

Review

Not peer-reviewed version

Spectroscopy-Based Methods and Supervised Machine Learning Applications for Milk Chemical Analysis in Dairy Ruminants

Aikaterini-Artemis Agiomavriti , [Maria P. Nikolopoulou](#) , [Thomas Bartzanas](#) , [Nikos Chorianopoulos](#) , [Konstantinos Demestichas](#) , [Athanasios I. Gelasakis](#) *

Posted Date: 4 November 2024

doi: 10.20944/preprints202411.0204.v1

Keywords: milk analysis; dairy ruminants; spectroscopy; supervised machine learning; precision livestock farming; classification and regression; near-infrared; mid-infrared; laser induced breakdown spectroscopy; neural networks



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

Spectroscopy-Based Methods and Supervised Machine Learning Applications for Milk Chemical Analysis in Dairy Ruminants

Aikaterini-Artemis Agiomavriti ^{1,2,†}, Maria P. Nikolopoulou ^{3,4,†}, Thomas Bartzanas ³, Nikos Chorianopoulos ⁵, Konstantinos Demestichas ⁶ and Athanasios I. Gelasakis ^{1,*}

- ¹ Laboratory of Anatomy and Physiology of Farm Animals, Department of Animal Science, School of Animal Biosciences, Agricultural University of Athens, Iera Odos 45 str., 11855, Athens, Greece; agiomavriti@aua.gr
- ² R&D Department, TCB Avgidis Automations S.A., Athens, 11744, Greece; aagiomavriti@tcb.gr
- ³ R&D Department, Telefarm S.A., Athens, 11744, Greece; mnikolopoulou@telefarm.gr
- ⁴ Laboratory of Computer Science, Department of Agricultural Economics, Agricultural University of Athens, Iera Odos 45 str., 11855 Athens, Greece; cdemest@aua.gr
- ⁵ Laboratory of Farm Structures, Department of Natural Resources Management & Agricultural Engineering, Agricultural University of Athens, Iera Odos 45 str., 11855 Athens, Greece; m.nikolopoulou@aua.gr, t.bartzanas@aua.gr
- ⁶ Laboratory of Microbiology and Biotechnology of Food, Department of Food Science and Human Nutrition, Agricultural University of Athens, Iera Odos 45 str., 11855, Athens, Greece; nchorian@aua.gr
- * Correspondence: gelasakis@aua.gr; Tel.: +30-210-529-4387
- † Equal contribution.

Abstract: Milk analysis is critical to determine its intrinsic quality, as well as its nutritional and economic value. Currently, the advancements and utilization of spectroscopy-based techniques combined with machine learning algorithms have made feasible the development of analytical tools and real-time monitoring and prediction systems in the dairy ruminant sector. The objectives of the current review were i) to describe the most widely applied spectroscopy-based and supervised machine learning methods utilized for the evaluation of milk components, origin, technological properties, adulterants, and drugs residues, ii) to present and compare the performance and adaptability of these methods and their most efficient combinations, providing insights into the strengths, weaknesses, opportunities, and challenges of the most promising ones regarding the capacity to be applied in milk quality monitoring systems both at the point-of-care and beyond, and iii) to discuss their applicability and future perspectives for the integration of these methods in milk data analysis and decision support systems across the milk value-chain.

Keywords: milk analysis; dairy ruminants; spectroscopy; supervised machine learning; precision livestock farming; classification and regression; near-infrared; mid-infrared; laser induced breakdown spectroscopy; neural networks

1. Introduction

It is projected that by 2050, the global population will exceed 9 billion people [1], a nearly 2-billion increase over the current population [2,3]. Most of this expansion will take place in developing countries, resulting in a sharp rise in the consumption of milk and products thereof. Indeed, in these countries the annual milk and dairy products consumption per capita is expected to increase by 1.47 times (from 45 to 66 kg), while a respective increase by 1.04 times (from 212 to 221 kg) is expected in developed countries [4]. Moreover, consumers are increasingly concerned about the environmental, public health, and animal welfare implications associated with the intensification of livestock production [3–5]. This situation has placed immense pressure on the dairy ruminant sector to find sustainable solutions for the optimization of the milk production systems and the minimization of

the environmental impact [e.g., rising water consumption, land and ecosystems degradation, increased greenhouse gas emissions, waste of natural resources, loss of biodiversity, etc.]. Towards this target, precision livestock farming (PLF) technologies have emerged as critical tools for the mitigation of the environmental impacts and essential components of sustainable production and evidence-based herd health management. Thus, sensors, animal-recording technologies, artificial intelligence (AI), and robotic systems, as well as life cycle assessment (LCA) methods are utilized by modern dairy farms, to significantly reduce their environmental footprint and improve their profitability [6].

It is undeniable that milk and products thereof are listed among the most valuable agricultural commodities, due to their high nutritional value in human diets [7]; therefore, milk production is an essential asset to global societies and economies. Moreover, milk quality is significant for the milk processing industry, directly affecting the technological properties, organoleptic traits, hygiene status, and overall acceptance of the derived dairy products, and subsequently the market value of milk. Therefore, milk quality and safety must be efficiently evaluated and managed to satisfy consumer demands, meet legal requirements, and ensure transparency and fair pricing for the farmers. Moreover, systematic assessment of milk chemical composition i) facilitates efficient monitoring of its intrinsic quality, ii) supports early detection and prevention of milk fraud, as well as intramammary infections and mastitis, and iii) decreases the time, effort, and expenses demanded for routine laboratory milk analyses [8,9].

Despite the efficiency of the traditional laboratory-based methods utilized for the assessment of milk quality, they require expensive equipment, specialized staff, well-organized logistics, time-consuming (>48h), labor intensive and destructive/invasive sampling, transferring, and analytical processes; thereby, crucial information sharing and farming decisions are delayed. The potential to monitor milk quality on-site has lately been made possible by the development of portable and handheld devices intended for use at the point-of-care (POC), following the recent advances in chemometric and optical sensor technologies. The primary advantages of these technologies is their capacity to collect enormous data volume, their accuracy and the real-time output [3]. This is consistent with the idea of PLF, which is defined by Tullo et al. as “the application of process engineering principles and techniques to livestock farming to automatically monitor, model, and manage animal production” [10]. These sensors offer accurate and reliable measurements of milk quality traits, its origin and adulteration, as well as the udder health status of animals, allowing the systematic, non-invasive monitoring, even when needed to be applied in situ and at the individual animal level [11,12]. Also, they collect data, which are then processed by algorithms and stored in databases for later use in decision support systems (DSS).

Among optical technologies, spectroscopy-based methods have emerged as promising tools for milk chemical analyses. Indeed, methods such as Raman spectroscopy, near-infrared spectroscopy (NIRS), mid-infrared spectroscopy (MIRS), and laser-induced breakdown spectroscopy (LIBS) provide rapid, non-invasive, and precise evaluation of milk components such as fat, protein, lactose, and others, as well as indications of the milk origin, adulteration, and occurrence of drug residues. Combining spectroscopic techniques with advanced machine learning (ML) algorithms like support vector machines (SVM), random forests (RF), logistic regression (LR), elastic net (EN), k-Nearest Neighbors (k-NN), neural networks (NN), and gradient boosting machines (GBM), can remarkably improve the diagnostic performance of these technologies. Hence, utilizing spectroscopic milk analysis and advanced ML methods, key-applications in the field of milk analysis have been released and expected to be further developed for the assessment and prediction of milk traits and properties.

The objectives of the current review were: i) to present the use of spectroscopy-based techniques for milk analyses, emphasizing on their specific applications and their potential integration into contemporary dairy management systems, ii) to compare the performance and adaptability of the available spectroscopic methods, providing insights into the strengths, weaknesses, opportunities, and challenges of the most efficient relevant technologies for upgrading milk quality monitoring systems in dairy farms, iii) to discuss the applicability and performance of ML techniques for milk data analysis systems and prediction models to facilitate quicker and better-informed DSS.

2. Spectroscopy Principles

2.1. Spectroscopy

The study of how light interacts with matter is known as spectroscopy [13]. The energy of light is proportional to its frequency and inversely proportional to its wavelength. The relationship is described by the following equation:

$$E = h * f \quad (1)$$

Where E is the energy of the light in J , h is the Planck's constant ($6.62 * 10^{-34} J * Hz^{-1}$) and f is the frequency of the light in Hz .

Alternatively, since the frequency is related to the wavelength (λ) by the speed of light (c) (where $c = \lambda * f$), the energy can also be expressed as:

$$E = \frac{h*c}{\lambda} \quad (2)$$

Where λ is the wavelength of the light in m and c is the speed of light ($\approx 3 * 10^8 m/s$).

In methodological terms, spectroscopy can be defined as the process of analyzing the spectrum of light that a substance absorbs, emits, or scatters in order to determine its physical structure, molecular composition, and other properties [14].

2.2. The Infrared Region of the Electromagnetic Spectrum

Located in the middle of the electromagnetic spectrum, the infrared (IR) region is divided into three main subregions: Far-Infrared (FIR), with wavelengths ranging from 1 mm to 10 μm , Mid-Infrared (MIR) from 10 μm to 2.5 μm and Near-Infrared (NIR) which refers to the portion of the electromagnetic spectrum closest to the visible region, with its wavelengths ranging from approximately 750 nm to 2500 nm [15]. Figure 1 illustrates the electromagnetic spectrum including the Visible (Vis) region from 380 nm to 750 nm and Ultraviolet (UV) region from 10 nm to 400 nm. William Herschel is credited with discovering near-infrared radiation in 1800. Herschel observed that light temperature rose from the blue (450 – 475 nm) to the red (620 – 750 nm) end of the spectrum while using a thermometer in his experiments. The temperature increased even after the thermometer was positioned beyond the visible red region, suggesting existence of energy beyond the visible spectrum [15,16]. The development of NIR spectroscopy, a potent analytical tool for analyzing the chemical and physical properties of materials, was made possible by this discovery.

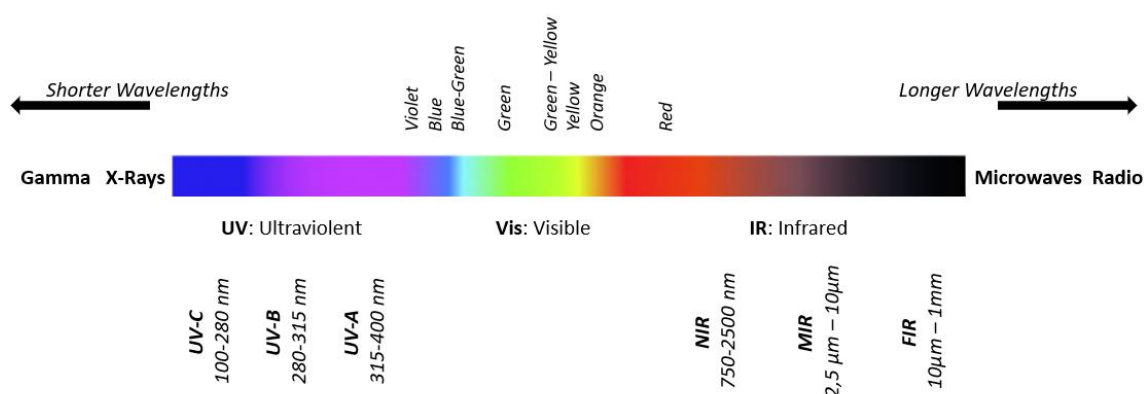


Figure 1. The electromagnetic spectrum and wavelength ranges of electromagnetic radiation.

2.3. Transmittance, Reflectance, Absorption and Emission of Light

Transmittance (T) indicates the amount of incident light that can pass through a material. Reflectance (R), is defined as the amount of incident light reflected by the material's surface, while absorbance is defined as the amount of incident light not reflected or transmitted but absorbed by the material (Figure 2). The conservation of energy requires that:

$$T + R + A = 1 \quad (3)$$

Where (T) is the transmittance, (R) is the reflectance and (A) the absorbance.

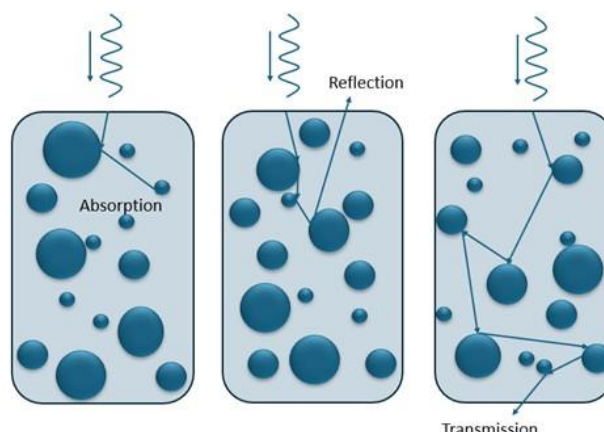


Figure 2. Example of light's interaction with matter [14] (modified).

Natural frequency of an atom is the frequency at which its electrons vibrate spontaneously. When atoms of a material vibrate at the same frequency with a light wave, their electrons absorb the wave's energy and start to vibrate as well. Objects vary in color because the electrons of different materials' atoms vibrate at different frequencies and therefore absorb different light frequencies. Electrons in atoms are confined to distinct energy levels or electron shells.

The lowest possible energy state is known as the ground state. According to the quantization of energy levels, electrons can move from a lower energy state to a higher one only by absorbing a discrete amount (quanta) of energy described by the laws of quantum mechanics. The difference between the two energy levels must be represented by the absorbed energy. When an electron absorbs energy, it is excited to a higher energy state and moves away from the nucleus of the atom. Electrons, however, do not remain excited for very long. After briefly being in this higher energy state, they return to their original ground state, releasing the absorbed energy in the form of photons. As stated in Kirchoff's radiation law, the energy of the emitted photons corresponds precisely to the amount of energy initially absorbed by the electrons. This process of light emission is called fluorescence and is a subcategory of luminescence.

Fluorescence is a rapid form of luminescence that starts very shortly after light absorption and ends almost immediately when the light source is removed. Luminescence is the general term for light emission without heat.

2.4. Light Scattering

Photons or light particles interact with matter in a process known as light scattering. When a light source illuminates a medium, its particles scatter light in different directions, causing it to deviate from its original optical path. Fundamentally, light scattering is the result of the interaction between matter and photons, or light particles. Light can disperse when photons encounter particles or irregularities in a medium. These particles or irregularities can absorb photons and reemit them in different directions. Decrease of the intensity of the light that passes through a medium can be caused by absorption or scattering; this is a basic phenomenon in optics since it gives rise to a variety of optical effects. Figure 3 demonstrates the scatter light effects in milk, caused by fat and protein particles.

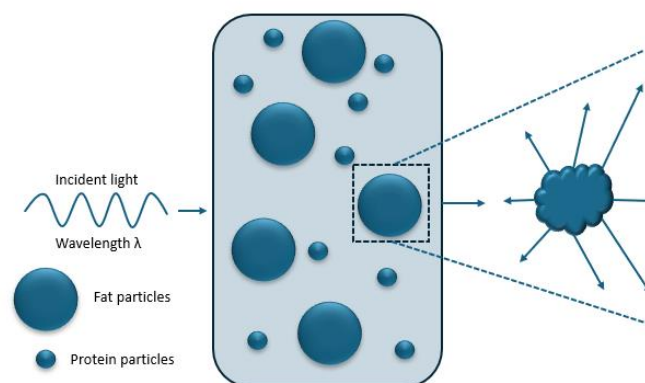


Figure 3. Scatter light effects generated by fat and protein particles in milk. The incident wavelength is smaller than the diameter of the particles, resulting in Mie scattering, which is demonstrated in the zoomed view [17].

