat Milk Coagulation Properties

Arianna Goi <sup>1</sup>, Silvia Magro <sup>1,\*</sup>, Luigi Lanni <sup>2</sup>, Carlo Boselli <sup>2</sup> and Massimo De Marchi <sup>1</sup>

<sup>1</sup> Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy; arianna.goi@unipd.it (A.G.); massimo.demarchi@unipd.it (M.D.M.)

<sup>2</sup> Istituto Zooprofilattico Sperimentale del Lazio E Della Toscana "M. Aleandri"—National Reference Centre for Ovine and Caprine Milk and Dairy Products Quality, Via Appia Nuova 1411, 00178 Rome, Italy; luigi.lanni@izslt.it (L.L.); carlo.boselli@izslt.it (C.B.)

\* Correspondence: silvia.magro.1@studenti.unipd.it

## Abstract

The assessment of milk coagulation properties (MCPs) is crucial for enhancing goat cheese production and quality. In this study, 501 bulk goat milk samples were collected from various farms to evaluate the MCPs. Traditionally, cheesemaking aptitude is evaluated using lactodynamographic analysis, a reliable but time-consuming laboratory method. Mid-infrared spectroscopy (MIRS) offers a promising alternative for the large-scale prediction of goat milk's technological traits. Reference MCP measurements were paired with mid-infrared spectra, and prediction models were developed using partial least squares regression, with accuracy evaluated through cross- and external validation. The ability of MIRS to classify milk samples by coagulation aptitude was evaluated using partial least squares discriminant analysis. Only the model for rennet coagulation time obtained sufficient accuracy to be applied for screening ( $R^2_{CrV} = 0.68$ ;  $R^2_{Ext} = 0.66$ ; RPD = 2.05). Lower performance was observed for curd-firming time ( $R^2_{CrV} = 0.33$ ;  $R^2_{Ext} = 0.27$ ; RPD = 1.42) and curd firmness ( $R^2_{CrV} = 0.55$ ;  $R^2_{Ext} = 0.43$ ; RPD = 1.35). Classification of high coagulation aptitude achieved balanced accuracy values of 0.81 (calibration) and 0.74 (validation). With further model refinement and larger calibration datasets, MIRS may become a resource for the dairy-goat sector to monitor and improve milk suitability for cheesemaking.

**Keywords:** mid-infrared spectroscopy; milk; goat; technological properties



Academic Editor: Rajka Božanić

Received: 6 June 2025

Revised: 1 July 2025

Accepted: 4 July 2025

Published: 7 July 2025

**Citation:** Goi, A.; Magro, S.; Lanni, L.; Boselli, C.; Marchi, M.D. Mid-Infrared Spectroscopy for Predicting Goat Milk Coagulation Properties. *Foods* **2025**, *14*, 2403. <https://doi.org/10.3390/foods14132403>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The demand for milk and dairy products has consistently increased on a global scale, with a particularly notable rise in developing countries [1]. Although goat milk production represents a relatively small share of the global dairy market, following cow and buffalo milk [2], it is continuously growing.

The majority of goat milk is dedicated to cheese production, with most being consumed within the country of origin. The efficiency of cheese production and the quality of the final product largely depend on the compositional properties of the milk and its cheesemaking aptitude; that is, its ability to coagulate upon the addition of a clotting agent [3–5]. The parameters that define the coagulation process and the cheesemaking potential of milk are typically evaluated using lactodynamographic analysis, collectively referred to as milk coagulation properties (MCPs). These properties include (i) rennet coagulation time (RCT), which represents the interval (min) between the addition of the clotting agent and the onset of coagulation, (ii) curd-firming time ( $k_{20}$ ), indicating the time (min) needed to reach a

curd firmness of 20 mm, and (iii) curd firmness ( $a_{30}$ ), defined as the coagulum firmness (mm) 30 min after the addition of the clotting agent [6]. Milk with good suitability for dairy processing is typically characterized by short RCT and  $k_{20}$  values, along with a relatively high  $a_{30}$  value. Generally, goat milk exhibits lower RCT and  $k_{20}$  values, and higher  $a_{30}$  values than cow milk [7]. This makes goat milk more advantageous for cheesemaking, as faster coagulation ensures more efficient processing, and rapidly reaching curd firmness indicates a firmer curd, which benefits both texture and yield [8,9]. However, goat milk shows a rapid decrease in curd firmness after reaching its maximum value [7].

Although the reference method provides reliable results, it remains impractical for large-scale milk recording programs. In fact, it requires considerable time, involves high costs, and depends on trained personnel to perform a labor-intensive analytical protocol [10]. To overcome these limitations, other approaches should be explored. Infrared spectroscopy, a rapid, cost-effective, and easily implementable technology widely used for food analysis worldwide, could serve as a potential alternative for this purpose. Specifically, mid-infrared spectroscopy (MIRS) exploits the vibrational absorption of infrared radiation by molecular bonds within matter in the spectral range of 2500–25000 nm (5000 to 900  $\text{cm}^{-1}$ ), enabling the quantification of constituent concentrations. The accuracy of this technique is dependent on calibration models, sample conditions, and other contributing factors [11]. MIRS is particularly recognized in the dairy sector, as it is employed for the routine official analysis of liquid milk composition traits, including protein, casein, fat, lactose, and urea contents [12]. Given its widespread use, extending MIRS to assess coagulation properties would require minimal additional costs, offering an efficient and economical solution for large-scale screening. Moreover, studies have demonstrated its application in predicting the MCPs of individual bovine and ovine samples with sufficient or moderate accuracy [13,14], further supporting its potential for implementation in the dairy industry. By contrast, lower predictive accuracies have been reported in the literature for individual goat milk samples [15,16].

Despite the relevance of bulk milk in routine collection and processing, to the best of our knowledge, the use of MIRS to predict the coagulation properties of bulk goat milk has not yet been investigated. This lack of information limits our understanding of how this technique might perform under real-world conditions, where milk from multiple animals is pooled together. Addressing this gap is essential for assessing the practical applicability of MIRS in cheesemaking-oriented strategies and enhancing the technological valorization of goat milk.