2.5. Other Optical Properties

Dispersion or refraction is defined as the process during which different colors bend at slightly different angles, (e.g., the formation of a rainbow), and is linked to the refractive index which is affected by the wavelength (color) of the light and indicates how much the light bends and slows down (Figure 4).

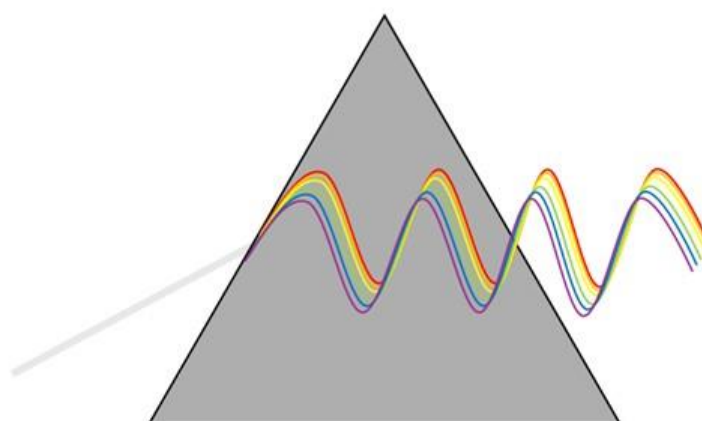


Figure 4. Different colors refract at different angles in a dispersive prism due to material dispersion, a wavelength-dependent refractive index divides white light into a spectrum [18].

Refractive-index-based sensors measure variations in a medium (such as gases, liquids, or solids) by measuring the refraction of the light which passes through the medium. Variations in the refractive index indicate changes in the material's composition, temperature or density. In some spectroscopic techniques, the refractive index itself can be used as a tool to identify or characterize samples. Different materials have different refractive indices at different wavelengths, so measuring a material's refractive index spectrum can provide information about its composition and structure.

When light is absorbed by a material, causing a localized increase in temperature, it is referred to as the optothermal effect. The substance increases in temperature when it absorbs light energy and converts it to thermal energy or heat. This effect is widely used in optothermal spectroscopy and other fields where the heat generated by light absorption is essential for material manipulation or detection. The distinct heat conductivity properties of graphene have been discovered through the application of the optothermal effect, and more precisely, the optothermal Raman technique, which measures the local temperature of the sample using Raman spectroscopy and uses the excitation laser as a heat source.

2.6. Optical Chemosensors

The devices that are specially designed to detect, identify, and quantify chemical compounds are defined as chemical sensors. Their operation is based on the exploitation of chemical reactions or physical changes to acquire measurable signals (e.g., optical signals) and achieve the quantification of the desired compounds [19]. Optical chemosensors are considered a subclass of chemical sensors, and they can be divided based on the optical properties used in the sensors to detect compounds [20] as illustrated in Figure 5.

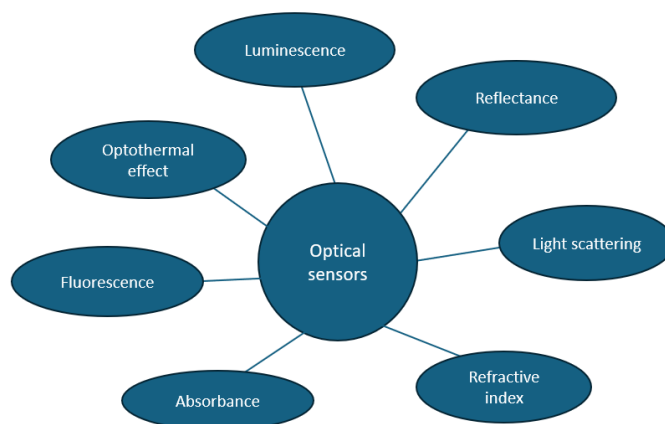


Figure 5. Optical sensors classification from International Union of Pure and Applied Chemistry (IUPAC) [21].

The same optical properties that govern the operation of optical chemosensors are also employed in a range of spectroscopic methods. Despite most of these methods being applied in the laboratory, they all operate on the same principles. Therefore, these types of optical sensors can also be utilized in portable spectroscopy sensors or analyzers. As technology advances, portable spectroscopic sensors are expected to be increasingly exploited for milk analyses in the dairy ruminant industry.

3. Milk Composition and Quantification Techniques

Milk is a complex biological fluid. In ruminants, its composition varies depending on the species of origin and several other physiological, genetic, and environmental factors, consisting of approximately 80 - 87% water, 3.6 - 7.9% fat, 3.2 - 6.2% proteins, and 4.1 - 4.9% lactose [17]. These chemical compounds are key-determinants of the milk quality, influencing both its nutritional value and technological properties [22]. Among milk components, milk fat is important as it affects the cheese-making capacity and the nutritional value of milk. Moreover, the fat content and the fatty acids (FAs) profile thereof contribute to important organoleptic traits such as taste, density, appearance, and flavor [22], while polyunsaturated FAs, such as oleic acid, trans-11, C18:2, cis-9, and α -linolenic acid, offer various health benefits including reinforcement of the immune system, hormone production, and cognitive health [23]. Proteins are also vital components of milk in terms of their nutritional and economic value. Between the main proteins in milk the casein fraction is significantly related to the cheese-making capacity of milk and the functional and technological properties of dairy products, while it is a significant source of amino acids [24]. Except for being a source of essential amino acids, milk proteins exhibit a range of biological functions supporting growth and maintenance, metabolic reactions, hormonal and immune system functions, energy storage, antioxidant and antimicrobial activities. Lactose is another main component of milk and the main carbohydrate in it, directly associated with the milk yield capacity of ruminants; it is a disaccharide made up of glucose and galactose and is vital as an energy source [23], while it is widely referred as “milk sugar” and it is the sole common sugar of animal origin. Urea is a metabolic product of proteins and amino acids. It is the primary source of non-protein nitrogen in milk. Its concentration

above the physiological threshold may be indicative of renal diseases or imbalanced nutrition, while it plays a significant role in overall milk quality, as it can also be an indicator of milk adulteration [22]. Somatic cells in milk refer to all the cells that can be found in milk and have nuclei (e.g., white blood cells, epithelial cells etc.). Somatic cell count, or SCC, is a crucial indicator of the milk quality and the animals' welfare. Determination of milk price, regulatory compliance monitoring, udder health and genetic assessment are some of the uses of SCC estimation in milk [25].

The concentration of certain compounds can provide insights into udder health status and other nutritional and physiological parameters of ruminants. For example, elevated SCC is one of the most reliable indicators of intramammary infections and mastitis [26], while the protein and urea contents of milk can be used to assess the balance between protein intake and energy supply in ruminants' diets [27]. The concentration of these milk compounds can be quantified through various analytical techniques, with spectroscopic methods gaining increasing popularity. Figure 6 illustrates the spectroscopy methods used for various milk applications.

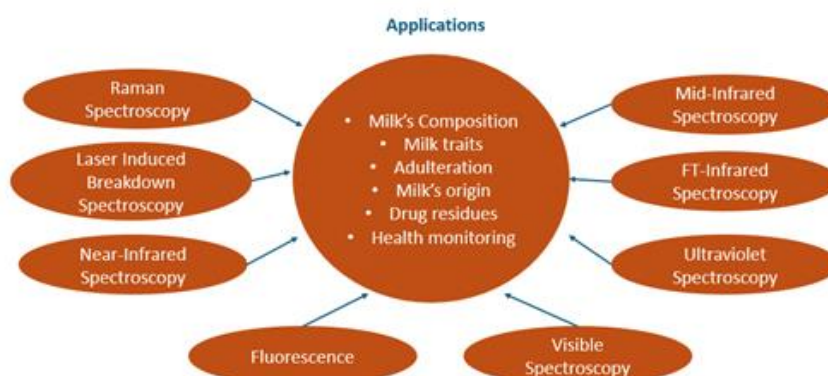


Figure 6. Milk application and spectroscopy methods [22].

4. Spectroscopy Applications

Conventional methods used for the estimation of milk compounds, as well as for the detection of adulterants and drug residues and the microbiological assessment of milk usually involve labor-intensive and time-consuming procedures performed by specialized staff. Considering the growing demand for real-time, non-invasive analyses, spectroscopy-based techniques have emerged as promising tools for ensuring the production of high-quality and safe milk. Various technologies, including Raman spectroscopy, LIBS, NIRS, and MIRS, have shown to be effective in providing rapid and reliable assessment of the milk composition and its microbiological status in dairy farms. Although in this review the chemical analysis of milk is examined, it should be noted that spectroscopy methods are also widely used for the identification of microbial and bacterial contamination in milk [28–30]. The application of the forementioned spectroscopy methods can be used to collect data and in combination with a variety of multivariate analysis techniques to extract analytical information and predict milk quality (Figure 7); this is achieved by correlating multiple analytical variables (as derived from the spectrum analysis) with the properties of the studied analytes content, such as milk components, adulterants, and drug residues [31].

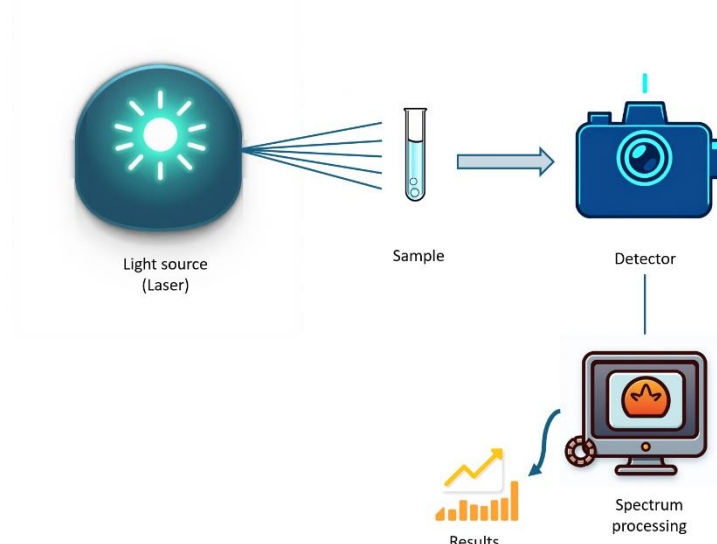


Figure 7. Illustration of the spectroscopy procedure.

In this section, the definitions and principles of various spectroscopy methods are summarized along with their application in milk chemical analyses. Also, studies that primarily utilized Raman, LIBS, NIR, and MIR spectroscopy are presented and discussed, regarding their spectral ranges, calibration models, and predictive capacity when applied for the measurement of milk components, the detection of adulterants, and drug residues, as well as the discrimination of milk origin (different ruminant species, organic from non-organic milk, etc.)

4.1. Reflectance, Absorption, and Emission Spectroscopy

Reflectance spectroscopy is defined as the study of the light reflected from a solid, liquid, or gas material, as a function of its wavelength. It is a process that quantifies the light or electromagnetic radiation that is reflected off the surface of the material of interest. By analyzing the spectrum of the reflected light, information about the material's composition, structure, and surface properties can be obtained.

The process of measuring light absorption in materials is carried out through absorption spectroscopy. A continuous band of color with black lines connecting them displays the material's absorption spectrum (Figure 8). Colored portions depict the entire amount of light directed onto the substance, while the areas of the spectrum where the electrons absorbed the light photons are indicated by the black lines, which depict the absence of the directed light. Absorption spectroscopy is further divided into molecular and atomic absorption spectroscopy. Atomic absorption spectroscopy is the process of generating a spectrum when free atoms absorb various light wavelengths; it is a method commonly used to analyze gases. Molecular absorption spectroscopy is the process of generating a spectrum when entire molecules absorb various light wavelengths, usually at the Vis or the UV region of the spectrum.

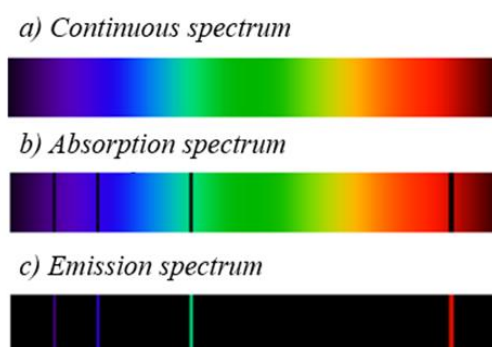


Figure 8. (a) Continuous spectrum: contains all wavelengths emitted by a light source, (b) Absorption spectrum: black lines where the electrons have absorbed the light photons, (c) Emission spectrum: color lines where photons have been released from the electrons when they fall to a lower energy level [32].

Emission spectroscopy counts the photons released when excited electrons return to their ground state. An emission spectrum is shown as a black background with distinct colored lines that represent the wavelengths of photons emitted as electrons release energy. Emission spectra can be categorized as either line emission spectra, which display discrete colored lines separated by black spaces, or continuous emission spectra, which show a continuous range of colors across wavelengths (Figure 8). Since different substances release energy in characteristic patterns, emission spectroscopy is a powerful tool for analyzing complex materials to identify their components.

4.2. Raman Spectroscopy

Raman spectroscopy is an analytical method that uses scattered light to quantify a sample's vibrational energy modes. It is named after C. V. Raman, an Indian physicist who, in 1928, together with K. S. Krishnan, made the first observation of Raman scattering [33]. Raman spectroscopy is a vibrational spectroscopic technique that uses a substance's distinctive "fingerprint", through which it can identify and provide structural and chemical information of any kind of material [34]. This information is extracted by Raman spectroscopy by detecting Raman scattering in the sample.

Raman spectroscopy is used in milk analysis for a variety of purposes, including the assessment of the content of the major milk compounds, as well as, for the detection of drug residues [23]. It does not require any special sample pretreatment, enabling real-time, *in situ* monitoring of milk components. Vaskova et al. [34], used Raman spectroscopy to measure lactose content in dried milk droplets, demonstrating the broad applicability of this technique, while Mazurek et al. [35] used Raman spectroscopy to analyze 64 bovine milk samples for the quantification of the fat, protein, lactose, and dry matter contents. The same technique was used by El-Abassy et al. [36] to determine milk fat content in different types of milk samples; in their study, measurements were made using the 514.5 nm emission line of an argon ion laser, specifically the Coherent Innova 308 Series, with 30 s recording time; results regarding the liquid milk fat content prediction capacity of the method were promising, showing low root mean square errors (0.16 and 0.06) and high correlation coefficients (0.97 and 0.97) for milk samples with fat from 0.3-1.55% and 0.3-3.8%, respectively. Concerning dried milk samples, the results were also very promising with $R^2 = 0.97$ and $RMSE = 0.18$. In a study by Rodrigues Júnior et al. [37], a combination of chemometric analysis and Raman spectroscopy were utilized to detect adulterants and to assure the quality of milk powder with regard to fraud involving the addition of maltodextrin and the classification of milk powder samples according to their lactose content. The detection of adulteration via Raman spectroscopy was also investigated by Khan et al. [38]; in that study, recording of the liquid samples' spectra, with 27 different values of urea concentrations, was performed using a 785 nm diode laser (CL-2000, CrystalLaser). It was found that urea concentration could be accurately predicted (> 97% accuracy) for concentrations above 100 mg/dl. However, the accuracy of the method decreases with the urea concentration (90-95% for 50-100 mg/dl and < 60% for 50 mg/dl).

Raman spectroscopy's non-destructive nature and its capacity to quickly and accurately analysing ruminants' milk demonstrate its potential application in routine milk quality assessment and dairy management systems [34]. Milk components can be efficiently identified and quantified due to the method's sensitivity, which makes Raman spectroscopy a useful tool for the systematic evaluation of milk quality status in a variety of applications, even in raw milk samples collected on-site, in dairy ruminant farms [34]. Nevertheless, despite its advantages, most applications of Raman spectroscopy in dairy systems are still under development, and there are still theoretical and technological issues that need to be resolved, such as the enhancement of its accuracy for different milk types and the minimization of sample preparation. Furthermore, the high cost of Raman systems limits their accessibility, particularly for small and medium-sized dairy farms, for which the initial investment cost may be prohibitive [23].

Table 1. Raman spectroscopy applications and performance.

Wavelength (cm^{-1})	Type of milk sample	No of samples	Origin of milk	Application	R^2	RMSE	Diagnostic performance	Ref.
300 – 1700	powder	ND	retail	lactose	0.91	-	-	[34]
250 – 3500	powder	136	retail	fat protein	-	0.21 – 0.31 % w/w _p 0.14 – 0.35 % w/w _p	-	[39]
800 – 3050	liquid* liquid** powder*	13	retail	fat	0.97 _v 0.97 _v 0.97 _v	0.16% _v 0.06% _v 0.18% _v	-	[36]
8, 16, 32	liquid	75	retail	fat protein carbohydrates dry matter	-	5.3 – 5.8% _{sp} 5.6 – 6.1% _{sp} 3.5 – 4.8% _{sp} 3.4 – 4.8% _{sp}	-	[35]
400 – 3500	powder	45	retail	lactose high/low classification maltodextrin adulteration	-	-	Se: 98.6% Sp: 100.0% Se: 88.6% Sp: 100.0%	[37]
750 – 1800	liquid	10 batches	retail	urea adulteration	> 0.95	-	Acc+ 100mg/dl: > 97% 50-100mg/dl: 90-95% <50mg/dl: ≈ 60%	[38]
600 – 1800	liquid	602	cow human buffalo goat	milk origin	-	-	Se: 93.0% Sp: 97.0% Acc: 93.7%	[40]

R^2 : Coefficient of determination, RMSE: root mean square error, ND: Not Defined, _p: RMSEP (root mean square error of prediction), _{sp}: RSEP (relative standard errors of prediction), *: 0.3-3.8%, **: 0.3-1.55%, _v: RMSEV (root mean square error of validation), + depending on urea concentration, Se: Sensitivity, Sp: Specificity, Acc: Accuracy.

4.3. Laser Induced Breakdown Spectroscopy (LIBS)

The optical emission method known as LIBS is used to ascertain the elemental composition of materials [41]. This process involves directing a focused, pulsed laser onto a sample, generating plasma which results from the ionization of the material's atoms. As the plasma cools, the recombination of atoms with free electrons produces light across the UV, Vis, and IR regions [42]. A small amount of the target material (solid, liquid, or gas) is vaporized by the high-energy laser pulses, and the light emitted from the excited atomic and ionic species in the plasma is gathered for spectroscopic analysis to determine the elemental composition of the sample [43].

Laser-Induced Breakdown Spectroscopy is a relatively new optical method that holds great promise for milk analysis. Indeed, it becomes increasingly popular due to its potential to provide quick, multi-elemental analyses, with high sensitivity and accuracy, in a variety of complex matrices, including liquid and solid milk samples, as well as due to its quick and easily adaptable methodology [44–46]. This method requires from minimal to zero sample preparation, offers real-time analysis, and operates as a non-contact technique, making it suitable for POC applications [45]. Laser-Induced Breakdown Spectroscopy has been utilized for the detection of minerals, trace elements, and adulterants in milk, to support quality control and nutritional evaluation processes in the dairy value-chain. To fully unravel its potential for widespread industrial application in dairy quality assurance systems, further optimization is needed, especially with regard to the calibration models and the improvement of the precision within complex milk matrices [45].

Liquid bovine, ovine, and caprine milk samples were analyzed using LIBS in the studies by Nanou et al. [46,47], resulting in unique spectral lines of specific milk compounds and accurate elemental profiles of milk. In particular, the spectral characteristics of major elements such as magnesium (Mg), calcium (Ca), sodium (Na), and potassium (K), as well as minor minerals like phosphorus (P), zinc (Zn), copper (Cu), and silicon (Si), were accurately detected and identified [46]. Notably, Nanou et al. [47] used milk ash for the analysis of minor minerals content in order to improve the trace element detection accuracy, while key inorganic spectral lines and LIBS spectra were utilized in the same study to differentiate milk samples based on the animal species of origin. A variety of ML algorithms were exploited to classify the samples with remarkable precision; classification accuracy of up to 95.5% using the full LIBS spectra were achieved. Even when focusing on five specific spectral lines —magnesium Mg(II) at 279.8 and 280.3 nm, calcium Ca(I) at 422.6 nm, ionic calcium Ca(II) at 315.9, 317.9, 393.3, and 396.8 nm, sodium Na(I) at 589.0 nm, and potassium K(I) at 766.5 and 769.8 nm—the classification accuracy remained at approximately 93%. These results indicate that rapid and accurate milk origin assessment can be achieved by the combined implementation of LIBS and the appropriate ML algorithms.

Table 2. Observed spectral lines of the major minerals in Laser-Induced Breakdown Spectroscopy spectra and their corresponding wavelengths [46].

Element	Wavelength (nm)
H	486.1(H_{β}), 656.3(H_{α})
N (I)	742.4, 744.2, 746.8, 818.8, 821.6, 824.2, 862.9, 865.6
N (II)	500.5, 568.6
O (I)	715.6, 777.2, 777.4, 777.5, 844.6, 926.4
C (I)	247.8, 795.2, 906.2, 940.6
Mg (II)	279.8, 280.3
Ca (I)	422.6, 428.3*, 428.9*, 430.2*, 431.9*, 442.5*, 443.6*, 445.5*, 559.4*, 612.2*, 616.2*, 643.9*, 646.3*, 649.4*
Ca (II)	315.9, 317.9, 393.3, 396.8
Na (I)	589.0
K (I)	766.5, 769.8

*Spectral lines observed only in lyophilized powder milk; I: Atomic; II: Ionic.

In the study by Moncayo et al. [48], LIBS was combined with NN for both qualitative and quantitative analysis of milk adulteration. The authors applied chemometric tools, NNs and Principal Component Analysis (PCA), alongside LIBS data, which were collected using a Q-switched Neodymium-doped (Nd): Yttrium Aluminium Garnet (YAG) laser (Quantel Brio model) operating at 1064 nm. The application of NN on the LIBS data enabled the development of predictive models with high accuracy in detecting adulterated milk samples and for the estimation of the melamine content. Neural networks incorporation significantly enhances LIBS utility as a non-invasive, real-time technique for milk quality assessment and fraud detection, offering a powerful tool for dairy industry applications. Adulteration in whey milk powder was also investigated by Bilge et al. [49]; an 80.5% discrimination rate between powdered milk, whey, and demineralized whey was achieved, while, the correlation coefficients (R^2) for adulteration with sweet and acid whey were 0.981 and 0.985 respectively. In the study by Abdel-Salam et al. [50], the quality of maternal milk and commercial infant formulas were evaluated, using samples of maternal milk and formula samples from 6 popular commercial products. Using the acquired spectra and by comparing the intensities of the spectral lines in the samples, the authors concluded that maternal milk had higher overall nutritional value compared to the formulas, while it was found that younger mothers produced higher quality milk.

In a more recent study by Abdel-Salam et al. [51], quality traits of 300 milk samples, derived from 99 dairy cows (with and without mastitis), were assessed using LIBS. From these samples, forty samples were selected, based on the SCC measurements, to be further used for in-depth LIBS analysis. It was found that subclinical and clinical mastitis was associated with a lower milk quality, particularly regarding the protein and lactose content. Furthermore, a robust positive correlation between the LIBS spectral scores and SCC was observed, underpinning the potential exploitation of LIBS as a quick and efficient way to monitor milk quality on-site and as a diagnostic tool for the early detection of mastitis-induced changes in milk.

Table 3. Laser Induced Breakdown Spectroscopy (LIBS) applications and performance.

Wavelength (nm)	Type of milk sample	No of samples	Origin of milk	Application	R^2	RMSE/SEP	Accuracy (%)	Ref.
534.9	powder	23	retail	Ca	0.92	2614 mg kg ⁻¹ _{SEP}	-	[44]
766.5				K	0.80	1549 mg kg ⁻¹ _{SEP}		
285.2				Mg	0.91	91 mg kg ⁻¹ _{SEP}		
Laser excitation: 1064 & 532	liquid, ashed L/ph powder	ND	cow ^R , goat ^R , sheep ^R	major minerals ⁺ minor minerals ⁺⁺	-	-	-	[46]
181 – 904	powder	5	infant formula	Ca	0.85 _{pr}	0.68 mg/g _p	-	[52]
200 – 700	dried	60	maternal	composition quality (Mg, Ca, Fe, Na)	-	-	-	[50]
		ND	infant formula					
200 – 900	liquid	300	cow	fat, protein, lactose, SNF, density, SCC	-	-	-	[51]
200 – 1000	liquid	1296	cow, goat, sheep	milk origin	-	-	92.8	[47]
	L/ph powder	683					95.5	
Mg, Ca, Na, K spectral lines	liquid	1296	cow, goat, sheep	milk origin	-	-	87.6	[47]
	L/ph powder	683					92.9	
≈ 185 – 1048	powder	50	vetch root	milk origin	-	-	73.1	[53]
190 – 450	blended powder	12	cow ^R , goat ^R , sheep ^R	melamine ^A , p/b cls.	0.99 (melamine)	-	98 (cls. rate)	[48]
540 – 900	powder	36	cow	sweet whey ^A	0.981	-	-	[49]
				acid whey ^A	0.985			
186 – 900	gel	13	cow	caprine adult. with bovine ovine adult. with bovine	0.993	4.53 µg mL ⁻¹ _p	-	[54]
		13	goat					
		14	sheep					
196 – 874	powder	25	infant formula	exogenous protein	-	-	93.9 (SVM)	[55]
							97.8 (CNN)	

R^2 : Coefficient of determination, RMSE: Root Mean Square Error, SEP: Standard error of prediction, ND: Not Defined, L/ph: Lyophilized, SNF: Solids-Not-Fat, SCC: Somatic Cell Counts, SVM: Support Vector Machines, CNN: Convolutional Neural Network, p/b: pure/blended, cls : classification, ⁺ Major minerals: (Ca, Na, Mg, K), ⁺⁺ Minor minerals: (P, Zn, Cu, Si), ^R: retail, _p: RMSEP (root mean square error of prediction), _{pr}: prediction, ^A: adulteration.

4.4. Infrared (IR) Spectroscopy

Since the physicochemical properties of milk determine its spectrum, affect its intrinsic quality and nutritional value, and are related to the health and welfare of ruminants, IR spectroscopy provides a rapid and cost-effective method for measuring/predicting/diagnosing the above [56]. Over the past few decades, simple visible and NIR spectroscopy have been widely utilized to measure milk composition, as well as to monitor milk quality in dairy farms and milk-processing plants [57,58]; in particular, they have proven valuable technologies in laboratory settings for the evaluation of the fat, protein, and lactose content in raw milk [59]. Moreover, infrared thermography has been used as a diagnostic tool for udder health assessment and mastitis detection in dairy ruminants [60].

Milk exhibits absorption when it is illuminated; this phenomenon is governed by the Beer-Lambert's law (4) and is explained by Swinehart [61] as below:

$$A = \log_{10} \left(\frac{I_0}{I} \right) = d * \epsilon * c \quad (4)$$

The absorbance (A) depends on the optical path length (d) in (cm), molar absorptivity (ϵ) in (L/(mol * cm)), and analyte concentration (c) in (mol/L). The output of these elements can also be estimated by the logarithm response ($\log_{10}$) of the ratio between the intensity of incident light (I_0) and the intensity-transmitted light (I). Then, the concentration of different milk components (fat, protein, lactose, etc.) can be estimated by computing the absorbance. Absorption properties of milk in the IR region of the spectrum are determined by the presence of certain chemical groups, such as methylene group (-CH), hydroxyl group (-OH), and amino group (-NH), which are responsible for the vibration spectra in the NIR part of the spectrum; primary components of milk, such as fat (2340, 2310, 2270, 1780, 1730, 1720 nm), casein (2790, 2340, 2310, 2100, 1980, 1820, 1780, 1730, 1720, 1680, 1450 nm), and lactose (2340, 2100, 1820, 1450 nm) demonstrate distinct bands [57].

4.4.1. Near-Infrared Spectroscopy (NIRS)

Near-Infrared spectroscopy is the study of the light's emission, absorption, and reflection at the NIR region of the spectrum. This non-destructive technique uses the IR portion of the electromagnetic spectrum (which is approximately between 750 and 2500 nm), to analyze the physical, chemical, and other properties of various materials. Through a multi-analytical approach, NIRS allows the simultaneous and accurate prediction of multiple elements [8,62]. Thus, NIRS applications have increased significantly in the last few years compared to other traditional laboratory analytical methods due to its higher speed and accuracy, as well as its non-destructive nature and affordability [63].

4.4.1.1. Applications of Near-Infrared Spectroscopy in the Dairy Industry

Guidelines for utilizing NIRS as an offline analytical tool for the evaluation of milk quality were published in 2006 by the International Dairy Federation (IDF) and the International Organization for Standardization (ISO) [64]. These guidelines were updated in 2020 to cover a broader range of milk and dairy products, including liquid, semi-solid, and solid forms thereof [65]. Near-Infrared spectroscopy applications across the milk chain are divided into four categories, namely off-line (laboratories), at-line, on-line and in-line installations (Figure 9) [14].

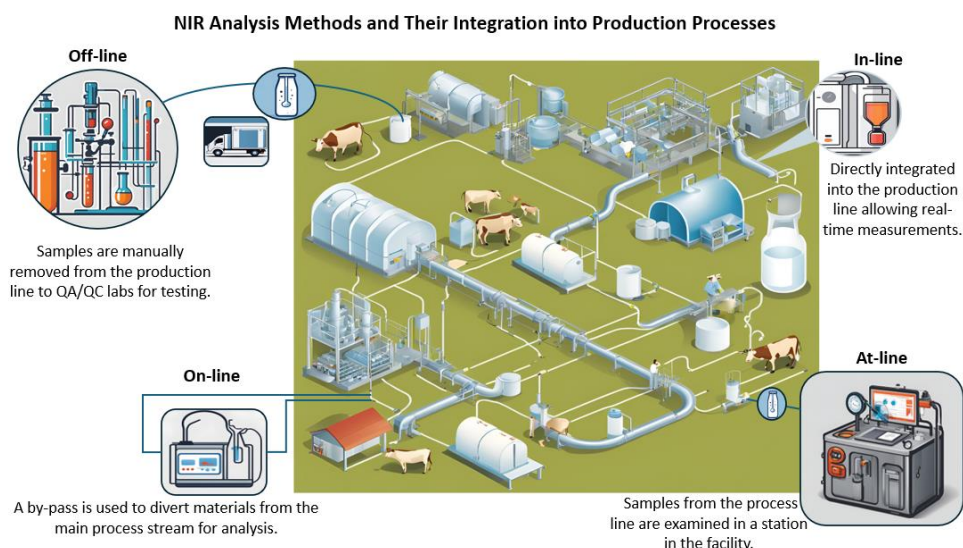


Figure 9. Near-Infrared Spectroscopy analytical methods and their integration into production processes.