The objectives of the present study were (1) to evaluate the accuracy of MIRS in predicting the MCPs of bulk milk samples from goats and (2) to assess the discriminant ability of spectra for the identification of samples with high coagulation aptitude.

## 2. Materials and Methods

### 2.1. Bulk Milk Sampling and Reference Standard Analysis

A total of 501 bulk milk samples (50 mL without preservative) were collected from 2021 to 2025 from goat farms located in central Italy, which reared different breeds, including Saanen, Alpine, Maltese, Murciano-Granadina, Sarda, Ciociara Grigia, and crossbreeds. The milking frequency on the farms ranged from once to twice per day, and some farms implemented out-of-season kidding practices to ensure year-round milk production. Bulk samples were collected from farm tanks, where milk from multiple goats was routinely pooled immediately after milking. Samples were obtained within the official milk quality system and transported to the laboratory under refrigeration at 4 °C. An analysis of samples was conducted within 36 h of collection at the quality milk laboratory of the Istituto Zooprofilattico Sperimentale del Lazio e della Toscana “M. Aleandri” (Rome, Italy). This

laboratory is accredited by Accredia, the Italian Accreditation Body (Laboratory No. 0201A), and operates in compliance with the ISO/IEC 17025:2017 standards of the International Organization for Standardization. The chemical composition of milk (i.e., fat, protein, casein, and lactose content) was quantified using a MilkoScan<sup>TM</sup> 7 RM (Foss Analytical A/S, Hillerød, Denmark) calibrated with the commercial goat standards obtained from the Milk Standard Laboratory of the Associazione Italiana Allevatori (Maccarese, Italy); the resulting spectrum of each sample, containing 1060 transmittance data points, was stored. The somatic cell count (SCC) was assessed using a Fossomatic<sup>TM</sup> FC system (Foss Analytical A/S, Hillerød, Denmark), and then converted to a somatic cell score using the formula of Ali and Shook [17]:  $SCS = 3 + \log_2 (SCC/100,000)$ , while the MCPs ( $RCT$ ,  $k_{20}$ , and  $a_{30}$ ) were measured using a Formagraph LDG 2.0 (Ma.Pe System Srl, Firenze, Italy). For the coagulation trait analysis, sample preparation involved heating 10 mL of milk to 36 °C, followed by the addition of calf rennet (200 µL) composed of 75% chymosin and 25% pepsin (175 international milk clotting units/mL; Clerici S.p.A., Sacco srl, Cadorago, Italy) diluted to a 1.6% ( $w/v$ ) concentration in distilled water. At the beginning of the analysis, an oscillating loop pendulum was placed in contact with the milk. As coagulation initiates, the viscosity of the milk increases, generating a resistance that is transmitted back to the loop. Consequently, the instrument produces a graphical representation of the curd firmness over time. Each measurement was performed within 30 min after adding the enzyme.

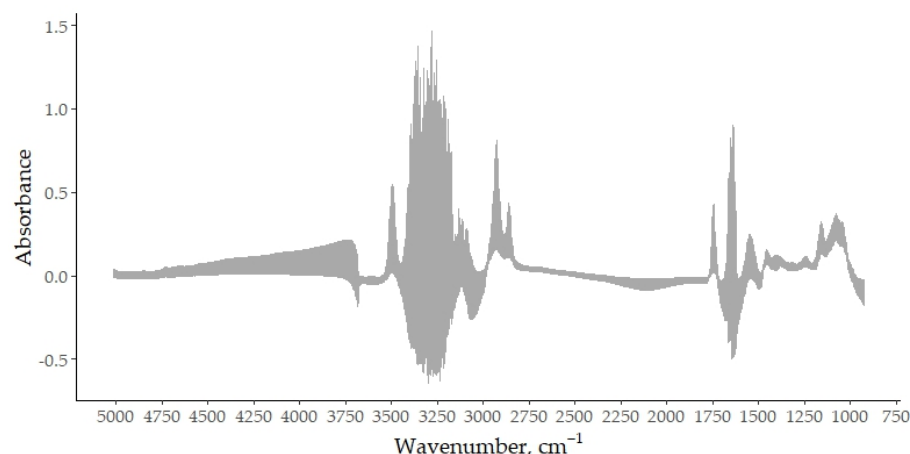
An index of milk aptitude to coagulate (IAC) standardized to mean = 100 and SD = 5 was built using the following formula [18]:

$$IAC = 100 + (a_{30} - \text{mean}_{a_{30}})/\text{sd}_{a_{30}} \times 2.5 - (RCT - \text{mean}_{RCT})/\text{SD}_{RCT} \times 2.5 \quad (1)$$

where  $\text{mean}_{a_{30}}$  and  $\text{mean}_{RCT}$  were the dataset averages of  $a_{30}$  and  $RCT$ , respectively, and  $\text{SD}_{a_{30}}$  and  $\text{SD}_{RCT}$  were their standard deviations.

## 2.2. Development of the MIRS Prediction Models

As a first step in preparing the spectral data for analysis, the raw mid-infrared spectra of all goat bulk milk samples were transformed into absorbance data by taking the  $\log_{10}$  of the reciprocal of the transmittance. This conversion allowed for the overall unprocessed spectral structure to be visualized and helped to identify the regions dominated by water absorption. Figure 1 shows the complete set of raw absorbance spectra prior to any exclusion or transformation, providing a comprehensive overview of the spectral variability within the dataset. In particular, the region between 3700 and 3000  $\text{cm}^{-1}$  is characterized by broad and intense bands mainly attributed to the O–H stretching from water. The interval from 3000 to 2800  $\text{cm}^{-1}$  includes bands related to the C–H stretching, primarily arising from lipids, as well as contributions from the region around 1800 and 1700  $\text{cm}^{-1}$  [19]. The range between 1700 and 1600  $\text{cm}^{-1}$  contains both the H–O–H bending vibration and the amide I band, mainly associated with the C=O stretching of the peptide bonds. The region from 1600 to 1200 includes the absorption bands corresponding to amide II and amide III, arising from protein and attributed to the C–N stretching vibrations in combination with the N–H bending, along with  $\text{CH}_2$  scissoring vibrations from lipid acyl chains. The range from 1200 to 900  $\text{cm}^{-1}$  is dominated by the C–O stretching primarily associated to lactose and the P=O vibrations from phosphate-containing compounds [20]. This preliminary inspection of the spectra guided the subsequent removal of specific wavelength intervals that are known to be affected by water interference.