- Off-line: NIRS systems are located in quality assurance/quality control (QA/QC) labs; samples are manually collected from the production line for testing.
- At-line: Samples are collected from the milk-processing line and tested using NIRS systems which are positioned near the line.
- On-line: NIRS systems are located at the sampling point; a sample bypass is used to divert materials from the main process stream to be analyzed by the NIRS systems.
- In-line: NIRS system is directly incorporated into the production line, utilizing various sampling techniques that allow real-time measurements.

In contrast to off-line and at-line methods, which involve manual sampling and subsequently delays between sampling and measurement [14], on-line and in-line installations of NIRS provide real-time automatic data collection, reducing manual handling and enabling continuous monitoring and data recording. Real-time NIRS systems can also be integrated into industrial control platforms like Supervisory Control and Data Acquisition (SCADA) systems, enabling the continuous optimization of the processes; however, this integration may encounter technical and cost-related challenges [14].

This review focuses mainly on off-line (benchtop) techniques and in-line (portable/handheld instruments).

4.4.1.2. Near-Infrared Spectroscopy Systems for Milk Analysis

Near-Infrared spectroscopy systems have been studied and extensively used in laboratories for analyzing key milk components. For instance, Albanell et al. [66] employed NIR reflectance spectroscopy to predict quality parameters in goat milk, analyzing 166 samples to determine fat, protein, casein, total solids (TS), and SCC. Similarly, Revilla et al. [67], evaluated the content of different FAs and vitamins A and E using NIR reflectance spectroscopy on 219 ovine milk samples while, Holroyd et al. [68] summarized the NIR bands linked to distinct chemical components in a range of dairy products. Table 2 shows the corresponding wavelengths for the measurement of specific compounds in liquid milk.

Table 4. Liquid milk NIR band assignments [68].

Compound Assignment	Wavelength (nm)
N-H, protein	904, 1014, 1031, 1720, 1758, 2196, 2296, 2334 [69,70]
O-H, C-H lipids	2076, 2376 [69]
Carotenoids	400 – 700 [69]
O-H, water	1454, 1984, 1953 [71]
O-H, N-H	1953, 2048 [71]
Attributed to high somatic cell count	782, 788, 908, 980, 1068 [72]

Aernouts et al. [59] evaluated two distinct spectroscopy measurement modes, reflectance and transmittance, as well as a range of Vis and NIR wavelengths to analyze raw cow’s milk composition in fat, protein, lactose, and urea. Based on their findings it was concluded that reflectance outperforms for measuring crude protein and fat, with R^2 reaching 0.997 and 0.959, respectively. The prediction capacity of lactose was weaker in the case of reflectance, $R^2 = 0.706$, while in transmittance mode the prediction reached $R^2 = 0.883$. However, neither mode provided acceptable predictions for the urea content. In another study, Coppa et al. [73] employed NIRS in reflectance mode to predict milk FAs profile in both liquid and dried milk samples, originating from 419 individual cows. The spectra were obtained by a Foss NIRSystems 6500 NIR scanning spectrometer (Foss NIRSystems, Silver Spring, MD, USA) and the scans were conducted in 2 nm intervals from 400 to 2498 nm. The total saturated fatty acids (SFA), total mono-unsaturated fatty acids (MUFA), and total unsaturated fatty acids (UNSAT) were predicted with success for liquid and dried milk samples. In that study, values of coefficient of determination in cross-validation (R^2CV), coefficient of determination in external validation (R^2V), and ratio of standard deviation of reference data in the calibration set to residual predictive deviation (RPD) ranged from 0.89 to 0.97, 0.86 to 0.95, and 2.93 to 6.25, respectively. Núñez-Sánchez et.al. [74] used both reflectance and tranflectance mode of NIRS to determine the milk fatty acid profile in goats. In reflectance mode, 805 oven-dried samples were used, with the fatty acids’ coefficients of determination of cross validation ranging from 0.80 to 0.47. On the other hand, for tranflectance mode 220 liquid and equal number of oven-dried milk samples were used. In that case, the coefficients of determination of cross validation ranged from 0.11 to 0.79 for liquid samples and from 0.23 to 0.78 for oven-dried samples, with the spectra for reflectance and tranflectance in spectral regions ranging from 400 to 2500 nm.

Table 5. Near-Infrared Spectroscopy (NIRS) applications and performance.

Wavelength (nm)	Type of milk sample	No of samples	Origin of milk	Application	R^2		RMSE/SEP		Accuracy (%)	Ref.
					refl	trans	refl	trans		
1000 – 1700 ^{refl} 1000 – 2500 ^{trans}	liquid	300	cow	fat	0.997	0.997	0.047% _p	0.043% _p	-	[59]
				crude protein	0.959	0.927	0.099% _p	0.133% _p		
				lactose	0.300	0.768	0.282% _p	0.162% _p		
				urea	-	-	-	-		
1445 – 2348	liquid HM liquid UM	166	goat	fat	0.98 _{HM, R}	0.98 _{UM, R}	-	-	-	[66]
				protein	0.96 _{HM, R}	0.95 _{UM, R}				
				casein	0.91 _{HM, R}	0.92 _{UM, R}				
				total solid	0.94 _{HM, R}	0.95 _{UM, R}				
				SCC	0.79 _{HM, R}	0.74 _{UM, R}				
851 – 1649	liquid	785	cow	fat	0.998		0.09% _{SEP}		-	[27]
				protein	0.98		0.05% _{SEP}			
				lactose	0.92		0.06% _{SEP}			
				urea	0.82		19.3 mg/L _{SEP}			
				SCC _{log}	0.85		0.18 _{SEP}			
1500 – 2500	powder	409	retail	protein	0.966 _p		0.547% _p		-	[75]
700 – 1100	liquid	384	cow	SCC	0.76		-		-	[72]
400 – 2500	liquid	242	cow	carotenoids	0.09 – 0.63		0.01 – 0.15 µg/mL _{SEP}		-	[76]
				vitamins	0.01 – 0.69		0.15 µg/mL – 611.82 pg/mL _{SEP}			
				FAs	0.07 – 0.96		0.12 – 4.13 g/100g _{SEP}			
400 – 2498 ^{refl}	oven dried	805	goat	FAs	0.80 – 0.47		0.06 – 2.99 g/100g _{SEP}		-	[74]
400 – 2498 ^{trans}	liquid	220	goat	FAs	0.11 – 0.79		0.05 – 2.81 g/100g _{SEP}		-	[74]
	oven dried	220			0.23 – 0.78		0.05 – 3.35 g/100g _{SEP}			
400 – 2498	liquid	468	cow, bulk	FAs	0.00 – 0.91 _v		0.11 – 3.93 g/100g _{SEP}		-	[73]
	oven-dried				0.20 – 0.95 _v		0.03 – 3.25 g/100g _{SEP}			
400 – 2498	liquid	215	cow	FAs	0.29 – 0.92 _v		0.08 – 2.34 g/100g _{SEP}		-	[77]
	oven-dried				0.46 – 0.97 _v		0.05 – 1.00 g/100g _{SEP}			
600 – 1100	liquid	ND	retail	pH	-		0.031 pH unit		88.0 – 93.0	[70]

≈1100 – 2500	powder	50	vetch root	milk origin	-	-	91.5	[53]
1100 – 2500	liquid	690	retail	melamine ^A	-	-	-	[78]
	powder	660						
	infant formula	660						
1000 – 2500	powder	110	infant formula	melamine ^A	-	0.28 – 0.31 % _p	-	[79]
1000 – 2500	liquid	150	cow	scattering in NIR absorption	-	-	-	[71]
1100 – 2498	liquid dried	219	sheep	summer milk winter milk	-	-	liquid: 79.0	[67]
							dried: 89.0	
							liquid: 78.0	
							dried: 93.0	
400 – 2498	oven-dried	486	cow	cow feeding-type classification	-	-	91.5 - 95.5	[69]

*R*²: Coefficient of determination, RMSE: Root Mean Square Error, ND: Not Defined, , SCC: Somatic Cell Count, FAs: Fatty Acids, Acc: Accuracy *HM*: Homogenized milk, *UM*: Unhomogenized milk, *R*: correlation coefficient, *SEP*: Standard Error of Prediction, *p*: RMSEP (root mean square error of prediction), *refl.*: NIRS reflectance, *trans.*: NIRS transfectance, *transms.*: NIRS transmittance, *v*: validation, ^A: adulteration.

4.4.1.3. Handheld and Portable Near-Infrared Spectroscopy Systems

Handheld and portable NIRS systems, have enabled real-time milk analysis in dairy farms, facilitating the rapid, non-destructive, monitoring of milk composition at the POC. They are small-sized, portable devices with a remarkable analytical capacity of critical milk compounds such as fat, protein, and lactose, without requiring sample pretreatment or extensive laboratory testing. For example, the Polychromix PHAZIR™ (PhIR, Phazir 1624, Polychromix Inc., Wilmington, MA, USA) is a MEMS (micro-electro-mechanical system) incorporating a digital transform spectrometer that operates in reflectance mode within the wavelength range of 1600 to 2400 nm. Llano Suárez et al. [80] used this spectrometer to monitor the FAs content in 108 raw, untreated, cow milk samples at room temperature. The standard normal variate and Savitzky-Golay derivatives (first and second) were used as mathematical pretreatment, while spectral pretreatment was applied, and PCA was employed to eliminate outliers. Partial least squares (PLS) were used to build the regression model, and the highest R^2 values for external validation were obtained by linoleic and capric acids (0.92 and 0.87, respectively). Another application of the portable NIRS devices refers to their capacity to discriminate between organic and non-organic milk; for this purpose, Liu et al. [81] used an ultra-compact spectrometer (Micro-NIR 1700, JDSU, Milpitas, CA, USA) operating between 908 and 1676 nm, with a sampling step of 6 nm. Although the results were useful for an initial on-site analysis, they were outperformed by the Fourier transform (FT)-NIR spectral data produced by benchtop NIRS instruments like the NIRFlex N-500 (Buchi AG, Flawil, Switzerland). Nevertheless, portable NIRS instruments exhibited promising performance for rapid evaluation of the composition and quality of milk at the POC. In another study, de la Roza-Delgado et al. [82] utilized a similar handheld spectrometer (MicroPHAZIR™ from Thermo Scientific) to measure protein, fat, and solids-non-fat (SNF) in cow milk. The calibration models, based on 552 milk samples, showed excellent predictive accuracy for fat, protein, and SNF content. A significant output of this research was the capability to successfully share calibration data between various operation units, demonstrating the suitability of portable NIRS instruments for applications related to the dairy industry. During an 8-week study, on a cattle farm, Diaz-Olivarez et al. [83] collected over 1000 NIR transmittance spectra demonstrating the technology's feasibility for extensive, real-time milk analyses; for that study, an online analyzer, operating between 960 and 1690 nm was used. Each milk sample was measured 100 times with 100-ms integration time, while an average spectrum was used for predictions. Two predictive models were developed: a post-hoc model trained on a representative set of samples ($n = 319$), and a real-time model using the first week's samples for training and the remaining seven weeks for testing the model's performance. For the post-hoc and the real-time models, the root-mean-squared error of prediction (RMSEP) was less than 0.080% and 0.092%, respectively. The post-hoc R^2 values for fat, protein, and lactose were 0.989, 0.689, and 0.947, respectively, while the real-time R^2 values were 0.989, 0.644, and 0.894, respectively. The integration of this system into automated milking systems appears promising, as it allows for the monitoring of each individual cow's milk quality during milking.

Table 6. Handheld Near-Infrared Spectroscopy (NIRS) applications and performance.

Wavelength (nm)	Type of milk sample	No of samples	Origin of milk	Application	R^2		RMSE/SEP		Diagnostic performance	Ref.
1600 – 2400	liquid	108	cow	FAs	0.01 – 0.92		0.01 – 1.57 g/100g _{SEP}		-	[80]
908 – 1676	liquid	87	retail	O / NO classification	-		-		Se: 59.0% Sp: 81.0% Acc: 73.0%	[81]
1600 – 2400	liquid	542	cow	fat	0.971		0.126 % _{SEP}		-	[82]
				protein	0.758		0.124 % _{SEP}			
				SNF	0.612		0.221% _{SEP}			
≈ 1600 – 2400	powder	110	infant formula	melamine ^A	-		0.33 – 0.35 % _p		-	[79]
≈ 1100 – 2200	powder	110	infant formula	melamine ^A	-		0.27 – 0.30 % _p		-	[79]
960 – 1690	liquid	1270	cow	fat	0.989 _{p-rl}	0.989 _{p-ph}	0.083 _{p-rl} *	0.078 _{p-ph} *	-	[83]
				protein	0.894 _{p-rl}	0.947 _{p-ph}	0.110 _{p-rl} *	0.080 _{p-ph} *		
				lactose	0.644 _{p-rl}	0.689 _{p-ph}	0.092 _{p-rl} *	0.077 _{p-ph} *		
800 – 1060	liquid	81	cow	fat	0.88		0.08 % wt _p		-	[84]
				casein	0.89		0.13 % wt _p			
				whey	0.91		0.07 % wt _p			

R^2 : Coefficient of determination, RMSE: Root Mean Square Error, ND: Not Defined, O: Organic, NO: Non-Organic, Se: Sensitivity, Sp: Specificity, Acc: Accuracy, SNF: Solids-Not-Fat, FAs: Fatty Acids, p : RMSEP (root mean square error of prediction), _{SEP}: Standard Error of Prediction, _{p-rl}: prediction real-time, _{p-ph}: prediction post-hoc, ^A: adulteration* % wt/wt.

4.4.2. Mid-Infrared Spectroscopy (MIRS)

Mid-infrared spectroscopy was one of the first methods employed for the analysis of milk to detect trace amounts of adulterants like urea and synthetic milk [85], due to its high predictive accuracy. Mid-Infrared Spectroscopy principles are similar to the ones described for NIRS with regard to absorption, emission, and reflection; however, they refer to the Mid-Infrared region of electromagnetic spectrum, from 10 μm to 2.5 μm . Mid-Infrared spectroscopy estimates the vibrational modes of molecules for the identification and measurement of a broad range of chemical compounds. It is based on the absorption of light energy by the molecular bonds, which makes them vibrate, bend, or stretch in the MIR spectrum in a process which reveals precise details about the chemical composition and structure of the tested substance.

Since MIRS was introduced as a useful tool for the chemical analysis of milk, several studies have exploited it for analytical purposes. For example, Etzion et al. [86] investigated the protein content in raw cow milk using MIR total reflectance spectroscopy; for their experiments they used 235 spectra of raw milk, the Foss Milkoscan 605/255 as “gold standard” and a Vector 22 spectrophotometer (Bruker, Inc. Ettlingen, Germany) to obtain their measurements. Finally, they used two statistical methods for protein content estimation i.e. i) PLS and ii) PCA followed by a NN. Their tests resulted in 0.22% prediction error using PLS and 0.20% using the NN based exclusively on the PCA, while they managed to reduce this error to 0.08% when they included the fat and lactose concentrations in the model. Moreover, Dabrowska et al. [87] utilized MIRS and an experimental set up to estimate the intensity reduction of the light transmitted through a milk sample at different frequencies; the goal was to identify and quantify proteins in the sample. A tunable quantum cascade laser was used (Hedgehog, Daylight Solutions Inc., San Diego, CA) to record the broadband absorption spectra in the region between 1470 and 1730 cm^{-1} . Finally, PLS was performed for the multivariate quantification of protein. The R^2 values obtained were > 0.98 , indicating a satisfying overall performance of the laser. Another study that used MIRS and specifically transmittance data points obtained by Milkoscan FT6000 (Foss Electronics) was performed by Frizzarin et al. [88]. The main objective of this research was to examine various technological properties of milk, such as detailed protein fraction, casein micelle size (CMS), and pH, with a particular emphasis on the utilization and assessment of ML techniques (NNs, SVM, Random Forest, etc.). The prediction accuracy was 0.62 with $R^2 = 0.08$, and 0.80 with $R^2 = 0.65$ for CMS and pH, respectively. For protein traits the accuracy and R^2 measurements ranged from 0.42 and 0.19 for β -lactoglobulin A (β -LG A) to 0.48 and 0.47 for α_{21} -CN, respectively. Mid-infrared spectroscopy was also employed by De Marchi et al. [89] to predict coagulation properties, titratable acidity, and pH of bovine milk. Spectral data were acquired from 1064 liquid samples in the spectral range of 900 to 4000 cm^{-1} using a Milko-Scan FT120 FTIR interferometer. The predictive models developed through this work were able to discriminate between high and low values of pH ($R^2 = 0.59$) and rennet coagulation time (RCT) ($R^2 = 0.62$). Finally, an approximate prediction was also given by the titratable acidity models ($R^2 = 0.66$).

In their review, De Marchi et al. [90] focused on the ability of MIRS to predict a variety of phenotypes by milk analysis such as i) milk FAs profile, ii) coagulation properties and acidity of milk, iii) milk protein fraction and mineral composition, and iv) health and energy status through ketosis prediction. Furthermore, the importance of chemometric analysis (e.g. PLS) is underlined for the successful prediction of the above-mentioned traits. Finally, the potential use of MIRS in the future for the prediction of additional traits is discussed, as well as the likelihood of being utilized for milk recording protocols integrated into selective breeding programs. In another review by Ceniti et al. [91] the use of MIRS for the determination of adulterants in milk as well as the plethora of other applications such as, identification of milk origin, detection of toxins, and detection of drug residuals and other chemical are thoroughly presented and discussed.

Table 7. Mid-Infrared Spectroscopy (MIRS) applications and performance.

Wavelength (cm^{-1})	Type of milk sample	No of samples	Origin of milk	Application	R^2	RMSE/SEP	Accuracy (%)	Ref.
1000 - 4000	liquid	235	cow	protein	-	PLS: 0.22% NN: 0.08%	-	[86]
1470 – 1730	L/ph powder	ND	cow	protein	0.974 _c	0.765 mg mL ⁻¹ _{cv}	-	[87]
400 – 4000	powder	409	retail	protein	0.990 _{pr}	0.294% _p	-	[75]
All MIR excluding: 1600 – 1710 2990 – 3690 > 3822	liquid	730	cow	CMS	0.08	25.286 mm _{cv}	0.62	[88]
				pH	0.65	0.061 pH unit _{cv}	0.80	
				protein traits	0.19 – 0.47	0.255 – 1.759 g/L _{cv}	0.41 – 0.48	
				RCT	0.50	6.397 min _{cv}	0.75	
525 – 4000	liquid	242	cow	carotenoids	0.01 – 0.50	0.01 – 0.19 µg/mL _{SEP}	-	[76]
				vitamins	0.02 – 0.40	0.15 µg/mL – 907.3 pg/mL _{SEP}		
				FAs	0.01 – 0.34	0.13 – 12.63 g/100g _{SEP}		
1000 – 5000	liquid	215	cow	FAs	0.33 – 0.94 _v	0.06 – 1.14 g/100g _{SEP}	-	[77]
900 – 4000	liquid	1064	cow	RCT	0.62	2.36 min _{cv}	-	[89]
				titratable acidity	0.66	0.26 SH ^o /50 mL _{cv}		
				pH	0.59	0.08 Ph unit _{cv}		
500 – 4000	liquid	690	retail	melamine ^A	-	-	-	[78]
	powder	660						
	infant formula	660						
1450 – 1600	liquid	310	retail	(w, sm, su, u, hp) ^A	0.96, 0.94, 0.98, 0.98, 0.90	(2.33, 0.06, 0.41, 0.30, 0.01) g/L _{SEP}	-	[85]

R^2 : Coefficient of determination, RMSE: Root Mean Square Error, ND: Not Defined, PLS: Partial Least Squares, NN: Neural Networks, L/ph: Lyophilized, RCT = rennet coagulation time, FAs: Fatty Acids, Acc: Accuracy, CMS: Casein Micelle Size, RCT: Rennet Coagulation Time _p: RMSEP (root mean square error of prediction, _{cv}: RMSE of cross validation, _{SEP}: Standard Error of Prediction, _c: cross-validation, _v: validation, _{pr}: prediction, w: whey, sm: synthetic milk, su: synthetic urea, u: urea, hp: hydrogen peroxide, ^A: adulteration.

4.5. Other Spectroscopy Methods

Beyond the above-mentioned methods, there are more spectroscopic methods that have been utilized for milk analyses such as Fourier transform infrared (FTIR), fluorescence, and UV absorption. As noted by Fox et al. [92], milk absorbs light between 200 and 380 nm due to its protein content. Furthermore, there is a correlation between the percentage of fat in the milk and the light absorption measured between 400 and 520 nm. These wavelengths are in the UV/Vis region of the electromagnetic spectrum and consist of a primary example of how different techniques can be applied to milk analyses.

Similar to other IR spectroscopy methods, FTIR spectroscopy is used to measure the IR spectrum of materials' absorption or emission. It is a type of MIRS that enables the quick scanning of the MIR region of the spectrum [91]. The technique is called FTIR spectroscopy due to the Fourier transformation used to convert the raw data into the actual spectrum. This method has been used by several researchers for the study of milk compounds. Among them, Nicolaou et al. [93] used FTIR to detect and quantify milk originating from different ruminant species. In that study, a Bruker Equinox 55 infrared spectrometer was used to acquire approximately 400 spectra of milk mixtures; after developing a set of multivariate analyses FTIR demonstrated promising results with regard to the measurement of milk compounds such as casein and urea [94].

In the study by Fragkouli et al. [95], FTIR reflectance, fluorescence, and UV absorption, were applied to determine the milk fat content and the ruminant species of milk origin in 23 commercial milk samples, including 11, 9, and 3 bovine, caprine, and ovine samples, respectively. The study achieved 96% accuracy in determining milk fat content using UV absorption, and 91% accuracy when combining UV absorption and fluorescence for identifying the ruminant species the milk originated from.

Fluorescence has a variety of applications in the dairy industry mainly on dairy products rather than in the raw milk analyses [96,97]; for example, an application on the determination of melamine used to reveal adulteration in milk has been studied by Barreto et al. [98]. Melamine's use as milk adulterant for testing, specifically targets to evaluate the performance of milk protein adulteration, due to the melamine's high nitrogen content and water solubility.

Visible spectroscopy has also been exploited by Aernouts et al. to evaluate milk's composition [59], while in a relevant study by Bogomolov et al. [99] visible light scatter was applied to quantify milk fat and protein content; RMSE values equal to 0.05% and 0.03% were observed for milk fat and protein contents, respectively, concluding that visible spectroscopy could be successfully applied in both laboratory and in-line/POC measurements to replace traditional NIR methods.

Moreover, Yang et al. [100] designed and evaluated a portable milk analyzer using a miniature UV/Vis spectrometer. The UV/Vis absorption spectra were collected, and PLS algorithms were developed for the prediction of fat, protein, lactose, and TS contents in high-pressure homogenized and in raw milk samples. Concerning raw milk, the results were promising but obviously less accurate compared to the ones achieved by homogenized samples.

Table 8. Applications and performance of other spectroscopic methods.

Spectroscopy Method	Wavelength (nm)	Type of milk sample	No of samples	Origin of milk	Application	R^2		RMSE	Accuracy (%)	Ref.	
FT-IR	600 – 4000 (cm^{-1})	liquid	63	cow ^R , goat ^R , sheep ^R	composition	0.92		6.40* _{<i>p</i>}	-	[93]	
						0.93		5.61* _{<i>p</i>}			
						0.96		3.98* _{<i>p</i>}			
FT-IR	400 – 4000 (cm^{-1})	liquid	23	cow ^R , goat ^R , sheep ^R	fat content animal of origin	-		-	78.0 74.0	[95]	
Ultraviolet	220 – 400	liquid	23	cow ^R , goat ^R , sheep ^R	fat content animal of origin	-		-	96.0 91.0	[95]	
Fluorescence	240 – 500 _{exc} 290 – 750 _{em}	liquid	23	cow ^R , goat ^R , sheep ^R	fat content animal of origin	-		-	70.0 91.0	[95]	
Fluorescence	250 – 380 _{exc} 280 – 640 _{em}	liquid	40	cow	milk origin cls.	-		-	76.9 [†] 70.4 ^{††}	[101]	
Fluorescence	250 – 550 _{exc}	liquid	242	cow	carotenoid vitamins FAs	0.01– 0.54 0.03 – 0.17 0.01 – 0.50	0.01 – 0.17 µg/mL _{SEP} 0.17 µg/mL – 918.32 pg/mL _{SEP} 0.15 – 13.76 g/100g _{SEP}		-	[76]	
Fluorescence	240 – 260 _{exc} 320 – 440 _{exc}	liquid	12	retail	melamine ^A	0.97 ⁺⁺⁺ 0.95 ⁺⁺⁺	PARAFAC: 68.6 ppm _{<i>p</i>} U-PLS/RBL: 81.9 ppm _{<i>p</i>}		-	[98]	
Fluorescence	330 _{exc} 420 _{em}	liquid	23	ND	heat treatment discrimination	> 0.95		-	-	[102]	
Fluorescence	250 – 350 _{exc} 260 – 500 _{em}	liquid	30	cow	characterization of pasteurized milk	-		-	-	[103]	
Visible	400 – 1000 _{refl} 400 – 1000 _{trans}	liquid	300	cow	fat crude protein lactose	refl	trans	refl	trans	-	[59]
						0.978	0.395	0.11% _{<i>op</i>}	0.629% _{<i>p</i>}		
						0.861	0.687	0.18% _{<i>op</i>}	0.274% _{<i>p</i>}		
						0.557	0.111	0.22% _{<i>op</i>}	0.317% _{<i>p</i>}		

urea									
-									
-									
-									
Visible light scatter	400 – 1000	liquid	21	retail	fat protein	0.973	0.047%	-	[99]
0.964									
0.032%									
UV/Vis	183 – 667	liquid FR	240	cow	fat, protein, lactose,	-	Liquid FR 0.13% _p – 0.46% _p	-	[100]
liquid HPH									
240									
TSC									
HPH FR 0.09% _p – 0.27% _p									
Fusion NIRS-LIBS	≈ 185 – 2500	powder	50	vetch root	milk origin	-	-	95.8	[53]

R^2 : Coefficient of determination, RMSE: Root Mean Square Error, ND: Not Defined, Acc: Accuracy, FR: Fresh Raw, HPH: High-pressure Homogenized, TSC: Total Solids Concentration FAs: Fatty Acids, clss: classification ^R: retail, * percentage volume, _p: RMSEP (root mean square error of prediction), ^{exc}.: excitation, ^{em}.: emission, ^{refl}.: Visible reflectance, ^{trans}.: Visible transmittance, _v: validation, ^Δ: adulteration [†]: based on aromatic amino acids and nuclei acids fluorescence spectra, ^{††}: based on rivo flavin fluorescence spectra, ^{†††}: predicted x reference concentration correlation coefficient, [·]

4.6. Benchmarking of Spectroscopy Methods

To accurately assess the efficiency of each spectroscopy method benchmarking on the same implementation is critical. For instance, Domingo et al. reviewed in their paper [104] the capacity to detect melamine in milk using MIR, NIR and Raman spectroscopy; they concluded that Raman must be further studied, as it is likely to effectively detect and quantify melamine content, but there is still a demand for further work to elucidate its diagnostic value. Concerning MIRS and NIRS, they both produced similar results with the PLS being the most used method for analyzing the data and comparing the methods. Similarly, melamine detection using MIRS and NIRS has been examined by Balabin et al. [78], who also concluded that both techniques are suitable for this application resulting to a limit of detection lower (LOD) than 1 ppm (0.76 ± 0.11 ppm), while Wu et al. [75] found NIRS and MIRS to have very similar performance when utilized for milk protein measurement, with R^2 being 0.966 and 0.990 and the RMSEP being 0.5473 and 0.2944, respectively.

Comparisons between NIRS, MIRS, and molecular fluorescence was the primary focus of the study by Soulat et al. [76]; in their research they aimed to determine the best method to predict carotenoid, vitamin and FAs content in bovine milk. Fatty acids and some carotenoids (cis9- β -carotene, β -cryptoxanthin and zeaxanthin) were more efficiently predicted using NIRS, whereas other carotenoids (13- β -carotene, the sum of β -carotenes) were better predicted by fluorescence. Nevertheless, the prediction capacity of vitamins was relatively poor, irrespective of the method used. Moreover, in the same study, MIRS outperformed the other methods when used for the prediction of lutein and α -tocopherol. A broader comparison between fluorescence, MIR, and NIR spectroscopy was presented in the review by Loudiyi et al. [94] who concluded that fluorescence spectroscopy is more sensitive compared to absorption measurements due to the zero background of the measured signal. However, it is worth noting that only one device has been developed by Spectralys Innovation (Amaltheys®) and has been proposed for the dairy industry, in contrast to IR spectroscopy where more industrial applications are available as being faster and cheaper. Indeed, as Loudiyi et al. [94] discussed, the objective of many of the available studies was to create a real-time milk analysis system, rather than actually test its capacity to perform the measurements under real-world conditions.

The idea of combining spectroscopy methods is an innovative approach that expands the applicability of spectroscopy and marks new research pathways to explore, with the fusion of spectroscopy techniques being already exploited in some cases. A successful example is described by Eum et al. [53] who combined NIRS and LIBS to identify the origin of milk, with remarkably positive results. When the two methods were individually considered, accuracy values were 91.5% and 73.1% for NIRS and LIBS, respectively, whereas when the two methods were jointly considered, the accuracy reached 95.8%.