**Figure 1.** Raw mid-infrared absorbance spectra of bulk goat milk samples, recorded in the spectral range from 5000 to 900  $\text{cm}^{-1}$ , with a resolution of 1  $\text{cm}^{-1}$  (1060 total wavenumbers).

To minimize the inclusion of random noise, spectral regions related to water absorption (5011 to 2974  $\text{cm}^{-1}$ , 2503 to 1929  $\text{cm}^{-1}$ , 1712 to 1585  $\text{cm}^{-1}$ , and 964 to 925  $\text{cm}^{-1}$ ) were excluded prior to the chemometric analysis in accordance with the standard practice reported by Grelet et al. [21]. To optimize the calibration accuracy, spectral outliers were first identified and removed based on the Mahalanobis distance (Global  $H > 3.0$ ). The spectra were then coupled with the reference values of RCT,  $k_{20}$ , and  $a_{30}$  to develop the respective prediction equations. Predictive equations for each MCP trait were built using modified partial least squares (mPLS) regression in the WinISI 4.10 software (Infrasoft International, Port Matilda, PA, USA). The mPLS regression analysis was followed by the elimination of chemical outliers setting a critical T-value of 3, removing samples whose predicted value deviated more than 3 standard errors of cross-validation from the reference value. Subsequently, a second mPLS was performed, followed by another round of outlier removal. Then, for each trait, samples were ordered by their reference values, and every fourth sample was assigned to the validation set, with the remaining 75% comprising the calibration set. The two subsets were ensured to exhibit similar means and SD for each target trait. Model calibration was thus performed on the calibration subset, internally tested through five-fold cross-validation, and externally validated on the reserved validation set. For the development of predictive models, spectral data were used either in their raw form or subjected to scatter correction techniques, including standard normal variate (SNV), detrending (D), a combination of SNV and D (SNV + D), and multiplicative scatter correction (MSC). These preprocessing techniques were further combined with the following mathematical treatments: 0,0,1,1; 1,4,4,1; 1,8,8,1; and 2,5,5,1 [22]. In these notations, the first digit represents the derivative order, the second indicates the gap over which the derivative is computed, the third specifies the number of data points used in the first smoothing, and the fourth corresponds to the number of data points in the second smoothing [22]. To prevent overfitting, the number of latent variables in the mPLS model was selected by minimizing the root mean square error computed at each cross-validation iteration. The optimal prediction equation was determined based on the coefficient of determination for cross-validation ( $R^2_{\text{CrV}}$ ) and external validation ( $R^2_{\text{Ext}}$ ), as well as the ratio of performance to deviation (RPD) during external validation. The variable importance in projection (VIP) scores from the final round of mPLS were retained for analysis. A *t*-test was performed using the TTEST procedure of SAS software v. 9.4 (SAS Institute Inc., Cary, NC, USA) to assess whether the bias, calculated as the average difference between the predicted and reference values, did not statistically differ from zero. The null hypothesis was tested at a significance level of 0.05, and *p*-values were used to determine the statistical significance.

### 2.3. Discriminant Analysis

The IAC variable was transformed into a binary trait using the mean value (100) as the threshold. Samples with  $IAC > 100$  were classified as having a ‘high aptitude to coagulate’, while those with  $IAC \leq 100$  were classified as having a ‘low aptitude to coagulate’. A partial least squares discriminant analysis (PLS-DA) was performed using the ‘caret’ package [23] of R software v 4.4.2. to classify the milk spectra into these two categories. Model tuning was performed using 10-fold cross-validation repeated three times, and the number of components was set automatically but capped at a maximum of 15 to avoid overfitting. The spectral data points were mean-centered and scaled, and discrimination was performed based on class probabilities. The PLS-DA performance included sensitivity, specificity, positive predictive values, negative predictive values, balanced accuracy, and area under the curve (AUC) in both the calibration and validation. Balanced accuracy is the mean of sensitivity and specificity, and positive and negative predictive values are the proportions of positive and negative results that are true positive and true negative, respectively. For each metric, a 95% confidence interval was also reported.

## 3. Results and Discussion

### 3.1. Descriptive Statistics

The descriptive statistics of the MCPs and bulk milk chemical composition are shown in Table 1. The bulk milk samples used in the present study had fat, protein, and casein contents of 4.22%, 3.44%, and 2.64%, respectively. The average lactose content was 4.34%, whereas the average SCS was 6.77 (Table 1). The chemical composition of the evaluated milk samples was consistent with previous studies on bulk [8] and individual milk [16,24] of goats of the same breeds used in this study and reared in Italy.

**Table 1.** Descriptive statistics of milk coagulation traits and chemical composition for goat bulk milk.

Trait <sup>1</sup>	N <sup>2</sup>	Mean	SD	Range	CV (%)
Milk coagulation properties					
RCT, min	501	9.59	2.94	1.00–26.30	30.70
k <sub>20</sub> , min	456	3.10	1.74	0.52–11.15	44.30
a <sub>30</sub> , mm	501	29.78	11.96	2.90–62.62	40.16
IAC	501	100.00	3.48	84.06–113.06	3.28
Milk quality traits					
Fat, %	476	4.22	1.15	1.06–12.1	27.31
Protein, %	479	3.44	0.54	2.26–6.41	15.76
Casein, %	479	2.64	0.59	1.39–6.23	22.40
Lactose, %	479	4.34	0.31	3.37–6.04	7.14
SCS	479	6.77	1.34	0.88–11.07	19.79

<sup>1</sup> RCT, rennet coagulation time; k<sub>20</sub>, curd-firming time; a<sub>30</sub>, curd firmness; IAC, index of milk aptitude to coagulate, defined as  $IAC = 100 + (a_{30} - \text{mean}_{a_{30}}) / \text{sd}_{a_{30}} \times 2.5 - (RCT - \text{mean}_{RCT}) / \text{SD}_{RCT} \times 2.5$ ; SCS, somatic cell score.

<sup>2</sup> N, number of samples prior to reference outliers’ elimination.