5. Machine Learning Principles

Sophisticated analytical tools are necessary to extract meaningful insights from the massive amount of data generated by spectroscopy techniques such as Raman, LIBS, NIRS, and MIRS. Among them, ML algorithms have been efficiently utilized for the improvement of these spectroscopic techniques' predictive capacity. Indeed, with the integration of ML models, including regression analysis, NN, and SVM in spectroscopy systems, scientists have achieved more precise and effective predictions for milk composition, quality, and adulteration. The following sections will explore how ML algorithms are applied to the data acquired from spectroscopy techniques, offering new potential for the precision management of dairy farms and real-time milk analysis.

Automated monitoring and recording tools and AI are basic components of the PLF systems and can be used to efficiently address production, health, and welfare challenges by indicating early signs of potential production challenges, management errors, and diseases in dairy farms [105]. Artificial Intelligence, is defined by Kaplan and Haenlein as 'the ability of a system to accurately interpret external data, learn from it, and apply that knowledge to achieve specific goals and tasks through flexible adaptation' [106]. Therefore, AI employs knowledge-based rules (supplied by developers) or

recognizes the rules and patterns that underpin the application of ML to drive systems to predefined objectives. It also acts on external information from Internet of Things (IoT) platforms and other big data sources [107].

The two types of data modeling currently utilized by PLF systems as AI components are the predictive and the exploratory ones. Predictive models use data to forecast future events based on predefined criteria, while exploratory models analyze past events to identify key determinants [3].

Modeling-based approaches that involve the collection and analysis of data, risk assessment, and ML are frequently seen, and ML algorithms have been extensively integrated into modeling and simulation modules for the analysis of data collected by livestock sensors. Therefore, the volume of data being collected by livestock farms via PLF monitoring systems has significantly increased lately, necessitating the training of ML algorithms to automatically generate efficient DSS [107].

Data processing and analysis techniques are divided into two primary categories: 1) modeling and simulation-based techniques, and 2) ML and data analytics algorithm-based techniques. Combining these techniques significantly improves the efficiency and reliability of DSS. In fact, the integration of data analysis, ML, simulation, and modeling tools broadens the scope of this data-driven strategy; once data are collected, they are appropriately analyzed to produce information about the current state of the farm and support relevant management interventions (Figure 10). The process begins with simulating a ruminant farm in a controlled environment. However, simulation on its own is insufficient because of the complexity of actual ruminant farms [108]. In order to bridge the gaps and provide holistic and targeted solutions, ML and other data analysis techniques are used [107].

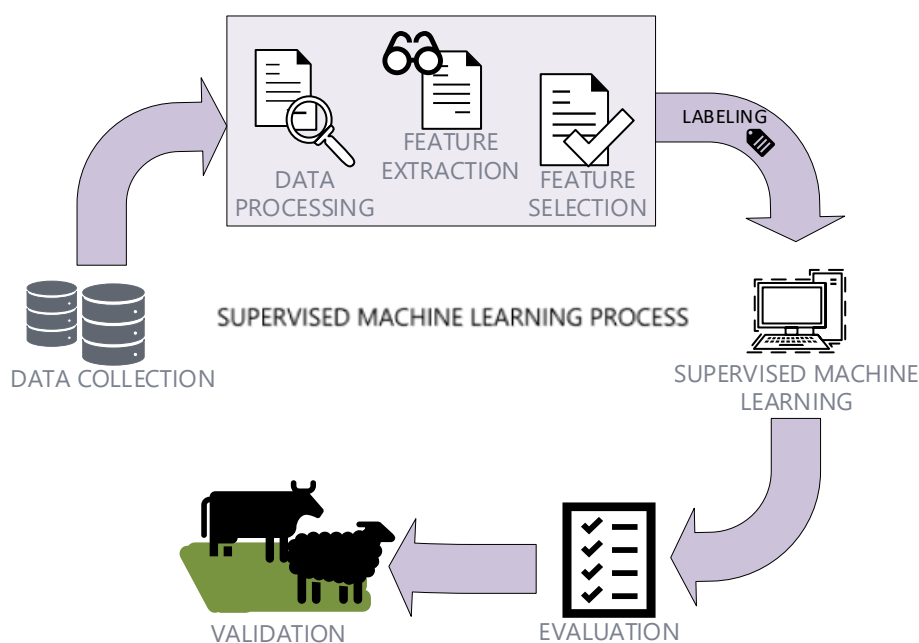


Figure 10. Supervised ML process of data.

Digital livestock farming systems support evidence-based animal production, as well as the health and welfare of farm animals, relying on data, collected from biometric and biological sensors, which are then appropriately analyzed to create predictive models [3]. Farmers may increase the health status of their animals and the sustainability of their farms by using real-time data analysis to make informed decisions based on the processing of large-scale, sensor-derived data [109]. These datasets function as the foundation for ML algorithms, which analyze them to improve the diagnostic and predictive system performance and enable the development of automated DSS [3,110,111].

In PLF, the main categories of ML refer to supervised learning, unsupervised learning, active learning, generative adversarial networks (GANs), and few-shot learning. Among these ML categories, supervised learning has been mostly utilized on dairy ruminant and milk analysis applications. This is associated with the capacity of the supervised models to be trained on a dataset

that includes both inputs and their corresponding outputs (labels), aiming in achieving the correct interrelation mapping between them.

Supervised learning comprises various methods concerning model training with labelled data. Regarding milk analysis, the most applied methods are linear regression, LR, DTs, Random Forest (RF), SVM, k-NN, Naive Bayes (NB), GBM, AdaBoost, NN, Linear Discriminant Analysis (LDA), PLS, and Partial Least Square Regression (PLSR). The main supervised ML algorithms used in dairy ruminant research are illustrated in Figure 11, with those framed in red being the focus of the following sections.

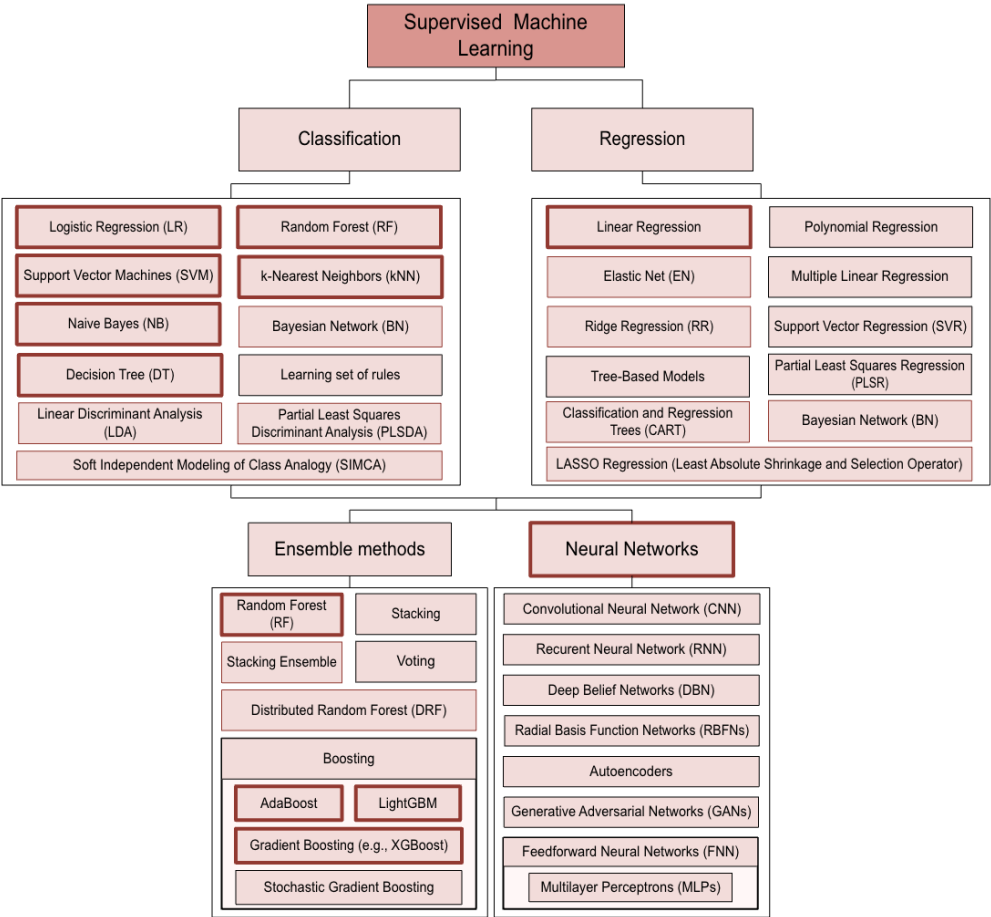


Figure 11. Supervised ML methods applied in dairy ruminants and milk analysis research [112–114].

5.1. Logistic Regression (LR)

Logistic regression is a statistical method used for building ML models that predict the probability of a discrete outcome, typically a binary one, based on a set of independent (explanatory) variables [115]. It estimates the relationship between a categorical dependent variable and the explanatory factors, allowing for the prediction of the likelihood of an event’s occurrence. As a supervised ML algorithm for solving classification problems, LR aims to find the minimum value of the loss function to enhance the accuracy of the prediction function, thereby solving the classification problem [116].

5.2. Decision trees (DTs)

Decision trees is a non-parametric supervised learning method that can be applied in both classification and regression tasks. It has a tree-like and hierarchical structure, with internal, leaf, branch nodes, and a root node [117]. Decision trees algorithms do not require much data preprocessing and can work with both numerical and categorical features. Additionally, while DTs are useful in many different applications, they are frequently insufficient for properly predicting continuous values in regression analyses, while their training can be a challenging and expensive

task. Moreover, although DTs can automatically handle feature selections and inference, they can be sensitive to small data variations, which may lead to significant changes in the tree structure, affecting its stability [115].

5.3. *Random Forest (RF)*

The RF method is an ensemble learning method used for both regression and classification tasks. By choosing random subsets of covariates, it constructs multiple DTs, improving the predictive accuracy and reducing overfitting. The final prediction arises from the weighted average or the majority vote of these trees [115]. Random Forest accomplishes implicit feature selection to generate uncorrelated DTs, making it this way an effective method especially in datasets with numerous features [88,118]. In contrast to linear regression, RF offers insights into features' importance but does not provide thorough coefficient analysis; however, it can be computationally demanding for big datasets. Random forest demonstrates great performance when both numerical and categorical data are analyzed, and usually does not require scaling or variable transformation. Despite its complexity, RF is characterized by strong resistance to noise and overfitting [119].

5.4. *Support Vector Machine (SVM)*

Support Vector Machine is a discriminative ML method that may be equally applied to regression and classification problems. It works by building a hyper-plane to reduce errors, and performs effectively in high-dimensional feature spaces, particularly when there is a distinct separating boundary between the data classes. This makes SVM suitable for problems where the decision boundary is well-defined. Since it uses a subset of training points in the decision function, known as support vectors, it is also memory-efficient. However, due to the longer required training period, it performs poorly with large datasets and particularly when there is extra noise in them, such as target class overlap [115,120].

5.5. *k-Nearest Neighbor (k-NN)*

K-nearest neighbors is a commonly used classification algorithm characterized by its simple implementation and flexibility. It is based on the principle of proximity, where the most common category among the nearest neighbors in the feature space defines the classification of a studied sample [121]. K-nearest neighbors assumes that class conditional probabilities are locally constant, which can introduce bias, particularly in high-dimensional spaces [9]. A key benefit of k-NN is that it does not require any preprocessing of the training data, providing both space and speed advantages when applied in very big datasets. Nonetheless, k-NN usually assumes an equal distribution of training samples among different classes [122]. In numerous practical scenarios, datasets present an imbalanced distribution, where the major class is represented by a large number of observations while the minority class by a few [123]. The imbalanced distribution highlights the significance of choosing the k parameter thoroughly, since it has a direct impact on the classification performance. If the k parameter has a predefined value, it may lead to bias in favor of the major class, especially in cases of uneven distribution of observations assigned to different classes [121,124].

5.6. *Naïve Bayes (NB)*

Naive Bayes is an efficient, incremental ML classifier known for its strong performance in everyday applications, since it can handle both discrete and continuous variables. Despite the assumption of feature independence, an often unrealistic condition which may result in poor performance in domains where attributes are highly interdependent, NB can still effectively compete with more sophisticated classifiers, particularly in scenarios with minimal feature interdependencies [125]. Since it explains its decisions through the total amount of information acquired, the algorithm is especially useful for its transparency. When employed iteratively, NB can solve non-linear problems while retaining its inherent advantages [126]. Due to its efficiency and simplicity, this method is particularly used in behavioral models within livestock [117].

5.7. Linear Regression

Linear regression is a statistical and ML method where the value of a dependent variable y is predicted by one or more independent variables x_i (where $i = 1, 2, 3$ etc.) and a .

A simple linear regression model is expressed as below:

$$y = a_0 + a_1x + e \quad (5)$$

where (y) is the dependent variable and (x) is the independent variable. The constant term, a_0 represents the vertical axis intercept of the regression line, a_1 is the regression coefficient that refers to the slope of the regression line, and e is the random residual error [127].

5.8. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis is a frequently used method for reducing dimensionality problems as a preprocessing step for ML and pattern classification applications. The objective of the LDA technique is to project the original data matrix onto a lower dimensional space. When high dimensional feature vectors from various classes are reduced to a lower dimensional feature space, the LDA technique identifies an orientation W that allows the projected feature vectors of one class to be clearly distinguished from those of other classes. For example, two-dimensional feature vectors are reduced to one-dimensional feature vector [128]. Despite the fact that LDA is one of the most commonly used data reduction techniques, small sample size (SSS) and linearity emerge as the two main disadvantages. A linear transformation that discriminates between various classes is found using the LDA technique. Nonetheless, if the classes are not separated linearly, LDA cannot find a lower dimensional space. Thus, when the discriminatory information is not in the means of classes, LDA fails to find its space. Also, SSS as one of the big problems of LDA technique, results from high-dimensional pattern classification tasks or a low number of training samples available for each class compared with the dimensionality of the sample space. Due to the high number of features or dimensionality, LDA technique has been applied in biometric applications, agriculture applications, medical applications etc. [129]. Linear Discriminant Analysis constitutes a specific type of Discriminant Analysis (DA) [130].

5.9. Boosting

Boosting is an ML method based on the idea that a combination of simple classifiers (obtained by a weak learner) can achieve a better performance than any other of the simple classifiers alone. A weak learner is a learning algorithm able to produce classifiers with error probability strictly less than that of random guessing, while, a strong learner is able (given enough training data) to produce classifiers with arbitrarily small error probability [131]. Boosting is a general method for improving the accuracy of any given learning algorithm, focusing primarily on the AdaBoost algorithm [132].

5.9.1. Adaptive Boosting/Adaboost

Adaptive Boosting or AdaBoost is a ML algorithm introduced by Freund and Schapire in 1996, and marked a significant enhancement in boosting techniques [133]. The algorithm repeatedly combines weak learners, typically decision stumps, to create a strong classifier. Each time, the weights of the data points are adjusted to provide misclassified cases, a higher weight and support direct learning towards more demanding cases [134–136]. The algorithm's classification performance is improved in a variety of applications by using this adaptive weighting mechanism. Regression tasks are another area in which Adaboost outweighs, highlighting its ability to efficiently adjust [137]. The mathematical formulation includes calculating a weight for each instance of equation x , as shown by $w_i = \frac{1}{n}$ [138], of n pieces of x data. The algorithm splits input values into binary classes, which are commonly represented by the numbers -1 and 1. AdaBoost's strong classification ability has made it a popular ML method in many fields including livestock farming [139].

$$x_i, x_{i+1}, x_{i+2} \text{ and } y_n \in \{-1, 1\} \quad (6)$$

Although AdaBoost is a well-established boosting technique, it has been used less frequently in dairy applications compared to GBM.

5.10. Gradient Boosting Machine (GBM)

Gradient boosting machine is an upgraded extension of the AdaBoost method, supporting any differentiable loss function. Fitting the tree models to the negative gradient of the loss function yields the difference between the expected and actual values of the outcome variable. According to Friedman [140], this enables GBM to optimize any differential loss function. Utilizing an ensemble model called gradient boosting, a set of poor prediction models is "boosted" to produce a more trustworthy model, with its current base learner being trained primarily on the mistakes that prior base learners have made [135,141].

5.11. Neural Networks (NN)

A NN is an ML method that enables computers to process data in a manner inspired by the human brain's functioning. Five basic components are usually used to analyze NN: the activation function, weight coefficients, bias (constant term), input values, and output values. A NN can be represented mathematically in the form of equation (7).

$$y = w_i \cdot x_i + b \quad (7)$$

where (y) is the output value, (w_i) is the weight coefficient, x_i is the input value, , and (b) is the constant term (Figure 12) [139].

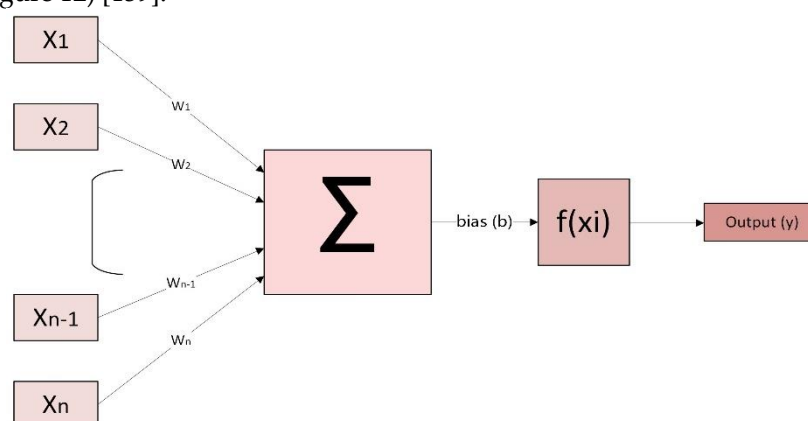


Figure 12. Neural networks model representation.

5.12. Partial Least Square (PLS)

Partial least square analysis is a multivariate statistical method that allows comparison between multiple response variables and multiple explanatory variables. It is considered especially useful for constructing prediction equations when there are many explanatory variables, but comparatively small sample data. Partial least squares analysis was designed to deal with multiple regression when the data has a small sample, missing values, or multicollinearity. It has been widely used in fields like chemistry and chemometrics, where there is a big problem with a high number of intercorrelated variables and a limited number of observations [142].

The intention of PLS is to form components that capture most of the information in the X variables that are useful for predicting $Y_1, \dots, Y_l$, while reducing the dimensionality of the regression problem by using fewer components than the number of X variables [143].

Despite its benefits, it cannot provide significance testing unless bootstrapping is used, while there is a lack of model test statistics [142].

5.13. Partial Least Square Regression (PLSR)

Partial least square regression is the PLS approach in its simplest and most commonly used form. Partial least square regression is a method that advances beyond common regression by modeling the structure of X and Y . For this purpose, it uses two-block predictive PLS models to describe the relationship between two data matrices, X and Y , through a linear multivariate model. The ability of PLSR to analyze data with numerous, noisy, collinear, and even incomplete variables in both X and

Y makes it a useful method, with its precision improving with the increasing number of relevant variables and observations [144]. The PLSR method is commonly used for the construction of rapid, online spectroscopic and image analysis systems in food quality and safety evaluation and control applications. Compared to other linear methods PLSR is simpler, easier to fit models, and allows for determining statistical properties; compared to nonlinear methods, it is more suitable and efficient for analyzing complex problems. However, PLSR is less powerful for predicting complex problems compared to other linear methods, while it has higher computational complexity compared to nonlinear methods [145].

6. Application of Machine Learning Methods in Milk Quality Assessment

Lately, ML methods have been increasingly exploited in the dairy ruminant sector with various applications developed within the framework of the PLF. Among these applications, the ones exploited to assess milk quality and properties are emerging as valuable tools for both farms and the dairy industry. In the current review, we focus on examples of milk chemical analyses, but it is worth mentioning that machine learning techniques have also shown promise in detecting microbiological contaminants such as *Brucella* spp., *E. coli* O157, *Bacillus cereus*, and *Listeria* spp. in milk samples, illustrating their potential in microbiological safety assessments, although we will not address this aspect further [30,146]. Research studies using milk chemical analyses applications are summarized and discussed in the following sub-sections.

6.1. Milk Quality and Composition Assessment

The predictive ability of different regression and classification techniques documented by Frizzarin et al. [88] involved testing 730 milk samples from 622 individual crossbred cows to estimate the protein content, technological properties, and the produced MIR spectra. Samples were collected from animals at different lactation stages and of various parities. Milk technological properties assessed in this study were CMS, heat stability, RCT, curd-firming time (k20), curd firmness at 30 and 60 minutes (a30, a60), and pH, while milk protein fraction was also analyzed to estimate the α S1-casein (α S1CN), α S2-casein (α S2-CN), β -casein (β -CN), κ -casein (κ -CN), α -lactalbumin (α -LA), β -LG A, and β -lactoglobulin B (β -LG B) contents. For the statistical analyses, both regression-based and classification methods were used. Utilizing regression-based approaches, a total of 11 different ML methods were employed, with RF, Boosting, and NN being the most pertinent to this study. Besides regression methods, several classification approaches were also applied, with RF, Boosting Decision Trees, and SVM being the most significant ones. Key results of the regression-based methods showed that NN had the best performance concerning RCT, k20, and heat stability. LASSO (Least Absolute Shrinkage and Selection Operator) and Ridge Regression (RR) showed exceptional performance in forecasting distinct proteins. RF, along with PLSR and RR, assigned the highest coefficients to key wavelength space regions, which were crucial for several traits, including RCT, pH, and α -LA. Based on the classification output by Frizzarin et al. [88], it was evidenced that SVM was the most accurate model to assess binary technological attributes, whereas RF and Partial Least Squares Discriminant Analysis (PLSDA) performed in a similar way in predicting protein fractions. In terms of sensitivity, Boosting performed well but exhibited lower specificity. In particular, SVM achieved the highest accuracy in 6 out of the 7 binary technological traits, including RCT, k20, a30, CMS, pH, and heat stability. For RCT, pH, and heat stability, SVM's accuracy was similar to that of PLSDA's. Moreover, SVM showed the highest accuracy for the prediction of α -LA and β -CN contents. On the other hand, RF was the most accurate method to predict α S1CN and κ -CN contents. For the recognition of coagulating properties, Boosting Decision Trees demonstrated a good sensitivity of 0.98 but poor specificity (0.50), indicating that Boosting could identify coagulating samples with satisfying accuracy, but struggled with false positives [88].

By applying various ML techniques and PLSR, Mota et al. [147] examined the use of FTIR spectroscopy to predict different milk κ -CN phenotypes in Holstein cattle. The research aimed to evaluate the predictive power of RF, GBM, and EN, in comparison to PLSR. The study used phenotypic data from 471 cows, and two cross-validation techniques were applied to evaluate the

models' performance. The κ -CN phenotype was evaluated in terms of predictive performance across the four different models. In the training set, the average predictive ability across 10 replicates was 0.96 for EN, 0.97 for GBM, 0.96 for RF, and 0.90 for PLS. Similarly, for the validation set, the average predictive ability across 10 replicates was 0.79 for EN, 0.81 for GBM, 0.80 for RF, and 0.77 for PLS; the respective RMSE values in the validation set were 1.25, 1.08, 1.18, and 1.41. The study's main findings demonstrated that the GBM continuously produced the best predictions across all phenotypes, with predictive capacities ranging from 0.58 to 0.77 in the herd/date-out cross-validation and from 0.63 to 0.81 in the samples-out random cross-validation. Random forest followed closely, with similar accuracy levels but slightly higher bias. In terms of accuracy, GBM performed better than PLS, RF, and EN (by 7%, 1%, and 4%, respectively). In comparison to PLS, GBM reduced predictive errors by 33%, RF by 26%, and EN by 25%, according to the RMSE analysis. Gradient boosting machine and RF significantly outperformed PLS in predictive accuracy ($P < 0.05$), especially in the samples-out random cross-validation, according to the Hotelling-Williams test. While both ML models outperformed PLS in the herd/date-out scenario, the difference was not statistically significant due to increased variability [147].

The study of Frizzarin et al. [148], provides a comprehensive evaluation of MIR spectra and the effectiveness of various ML methods to evaluate and discriminate milk samples derived from grazing and non-grazing cows. Various performance metrics, including F1 score, accuracy, sensitivity, specificity, and Cohen's kappa coefficient, were estimated. Over a three-year period, the authors collected and analyzed 4320 milk samples, which served as the basis for the comparisons between ML methods. The study tested 11 ML and statistical methods, namely, RR, LASSO, EN, LDA, model-based discriminant analysis (MB-DA), PLS-DA, variable-selection discriminant analysis (VarSel-DA), RF, boosting, principal components linear regression (PC-LR), and SVM. Linear discriminant analysis and PLS-DA emerged as the top-performing models accomplishing the highest accuracy (0.968) and F1 score (0.975); LDA had the highest specificity (0.980), while PLS-DA had the highest sensitivity (0.962). Additionally, MB-DA had also a good performance, with high accuracy (0.968), sensitivity (0.959), and specificity (0.972), indicating its strong discriminatory power, with VarSel-DA closely following, with an accuracy of 0.961 and a strong F1 score (0.971). Principal Components Linear Regression yielded the lowest accuracy (0.667) and specificity (0.117), making it the least effective method. It also had the second-lowest F1 score (0.790), after RF which had the lowest one (0.781), and the lowest sensitivity (0.827), indicating a poor performance on identifying milk from grazing cows. While they did not match the top-performing models in terms of overall accuracy and sensitivity, other techniques like SVM and EN performed quite well [148].

Strong evidence in favor of combining NIR spectroscopy with Artificial Neural Networks (ANN) and Stacking Ensemble to predict blood metabolites from milk samples has been provided by Giannuzzi et al. [149]. Several ML methods were tested to build prediction models for blood metabolites including RF, GBM, ANN, PLS, and Stacking Ensemble model. A total of 385 Holstein dairy cows made up the studied animal population. The AfiLab device was used to analyze NIR spectra of milk samples collected by the cows. Among the studied ML methods, Stacking Ensemble and Multi-Layer Feedforward ANN, outperformed other methods in predicting blood metabolites from milk samples. In that study, moderate correlations between observed and predicted values for key metabolites, like γ -glutamyl transferase ($r = 0.58$), haptoglobin ($r = 0.66$), and total reactive oxygen metabolites ($r = 0.60$) were observed. [149].

Another research by Giannuzzi et al. [150], aimed to predict 29 blood metabolites by using FTIR spectra and ML algorithms applied to milk samples of dairy cows. Primarily, 1204 Holstein cows made up the data set, but the sample size increased to 2701 cows for β -hydroxybutyrate (BHB). The authors used an automatic ML algorithm that tested various prediction methods, including EN, distributed random forest (DRF), GBM, ANN, and Stacking Ensemble. Additionally, the ML algorithms were comparatively assessed with PLSR. Two types of cross-validation (CV) scenarios were used; in the first, data was randomly split into 5 parts (CVr), while in the second data was split by herds (CVh). The Stacking Ensemble method outperformed most of the blood metabolites studied across both CVr and CVh scenarios. In particular, EN and Stacking Ensemble showed up to 75% and

150% improvement in prediction accuracy for CVr and CVh respectively, compared to PLS. The Stacking Ensemble model had the best R^2 values for glucose, urea, total reacting oxygen metabolites, globulins, Na, and ceruloplasmin. However, EN achieved the best R^2 values for albumin and total proteins. The CVh scenario had lower prediction abilities than the CVr, suggesting that herd-specific factors might have influenced the model's performance [150].

In their study, Mota et al. [56], evaluated the potential use of the AfiLab, real-time NIR milk analyzer measurements for the prediction of key cheese-making traits in Holstein cows, namely, k20, CFp (curd firmness peak), CYCurd (cheese yield from curd), CYWater (percentage cheese yield based on water content), and a45 (curd firmness in millimeters at 45 minutes) [56]. The study involved 499 cows from two farms. Several ML methods were applied and compared, including ANN, EN, GBM, XGBoost (eXtreme Gradient Boosting), and Stacking Ensemble. Artificial Neural Networks achieved the highest predictive capacity, with R^2 values ranging from 0.45 (CFp) to 0.71 (CYCurd). Gradient boosting machine followed by R^2 values ranged from 0.45 (%CYWater) to 0.70 (CYCurd), while EN was the last one with satisfying results, with R^2 values ranging from 0.46 (CFp) to 0.70 (CYCurd). XGBoost exhibited the lowest R^2 values, ranging from 0.43 (a45) to 0.63 (k20) [56].

In their study, Samad et al. [9], evaluated milk quality using the k-NN algorithm. Milk samples were classified into three categories based on their quality status (low, medium, and high). The dataset included a total of 1,059 entries, but the method of the milk analysis was not defined. To detect milk quality, traits such as pH, temperature, taste (acceptable or poor), odor (foul or no foul), colour, fat and turbidity (high or low) were recorded and forced into the algorithm. As shown by the confusion matrices, the standard k-NN classifier achieved a high overall accuracy of 98.58%, with particularly strong performance in detecting high-quality milk [9].

Soyeurt et al. [151], utilized four ML algorithms on data collected by MIR spectroscopy for the prediction of lactoferrin (LF) in bovine milk. They collected 6619 milk samples from various herds, breeds, and regions, creating a large dataset consisting of 5541 and 836 records for the training and the validation sets, respectively. Each ML model was evaluated using RMSE and R^2 metrics for both calibration and validation sets. Partial least squares regression, PLS with linear support vector regression (PLS + SVR), PLS with polynomial support vector regression (PLS + Polynomial SVR), and PLS + ANN were the ML methods applied. Among them, PLS + ANN was the most appropriate and reliable model to predict LF content, demonstrating the highest R^2 value (0.60), the lowest RMSE (162.17 mg/L), and the best residual distribution. Moreover, the PLS + ANN model predicted expected LF trends related to milk yield, somatic cell score, and lactation stage, however it tended to underestimate LF values above 600 mg/L. It was concluded that the PLS + ANN model provided the best balance between prediction accuracy and robustness, particularly when predicting extreme values of LF, showcasing its realistic potential to be applied for the LF monitoring as an udder health indicator [151].