Regarding the MCPs, in this study, the RCT and k<sub>20</sub> averaged 9.59 min and 3.10 min, while the a<sub>30</sub> averaged 29.78 mm (Table 1). A total of 8.98% of the samples did not coagulate during the analysis, and the lactodynamographic curve did not reach 20 mm. Consequently, for these samples, the k<sub>20</sub> was considered missing data. Pazzola et al. [8] reported higher MCP values than those found in the present study, based on 432 bulk milk samples from 161 commercial goat farms with different breeds in Sardinia (Italy). Specifically, they reported RCT, k<sub>20</sub>, and a<sub>30</sub> values of 13.20 min, 4.91 min, and 32.10 mm, respectively, in an extensive farming system, and 15.0 min, 4.50 min, and 32.80 mm, respectively, in an intensive farming system. Stocco et al. [16] also reported greater MCP values than those observed in the present study, based on 611 individual goat milk samples from



different breeds (Alpine, Murciano-Granadina, Maltese, and Sarda) reared on 19 farms in Sardinia (Italy).

The chemical composition and the MCPs exhibited moderately high variability, likely due to the fact that the milk was bulk milk, originating from a multitude of farms and a diverse range of sampled breeds. In particular, the coefficient of variation (CV) for the MCPs ranged from 31% (RCT) to 44% ( $k_{20}$ ). A high degree of data variation is desirable when developing the MIRS prediction models for use in population-level phenotyping [15].

### 3.2. Predicting Milk Coagulation Properties Using MIRS

The prediction statistics of the models are presented in Table 2. The outliers identified and removed during model development were 6.0% for RCT, 7.5% for  $k_{20}$ , and 7.2% for  $a_{30}$ . The scatter correction applied was SNV combined with D for RCT, only SNV for  $k_{20}$ , and D for  $a_{30}$ , while the first derivatization order was used for each trait. The highest prediction accuracy was achieved for the RCT, whereas the least accurate model was for the  $k_{20}$ , as confirmed by the RPD values of 2.05 and 1.35, respectively.

**Table 2.** Fitting statistics <sup>1</sup> of modified partial least square regression models using external validation for coagulation properties <sup>2</sup> of goat bulk milk with mid-infrared spectroscopy (MIRS).

Trait	Calibration Set (75%)					Validation Set (25%)				
	LF	SC	MT	R <sup>2</sup> <sub>CrV</sub>	SE <sub>CrV</sub>	Bias	Slope	R <sup>2</sup> <sub>ExV</sub>	SE <sub>P</sub>	RPD <sub>ExV</sub>
RCT, min	11	SNV + D	1,4,4,1	0.68	1.57	−0.12	0.87	0.66	1.58	2.05
$k_{20}$ , min	9	SNV	1,4,4,1	0.33	1.09	−0.01	0.76	0.27	1.13	1.42
$a_{30}$ , mm	8	D	1,4,4,1	0.55	7.90	0.24	0.91	0.43	9.03	1.35

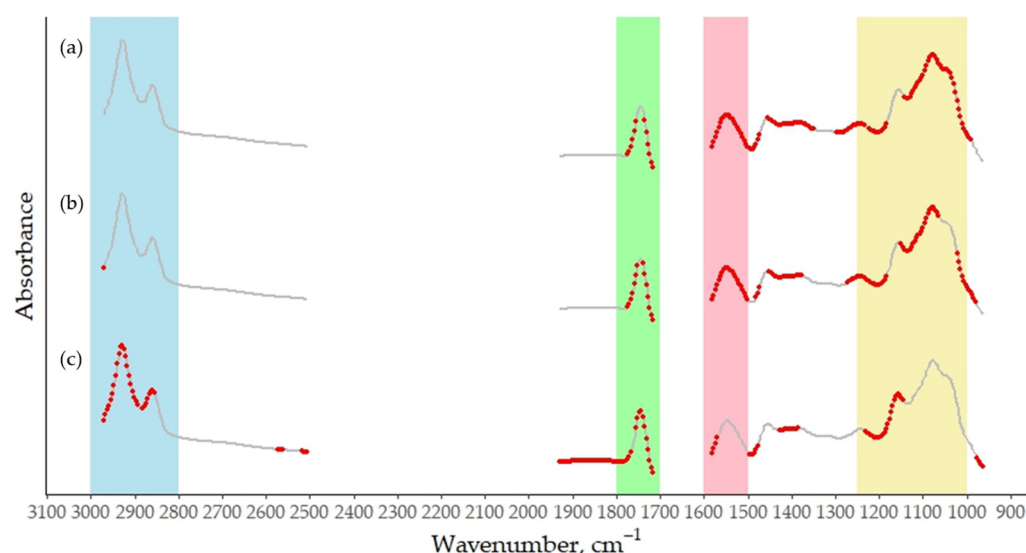
<sup>1</sup> LF, latent factors; SC, scatter correction; MT, mathematical treatment; R<sup>2</sup><sub>CrV</sub>, coefficient of determination of cross-validation; SE<sub>CrV</sub> = standard error of cross-validation; R<sup>2</sup><sub>ExV</sub> = coefficient of determination of external validation; SE<sub>P</sub>, standard error of external validation; RPD = ratio of prediction to deviation calculated as the ratio between the standard deviation of the observed trait and the SE<sub>P</sub>. <sup>2</sup> RCT, rennet coagulation time;  $k_{20}$ , curd-firming time;  $a_{30}$ , curd firmness at 30 min after rennet addition to milk.

Under external validation, the models in the present study yielded R<sup>2</sup><sub>ExV</sub> values of 0.66, 0.27, and 0.43 for RCT,  $k_{20}$ , and  $a_{30}$ , respectively. Dadousis et al. [15] obtained lower R<sup>2</sup> values for RCT and  $a_{30}$  (0.42 and 0.27, respectively) and a comparable value for  $k_{20}$  (0.29) in individual Sarda goat milk. In a multi-breed goat study, Stocco et al. [16] reported R<sup>2</sup> values of 0.42, 0.47, and 0.48, showing better predictive accuracy for  $k_{20}$  and similar accuracy for  $a_{30}$ , but still below the approximate screening benchmark of 0.66 proposed by Williams [25]. Given the limited literature on MIRS-based prediction of the MCPs in goats, parallelism can be drawn from the findings for other dairy species (i.e., cow, sheep, and buffalo). Moderate R<sup>2</sup> values for RCT have also been described in ovine milk (0.59; [14]) and bovine milk (0.69; [13]). When considering the cross-validation approach, the R<sup>2</sup><sub>CrV</sub> value obtained for RCT in this study (0.68) was higher than that reported by Manuelian et al. [26] for individual buffalo milk (0.45), and under cross-validation was also higher than the value reported by Guerra et al. [27] for buffalo bulk milk (0.40), and by Boselli et al. [28] for sheep bulk milk. By contrast, the models developed for the  $k_{20}$  and  $a_{30}$  in the present study showed lower predictive ability than the RCT model, yielding R<sup>2</sup><sub>CrV</sub> values of 0.33 and 0.55, respectively (Table 2). These results are consistent with the cross-validation outcomes reported by Manuelian et al. [26], who obtained R<sup>2</sup><sub>CrV</sub> values of 0.39 for  $k_{20}$  and 0.51 for  $a_{30}$ , and by Guerra et al. [27] for  $k_{20}$  (R<sup>2</sup><sub>CrV</sub> = 0.39). Meanwhile, higher R<sup>2</sup><sub>CrV</sub> values were reported for  $a_{30}$  by Guerra et al. ([27]; 0.61).

The differences in predictive accuracy may reflect methodological, technical, and biological factors. These include the frequency and quality of instrument calibration, size and representativeness of the calibration dataset, the choice and consistency of the spectral preprocessing methods, the chemometric approaches applied, and the rigor of both the

calibration and the validation procedures [16]. Moreover, the biochemical expression of the target traits, the strength of their correlation with milk composition, and the extent to which relevant information is captured by mid-infrared spectral data should be considered. Fat, total protein, and casein fractions—quantified by MIRS—account for a sizable portion of bovine milk coagulation variability [6,7,29]. Studies on goat milk indicate that coagulation properties are less strongly associated with infrared-detectable components. In fact, in goats, breed effects are pronounced, and factors such as ionic calcium and pH contribute more significantly to coagulation behavior than the fat/protein ratio captured by MIRS [15,16,30].

The RPD value closest to 2, which is the commonly accepted threshold for a prediction model to be considered suitable for quantitative screening [31], was that of the RCT. From a methodological perspective, the higher predictive performance observed for the RCT may be partially attributed to its association with the initial enzymatic phase of milk coagulation, which is more directly and consistently reflected in the mid-infrared spectrum of milk. By contrast, traits related to the later stages of curd development involve both chemical and physiomechanical changes, making their spectral signatures in milk more indirect and overlapping, resulting in more complex information to predict [32]. Overall, the predictive performance of the  $k_{20}$  and  $a_{30}$  models remains insufficient to enable the reliable application of MIRS for quantifying these traits in a screening context. The  $R^2$  values in external validation, 0.27 and 0.43, respectively, and the RPD values, 1.42 and 1.35, respectively, fell below the thresholds generally considered acceptable for reliable phenotyping, reflecting poor predictive performance. In particular, the  $R^2$  results indicate that the models capture less than half—and in the case of the  $k_{20}$ , less than one-third—of the variance in the reference data. Therefore, the application of the models for rough large-scale phenotyping or exploratory purposes is currently unfeasible. However, combining MIRS data with additional phenotypic or genomic information could, in the future, yield more reliable prediction models than those based on a single source of information [33]. Nonetheless, the use of these models in routine milk recording or precise decision-making remains premature without further refinement. The development of new trait-specific preprocessing techniques or the use of advanced machine learning algorithms may enhance the predictive capacity of the MIRS models for MCP traits, particularly for complex traits, such as  $k_{20}$  and  $a_{30}$ . The most informative spectral regions for the MCP prediction were identified using the VIP scores (Figure 2), with wavenumbers exhibiting VIP values greater than 1 being considered the most significant [34]. As shown, every trait contributes—albeit to varying degrees—to the fat, protein, and lactose-related bands [19,20]. Notably, only  $a_{30}$  places particular weight on the “fat B” region (3000 to 2800  $\text{cm}^{-1}$ ) [19], which corresponds to the C–H stretching of the methyl group in milk [35]. By contrast, for all the MCPs, high VIP scores were reported at wavenumbers in the “fat A” region (1800 to 1700  $\text{cm}^{-1}$ ), corresponding to changes in the C=O stretching vibration of ester linkage in bovine milk [19].



**Figure 2.** Average mid-infrared absorbance spectrum of bulk goat milk samples used for prediction of (a) RCT, (b)  $k_{20}$ , and (c)  $a_{30}$ , based on 338 selected wavenumbers (from an original set of 1060), after removing regions dominated by water absorption. Red dots indicate the most informative wavenumbers, while the remaining wavenumbers are shown in grey. Regions in blue and green are known as the milk region “Fat B” and “Fat A”, respectively [20]. The pink and yellow band represents the spectral window associated with milk protein and lactose [19].

### 3.3. Discriminating High Milk Coagulation Aptitude Using MIRS

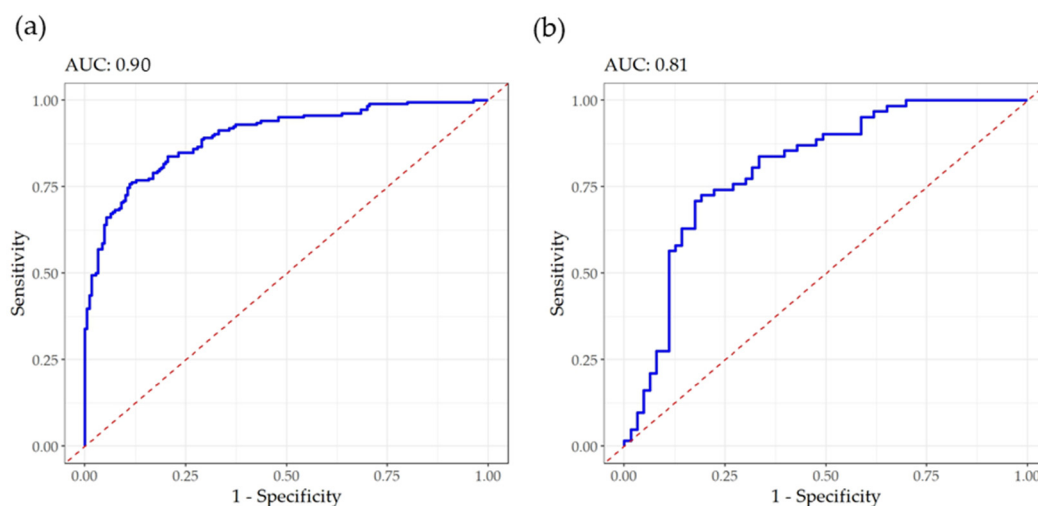
To assess the discriminant ability of the mid-infrared spectra in identifying the samples with high coagulation aptitude, a PLS-DA was performed. In the dataset, 50% of the milk samples had high aptitude to coagulate ( $IAC > 100$ ) (Table 3). Sensitivity, defined as the probability of correctly identifying milk samples with a high aptitude to coagulate ( $IAC > 100$ ), was 0.79 in the calibration and 0.68 in the validation (Table 3). Specificity, which is complementary to sensitivity, represents the probability of correctly identifying milk samples with low aptitude for coagulation ( $IAC \leq 100$ ). This parameter was 0.82 in the calibration and 0.81 in the validation. Martín-Gómez et al. [36] recommended an optimization based on sensitivity and specificity to assess the quality of the PLS-DA models that discriminate between the two classes. The balanced accuracy was 0.81 in the calibration and 0.74 in the validation. Moreover, the AUC values were 0.90 in calibration and 0.80 in validation (Figure 3), suggesting that the model has good discriminative ability during training and maintains an acceptable performance when applied to new data.

**Table 3.** Performance with 95% confidence interval of partial least square discriminant analysis for the prediction of coagulate aptitude <sup>1</sup> of goat bulk milk using mid-infrared spectroscopy (MIRS).

Performance	Calibration Set (75%)	Validation Set (25%)
Prevalence	0.51	0.50
Sensitivity	0.79 (0.73–0.85)	0.68 (0.55–0.79)
Specificity	0.82 (0.76–0.87)	0.81 (0.69–0.90)
Positively Predictive Value	0.82 (0.76–0.87)	0.78 (0.65–0.88)
Negatively Predictive Value	0.80 (0.73–0.85)	0.71 (0.59–0.82)
Balanced Accuracy	0.81 (0.75–0.86)	0.74 (0.62–0.85)

<sup>1</sup> Two classes: samples with ‘high aptitude to coagulate’ ( $IAC > 100$ ) and ‘low aptitude to coagulate’ ( $IAC \leq 100$ ).





**Figure 3.** Receiver operating characteristic curves for goat bulk milk coagulation aptitude in calibration (a) and validation (b), with the corresponding area under the curve (AUC). The samples classification was ‘high aptitude to coagulate’ ( $IAC > 100$ ) and ‘low aptitude to coagulate’ ( $IAC \leq 100$ ).

Considering that the predictive performance of the MCPs from the spectral data was moderate, binary phenotypes derived from the PLS-DA could serve as a valid indicator of traits for various applications, such as milk payment systems. Currently, in both the goat and cow milk industries, payment systems are primarily based on quantity and quality, with bonuses or penalties applied to farmers according to the fat and protein content as well as the SCC. Including a phenotype related to high or low coagulation aptitude could be a valuable tool for improving the suitability of goat milk for cheesemaking. This approach could incentivize the production of milk with better technological properties, thereby enhancing both the product quality and the economic returns for farmers.

### 3.4. Future Perspectives

Future research should aim to expand the calibration dataset by incorporating greater variability for factors such as lactation stage and farming system, which were not explicitly considered in the present study. The inclusion of multiple breeds already provided a degree of biological diversity, and further broadening of the management and production-related variables may help refine the model, enhancing its robustness and applicability across different contexts. As a future application, the development of such models with portable or handheld near-infrared and mid-infrared spectrometers—already explored for assessing milk composition, quality traits, and adulterants [37–39]—could open new opportunities for on-site screening of the MCPs, although their use to predict these parameters has not yet been established. Moreover, the prediction of the IAC, an index that incorporates both the RCT and  $a_{30}$ , could be a valid tool for milk payment systems, helping to distinguish and reward farms producing milk with a higher aptitude to coagulate compared to those with a lower aptitude. However, to be effectively implemented in such systems, it is necessary to expand the dataset to capture greater variability and improve the robustness and the generalizability of the predictive models.

## 4. Conclusions

This study highlights the potential of MIRS as a tool for predicting the RCT in goat milk and assessing the discriminant ability of spectra to identify samples with high coagulation aptitude. Although the models developed still require improvements to achieve full reliability across all the traits included, the promising results for RCT prediction suggest

that MIRS could serve as a valuable starting point for developing rapid and cost-effective screening methods. The other models ( $k_{20}$  and  $a_{30}$ ) require further refinement before practical application can be considered, including the incorporation of larger and more diverse datasets and the integration of a wider range of factors related to farm management and production conditions. Moreover, MIRS can be a reliable tool for classifying samples based on their coagulation aptitude (high or low). Beyond its analytical potential, this approach is relevant to the dairy industry, where fast and affordable methods for assessing milk quality are increasingly needed to support product standardization and valorization. Validated models with good accuracy could be implemented in milk collection centers or on farms to support rapid decision-making, and MIRS could gain the potential to become a resource for the dairy goat industry to monitor and enhance milk quality for cheesemaking and improve product consistency.

Future research should focus on refining the prediction models by expanding the calibration datasets to include a wider range of factors to fully exploit the potential of this technology in supporting quality-oriented strategies.

**Author Contributions:** Conceptualization, M.D.M.; methodology, A.G. and S.M.; software, A.G. and S.M.; validation, M.D.M.; formal analysis, A.G. and S.M.; investigation, C.B. and L.L.; resources, M.D.M. and C.B.; data curation, A.G., S.M., C.B. and L.L.; writing—original draft preparation, A.G. and S.M.; writing—review and editing, M.D.M.; visualization, A.G.; supervision, M.D.M.; project administration, M.D.M.; funding acquisition, M.D.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The original contributions presented in this study are included in the article, further inquiries can be directed to the corresponding author.

**Acknowledgments:** The authors would like to thank all the goat farmers who contributed to the realization of this study.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Cabral, C.F.S.; Veiga, L.B.E.; Araújo, M.G.; de Souza, S.L.Q. Environmental Life Cycle Assessment of goat cheese production in Brazil: A path towards sustainability. *LWT* **2020**, *129*, 109550. [\[CrossRef\]](#)
2. Moises, R.M.M.; Salgueiro, C.C.M.; Leitão, M.S.P.; Nunes, J.F. Exploring goat's milk cheese: A systematic review of production techniques and innovations (2013–2023). *Braz. J. Food Technol.* **2024**, *27*, e2023152. [\[CrossRef\]](#)
3. Andueza, D.; Rouel, J.; Chilliard, Y.; Leroux, C.; Ferlay, A. Prediction of the goat milk fatty acids by near infrared reflectance spectroscopy. *Eur. J. Lipid Sci. Technol.* **2013**, *115*, 612–620. [\[CrossRef\]](#)
4. Pazzola, M. Coagulation Traits of Sheep and Goat Milk. *Animals* **2019**, *9*, 540. [\[CrossRef\]](#)
5. Hovjecki, M.; Miloradovic, Z.; Barukčić, I.; Blažić, M.; Miocinovic, J. Rheological Properties of Goat Milk Coagulation as Affected by Rennet Concentration, pH and Temperature. *Fermentation* **2022**, *8*, 291. [\[CrossRef\]](#)
6. Bittante, G.; Penasa, M.; Cecchinato, A. Invited review: Genetics and modeling of milk coagulation properties. *J. Dairy Sci.* **2012**, *95*, 6843–6870. [\[CrossRef\]](#)
7. Bittante, G.; Amalfitano, N.; Bergamaschi, M.; Patel, N.; Haddi, M.L.; Benabid, H.; Pazzola, M.; Vacca, G.M.; Tagliapietra, F.; Schiavon, S. Composition and aptitude for cheese-making of milk from cows, buffaloes, goats, sheep, dromedary camels, and donkeys. *J. Dairy Sci.* **2022**, *105*, 2132–2152. [\[CrossRef\]](#)
8. Pazzola, M.; Amalfitano, N.; Bittante, G.; Dettori, M.L.; Vacca, G.M. Composition, coagulation properties, and predicted cheesemaking traits of bulk goat milk from different farming systems, breeds, and stages of production. *J. Dairy Sci.* **2022**, *105*, 6724–6738. [\[CrossRef\]](#)
9. Hoxha, M.; Mara, K.; Mara, V.; Shumka, S. Milk clotting properties of liqenasi goat, milk syneresis kinetics and recovery of protein and casein on curd. *J. Hyg. Eng. Des.* **2022**, *38*, 204–212.

10. Ferragina, A.; Cipolat-Gotet, C.; Cecchinato, A.; Pazzola, M.; Dettori, M.L.; Vacca, G.M.; Bittante, G. Prediction and repeatability of milk coagulation properties and curd-firming modeling parameters of ovine milk using Fourier-transform infrared spectroscopy and Bayesian models. *J. Dairy Sci.* **2017**, *100*, 3526–3538. [\[CrossRef\]](#)
11. Ozaki, Y. Infrared spectroscopy—Mid-infrared, near-infrared, and far-infrared/terahertz spectroscopy. *Anal. Sci.* **2021**, *37*, 1193–1212. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Rueda-Robles, A.; Lozano-Sánchez, J. Analysis of milk and dairy products by non-destructive techniques. In *Non-Invasive and Non-Destructive Methods for Food Integrity*; Jiménez-Carvelo, A.M., Arroyo-Cerezo, A., Cuadros-Rodríguez, L., Eds.; Springer Nature: Berlin/Heidelberg, Germany, 2024; p. 255.
13. Bonfatti, V.; Degano, L.; Menegoz, A.; Carnier, P. Mid-infrared spectroscopy prediction of fine milk composition and technological properties in Italian Simmental. *J. Dairy Sci.* **2016**, *99*, 8216–8221. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Cellesi, M.; Coreddu, F.; Manca, M.G.; Serdino, J.; Gaspa, G.; Dimauro, C.; Maciotta, N.P.P. Coagulation properties and individual cheese yield in sheep using partial least squares regression. *Animals* **2019**, *9*, 663. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Dadousis, C.; Cipolat-Gotet, C.; Stocco, G.; Ferragina, A.; Dettori, M.L.; Pazzola, M.; do Nascimento Rangel, A.H.; Vacca, G.M. Goat farm variability affects milk Fourier-transform infrared spectra used for predicting coagulation properties. *J. Dairy Sci.* **2021**, *104*, 3927–3935. [\[CrossRef\]](#)
16. Stocco, G.; Dadousis, C.; Vacca, G.M.; Pazzola, M.; Paschino, P.; Dettori, M.L.; Ferragina, A.; Cipolat-Gotet, C. Breed of goat affects the prediction accuracy of milk coagulation properties using Fourier-transform infrared spectroscopy. *J. Dairy Sci.* **2021**, *104*, 3956–3969. [\[CrossRef\]](#)
17. Ali, A.K.A.; Shook, G.E. An optimum transformation for somatic cell concentration in milk. *J. Dairy Sci.* **1980**, *63*, 487–490. [\[CrossRef\]](#)
18. Penasa, M.; De Marchi, M.; Ton, S.; Ancilotto, L.; Cassandro, M. Reproducibility and repeatability of milk coagulation properties predicted by mid-infrared spectroscopy. *Int. Dairy J.* **2015**, *47*, 1–5. [\[CrossRef\]](#)
19. Grewal, M.K.; Chandrapala, J.; Donkor, O.; Apostolopoulos, V.; Stojanovska, L.; Vasiljevic, T. Fourier transform infrared spectroscopy analysis of physicochemical changes in UHT milk during accelerated storage. *Int. Dairy J.* **2017**, *66*, 99–107. [\[CrossRef\]](#)
20. Grewal, M.K.; Huppertz, T.; Vasiljevic, T. FTIR fingerprinting of structural changes of milk proteins induced by heat treatment, deamidation and dephosphorylation. *Food Hydrocoll.* **2018**, *80*, 160–167. [\[CrossRef\]](#)
21. Grelet, C.; Dardenne, P.; Soyeurt, H.; Fernandez, J.A.; Vanlierde, A.; Stevens, F.; Gengler, N.; Dehareng, F. Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions. *Methods* **2021**, *186*, 97–111. [\[CrossRef\]](#)
22. Shenk, J.S.; Westerhaus, M.O.; Abrams, S. Protocol for NIR calibrations: Monitoring analysis results and recalibration. In *Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality*, 2nd ed.; Martens, G., Shenk, J., Barton, F., Eds.; USDA-ARS agriculture handbook; US Government Printing Office: Washington, DC, USA, 1989; pp. 104–110.
23. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [\[CrossRef\]](#)
24. Sandrucci, A.; Bava, L.; Tamburini, A.; Gislón, G.; Zucali, M. Management practices and milk quality in dairy goat farms in Northern Italy. *Ital. J. Anim. Sci.* **2019**, *18*, 1–12. [\[CrossRef\]](#)
25. Williams, P.C. A short course in the practical implementation of near-infrared spectroscopy for the user. In *Near-Infrared Technology: Getting the Best Out of Light*, 1st ed.; PDK Projects Inc.: Nanaimo, BC, Canada, 2003; p. 109.
26. Manuelian, C.L.; Visentin, G.; Boselli, C.; Giangolini, G.; Cassandro, M.; De Marchi, M. Short communication: Prediction of milk coagulation and acidity traits in Mediterranean buffalo milk using Fourier-transform mid-infrared spectroscopy. *J. Dairy Sci.* **2017**, *100*, 7083–7087. [\[CrossRef\]](#)
27. Guerra, A.; Boselli, C.; Galli, T.; Ciofi, L.; Fichi, G.; De Marchi, M.; Manuelian, C.L. Low Effectiveness of Mid-Infrared Spectroscopy Prediction Models of Mediterranean Italian Buffalo Bulk Milk Coagulation Traits. *Foods* **2024**, *13*, 1957. [\[CrossRef\]](#)
28. Boselli, C.; Guerra, A.; Costa, A.; De Marchi, M. Prediction of sheep bulk milk coagulation properties from mid-infrared spectral data. *Acta IMEKO* **2024**, *13*, 1–3. [\[CrossRef\]](#)
29. Tiezzi, F.; Pretto, D.; De Marchi, M.; Penasa, M.; Cassandro, M. Heritability and repeatability of milk coagulation properties predicted by mid-infrared spectroscopy during routine data recording, and their relationships with milk yield and quality traits. *Animal* **2013**, *7*, 1592–1599. [\[CrossRef\]](#)
30. Kotsiou, K.; Andreadis, M.; Manassis, G.; Lazaridou, A.; Biliaderis, C.G.; Basdagianni, Z.; Bossis, I.; Moschakis, T. Effects of Farming System on the Rheological Behavior of Rennet-Induced Coagulation in Milk from Skopelos Breed Goats. *Foods* **2025**, *14*, 1316. [\[CrossRef\]](#)
31. Williams, P. Statistical terms for evaluation of accuracy and precision. In *Near Infrared Technology—Getting the Best Out of Light*, 5th ed.; PDK Grain: Nanaimo, BC, Canada, 2007; Chapter 5; pp. 1–17.
32. Karoui, R.; Downey, G.; Blecker, C. Mid-infrared spectroscopy coupled with chemometrics: A tool for the analysis of intact food systems and the exploration of their molecular structure—quality relationships—A review. *Chem. Rev.* **2010**, *110*, 6144–6168. [\[CrossRef\]](#)

33. Baba, T.; Pegolo, S.; Mota, L.F.M.; Peñagaricano, F.; Bittante, G.; Cecchinato, A.; Morota, G. Integrating genomic and infrared spectral data improves the prediction of milk protein composition in dairy cattle. *Genet. Sel. Evol.* **2021**, *53*, 29. [\[CrossRef\]](#)
34. Caponigro, V.; Marini, F.; Scannell, A.G.; Gowen, A.A. Single-drop technique for lactose prediction in dry milk on metallic surfaces: Comparison of Raman, FT-NIR, and FT-MIR spectral imaging. *Food Control.* **2023**, *144*, 109351. [\[CrossRef\]](#)
35. Ahmad, A.; Ayub, H. Fourier transform infrared spectroscopy (FTIR) technique for food analysis and authentication. In *Nondestructive Quality Assessment Techniques for Fresh Fruits and Vegetables*; Springer Nature: Singapore, 2022; pp. 103–142.
36. Martín-Gómez, A.; Rodríguez-Hernández, P.; Cardador, M.J.; Vega-Márquez, B.; Rodríguez-Estévez, V.; Arce, L. Guidelines to build PLS-DA chemometric classification models using a GC-IMS method: Dry-cured ham as a case of study. *Talanta Open* **2023**, *7*, 100175. [\[CrossRef\]](#)
37. Gorla, G.; Mestres, M.; Boque, R.; Riu, J.; Spanu, D.; Giussani, B. ATR-MIR spectroscopy to predict commercial milk major components: A comparison between a handheld and a benchtop instrument. *Chemom. Intell. Lab. Syst.* **2020**, *200*, 103995. [\[CrossRef\]](#)
38. Pu, Y.; Pérez-Marín, D.; O'Shea, N.; Garrido-Varo, A. Recent Advances in Portable and Handheld NIR Spectrometers and Applications in Milk, Cheese and Dairy Powders. *Foods* **2021**, *10*, 2377. [\[CrossRef\]](#)
39. Guerra, A.; De Marchi, M.; Niero, G.; Chiarin, E.; Manuelian, C.L. Application of a short-wave pocket-sized near-infrared spectrophotometer to predict milk quality traits. *J. Dairy Sci.* **2024**, *107*, 3413–3419. [\[CrossRef\]](#) [\[PubMed\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.