Also, Bai et al. [136], in their research, proposed an algorithm which was based on multi-feature extraction and gradient boosting decision trees (GBDT)-Adaboost fusion model to identify different types of bovine milk somatic cells. For the identification and classification of bovine milk somatic cells, 392 cell images from four types of cells were initially identified; 65 were being identified as epithelial cells, 112 as lymphoid cells, 81 as macrophages, and 134 as neutrophils. After that, the images were preprocessed using the K-means clustering method. Afterwards, the extracted cell features were entered into the GBDT model for optimization and the optimized features were forced into the AdaBoost classifier for cell recognition. The model with the best recall rate across all cell types and F1-Score was the GBDT-AdaBoost. The model achieved 98.0, 96.8, 97.5 and 97.0% in classification accuracy, accuracy, recall rate, and F-value of comprehensive evaluation index, respectively. In the same study, the classification accuracy values for RF, ET, DTs and LightGBM models were 79.9, 71.1, 67.3 and 77.2%, respectively [136].

Table 9. Milk quality and composition prediction according to the utilized ML method.

ML	Tools	No and type of milk samples	Application	R ²	RMSE	Acc	Se	Sp	Ref.
NN	MIRS	730 b	RCT	0.50	⁽¹⁾ 6.397 min	-	-	-	[88]
			k20	0.36	⁽¹⁾ 2.770 min				
			heat stability	0.45	⁽¹⁾ 5.464 min				
			κ-CN	0.42	⁽¹⁾ 1.095 g/L				
MFFANN	NIRS	385 b	blood metabolites	-	-	-	-	-	[149]
ANN	NIRS	499 b	milk technological properties (CF _p , CY _{curd} , Rec _{protein} etc)	0.45 to 0.71	⁽²⁾ 0.02 % to 0.84 mm	-	-	-	[56]
	FTIR	2701 b	blood metabolites (hematocrit, myeloperoxidase, globulins etc)	0.09 to 0.81	0.03 L/L to 80.59 U/L	-	-	-	[150]
k-NN	sensors	1059 ND	milk quality	-	-	98.58%	-	-	[9]
PLS	FTIR	2701 b	blood metabolites (hematocrit, myeloperoxidase, globulins etc)	0.08 to 0.83	0.03 L/L to 106.37 U/L	-	-	-	[150]
	FTIR	471 b	κ-casein	⁽³⁾ 0.90 _{tr} 0.77 _v	⁽¹⁾ 1.41 g/L	-	-	-	[147]
			BCS	⁽³⁾ 0.95 _{tr} 0.57 _v	⁽¹⁾ 0.35				
			BHB	⁽³⁾ 0.88 _{tr} 0.76 _v	⁽¹⁾ 0.10				
PLS-DA	MIRS	730 b	technological & protein properties of milk	-	-	0.40 – 0.80	0.44	-	[88]
	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.968	0.977	0.962	[148]
LDA	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.968	0.980	0.961	[148]
SVM	MIRS	730 b	technological & protein properties of milk	-	-	0.43 – 0.80	0.44 (overall)	1.00 (overall)	[86]
	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.947	0.962	0.938	[148]
Boosting	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.754	0.587	0.842	[148]
Boosting DT	MIRS	730 b	coagulation	-	-	-	0.50	0.98	[88]

MB-DA	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.964	0.972	0.959	[148]
GBM	NIRS	499 b	milk technological properties (CF _p , CY _{curd} , Rec _{protein} etc)	0.45 to 0.70	⁽²⁾ 0.02% to 0.87 mm	-	-	-	[56]
	FTIR	471 b	κ-casein	⁽⁴⁾ 0.97 _{tr} 0.81 _v	⁽¹⁾ 1.08	-	-	-	[147]
			BCS	⁽⁴⁾ 0.91 _{tr} 0.63 _v	⁽¹⁾ 0.25				
			BHB	⁽⁴⁾ 0.90 _{tr} 0.77 _v	⁽¹⁾ 0.09				
	FTIR	2701 b	blood metabolites (hematocrit, myeloperoxidase, globulins etc)	0.10 to 0.83	0.03 L/L to 75.69 U/L	-	-	-	[150]
XGB	NIRS	499 b	milk technological properties (CF _p , CY _{curd} , Rec _{protein} etc)	0.43 to 0.63	⁽²⁾ 0.02 % to 0.90 mm	-	-	-	[56]
	FTIR	2701 b	blood metabolites (hematocrit, myeloperoxidase, globulins etc)	0.08 to 0.78	0.03 L/L to 80.23 U/L	-	-	-	[150]
RF	MIRS	730 b	αS1-CN, κ-CN	-	-	0.48 0.45	0.44	-	[88]
	FTIR	471 b	κ-casein	⁽³⁾ 0.96 _{tr} 0.80 _v	⁽¹⁾ 1.18	-	-	-	[147]
			BCS	⁽³⁾ 0.95 _{tr} 0.61 _v	⁽¹⁾ 0.26				
			BHB	⁽³⁾ 0.90 _{tr} 0.79 _v	⁽¹⁾ 0.10				
	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.696	0.447	0.827	[148]
DRF	FTIR	2701 b	blood metabolites (hematocrit, myeloperoxidase, globulins etc)	0.09 to 0.79	0.03 L/L to 82.49 U/L	-	-	-	[150]
EN	NIRS	499 b	milk technological properties (CF _p , CY _{curd} , Rec _{protein} etc)	0.46 to 0.71	⁽²⁾ 0.02 % to 0.78 mm	-	-	-	[56]
	FTIR	471 b	κ-casein	⁽³⁾ 0.96 _{tr} 0.79 _v	⁽¹⁾ 1.25	-	-	-	[147]
			BCS	⁽³⁾ 0.92 _{tr} 0.59 _v	⁽¹⁾ 0.27				
			BHB	⁽³⁾ 0.89 _{tr} 0.78 _v	⁽¹⁾ 0.10				
	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.951	0.960	0.946	[148]
	FTIR	2701 b	blood metabolites (hematocrit, myeloperoxidase, globulins etc)	0.12 to 0.87	0.03 L/L to 82.99 U/L	-	-	-	[150]
LASSO	MIRS	730 n	CMS, κ-CN	0.08 0.42	⁽¹⁾ 25.286 mm ⁽¹⁾ 1.095 g/L	-	-	-	[88]

	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.959	0.970	0.953	[148]
PC-LR	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.667	0.117	0.956	[148]
RR	MIRS	730 b	a30,	0.37	12.495 mm	-	-	-	[88]
			β-CN,	0.35	1.759 g/L				
			β-LG A	0.19	1.050 g/L				
	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.880	0.779	0.933	[148]
Stacking Ensemble	NIRS	385 b	blood metabolites	-	-	-	-	-	[149]
	FTIR	2701 b	blood metabolites (hematocrit, myeloperoxidase, globulins etc)	0.13 to 0.87	0.03 L/L to 76.33 U/L	-	-	-	[150]
VarSel-DA	MIRS	4320 b	grass-fed/ non-grass-fed milk classification	-	-	0.890	0.845	0.913	[148]
PLS+ANN	MIRS	6619 b	LF in milk	0.60 _c	130.59 _c mg/L	-	-	-	[151]
				0.55 _{cv}	139.01 _{cv} mg/L				
				0.60 _v	162.17 _v mg/L				
PLSR	MIRS	6619 b	LF in milk	0.53 _c	140.94 _c mg/L	-	-	-	[151]
				0.51 _{cv}	144.31 _{cv} mg/L				
				0.61 _v	163.76 _v mg/L				
PLS+SVM	MIRS	6619 b	LF in milk	0.53 _c	144.32 _c mg/L	-	-	-	[151]
				0.53 _{cv}	144.60 _{cv} mg/L				
				0.63 _v	174.92 _v mg/L				
PLS+ Polynomial SVM	MIRS	6619 b	LF in milk	0.64 _c	125.89 _c mg/L	-	-	-	[151]
				0.56 _{cv}	138.40 _{cv} mg/L				
				0.62 _v	166.75 _v mg/L				

⁽¹⁾RMSEV: root mean square error from the cross-validation data, ⁽²⁾RMSLE: root mean squared logarithmic error, ⁽³⁾r : average of predictive ability c: calibration, cv: cross-validation, v: validation, tr: training, b: bovine, MFFANN: multi-layer feedforward artificial neural network, NIR: Near-Infrared, Acc: accuracy, Sp: specificity, Se: sensitivity, ND: not defined, ML: Machine Learning, NN: Neural Networks, MIRS: mid-infrared spectroscopy, CMS: casein micelle size, RCT: rennet coagulation time, k20: curd-firming time, CN: casein, RMSE: root mean square error, ANN: artificial neural network, FTIR: Fourier transform infrared, CYCurd: weight of fresh curd, a45: curd firmness in millimeters at 45 min from rennet addition, PLS: Partial least squares, k-NN: k-Nearest Neighbors, PLS-DA: Partial least squares discriminant analysis, LDA: Linear Discriminant Analysis, SVM: Support Vector Machines, DT: Decision Tree, MB-DA: model-based discriminant analysis, GBM: gradient boosting machines, XGB: extreme gradient boosting, BCS: body condition score, BHB: blood β-hydroxybutyrate, RF: random forest, DRF: distributed random forest, EN: Elastic Net, CF_p: curd firmness peak, LASSO: least absolute shrinkage and selection operator, PC-LR: principal components linear regression, RR: ridge regression, VarSel-DA: variable-selection discriminant analysis, a30: curd firmness at 30 min, tr: training, PLSR: Partial least squares regression, GBDT: gradient boosting decision trees, LF = lactoferin.

6.2. Fraud Detection and Adulteration Identification

To detect fraud in caprine milk, Teixeira et al. [152], developed multivariate classification models using milk analyses performed by a NIRS system. The study focused on building models to classify authentic and adulterated goat milk samples and to recognize some fraud ingredients like water, urea, bovine milk and whey (classes). A total of 300 authentic caprine milk samples, and 300 adulterated ones were created by adding the studied adulterants to the authentic samples, in 5 concentrations, namely, 1%, 5%, 10%, 15%, and 20%. It was found that multivariate classification models could successfully identify fraud in caprine milk. Indeed, k-NN for a 2-class model (authentic goat milk and adulterated goat milk) achieved 95% to 99% sensitivity and 94% to 96.5% specificity, successfully distinguishing adulterated samples from authentic ones. For a 5-class model (authentic goat milk and four for the types of adulterants added (water, urea, bovine whey, and milk), k-NN achieved sensitivity and specificity, ranging from 76% to 100% depending on the adulterant type. Additionally, softs independent modeling of class analogies (SIMCA) sensitivity and specificity values ranged from 93.0% to 98.9%, for the 2-class model, and for the 5-class model SIMCA achieved slightly less consistent classification performance than PLSDA, with sensitivity and specificity ranging between 90.4% to 100.0%. In general, PLSDA outperformed the other methods, producing the most consistent results, with 100% sensitivity and specificity in distinguishing between authentic and adulterated samples for both the 2-class and the 5-class models [152].

In another study, a prediction model for the detection of cow and buffalo milk adulteration using synchronous front-face fluorescence spectroscopy and PLSR was developed by Ullah et al. [153]. In that study, ten raw cow and buffalo milk samples were collected and 30 distinct mixtures (0-100%) of cow with buffalo milk were prepared. The model using PLSR, achieved a high R^2 value (0.99), with a satisfactory performance for adulteration levels above 20%. The RMSE for CV was 1.16, showing low error for the validation phase, while RMSE for the prediction phase was 6.24 underpinning the poor performance of the model for adulteration levels below 20%. For further quantification of the model's sensitivity, Ullah et al., evaluated three detection limits, with Limit of Blank (LOB), Limit of Detection (LOD), and Limit of Quantification (LOQ) being equal to 9.22%, 18.45%, and 55.9%, respectively [153].

In a recent study, Sowmya and Ponnusamy [154], developed a spectroscopy-based sensor system for IoT applications to detect adulterants in milk. Ultraviolet, visible, and IR spectra were used as spectroscopic methods to enhance the adulterants detection accuracy. By using DTs, NB, LDA, SVM, and NN, the adulteration detection problem was formulated as a classification task. The dataset consisted of 16200 spectral data samples, with 70% being selected for the training, 15% for the validation and 15% for the testing of the models. The accuracy values of the utilized ML models were 92.7%, 91.7%, 90%, 90%, and 88.1%, for NN, DTs, SVM, NB, and LDA, respectively. Since NN outperformed the rest of the ML methods, a genetic algorithm was used to adjust the hyperparameters, increasing the accuracy from 92.7% to 100%. Additionally, two NN models were developed in the same study; a binary model to determine the presence of adulterants, and a multiclass model to categorize samples as pure milk or as milk containing one of the four studied adulterants (ammonium sulfate, sodium salicylate, dextrose, and hydrogen peroxide). After the hyperparameter tuning, the binary classification model produced 100% accurate results, which was verified with a confusion matrix. Furthermore, regarding the multiclass classification problem, which involved five classes (pure milk and four adulterants), the NN model also achieved 100% accuracy. Similar to the binary classification, the confusion matrix and ROC (Receiver Operating Characteristics) curve demonstrated the model's excellent performance in classifying adulterants in milk. In comparison to other methods, the proposed system was found to be superior due to its ability to i) work across multiple spectral wavelengths, ii) achieve an absolute accuracy, iii) detect multiple adulterants, and iv) offer a portable, cost-effective, and rapid solution for real-time milk adulterant detection [154].

Machine learning methods, such as classification and regression trees (CART) and multilayer perceptron (MLP) NN combined with FTIR spectroscopy, were used to detect addition of cheese

whey to milk by Lima et al. [155]. In total, 520 milk samples, adulterated with cheese whey in concentrations ranging from 1% to 30%, were tested and comparatively assessed with 65 control samples. The CART model identified lactose as being the most significant predictor with 100% relative importance. Moreover, the CART model demonstrated a remarkable predictive performance of the addition of cheese whey using compositional features like lactose and protein, achieving 96.2% and 0.994 and 97.2% and 0.980, accuracy and area under the curve (AUC) values for the training and the test sample, respectively. The MLP model, which employed inputs like protein, casein, lactose, SNF, TS, and freezing point, achieved an overall accuracy of 97.8% (97.4% and 97.8% for the training and the testing sample, respectively). Despite the satisfying accuracy in detecting adulterated samples from both approaches, MLP was the model with the best performance, in terms of prediction accuracy and misclassification rate [155].

Wang et al. [156], used FTIR absorption profiles along with ML techniques to determine the amount of heat treatment applied to milk. This study evaluated four ML classifiers, namely, RF, SVM, k-NN, and LDA. Random forest outperformed the other methods, achieving a mean accuracy of 0.92 ± 0.03 , followed by SVM with a mean accuracy of 0.90 ± 0.04 , k-NN with 0.86 ± 0.10 , and LDA with 0.84 ± 0.10 . Similarly, RF demonstrated the highest precision (0.90 ± 0.03), indicating fewer false positives and F1-score (0.90 ± 0.03), indicating an overall better classification performance. Further evidence that RF was the better model derived from the statistically significant ($p < 0.001$) differences on the performance between the models. [156].

Moncayo et al. [48] utilized NN for both qualitative analysis of milk blends and the detection and quantification of melamine in adulterated toddler milk powder, using LIBS. An aggregate of 10 pure milk samples were used (4 bovine, 4 caprine, and 2 ovine), alongside 12 mixtures of these pure samples (9 binary mixtures and 3 ternary mixtures). Neural networks method demonstrated increased accuracy in distinguishing between different types of milk and identifying adulteration in the mixtures; it also exhibited a strong generalization capacity, while avoiding overfitting and successfully classifying 100% of the pure and adulterated milk samples. Samples with melamine concentrations ranging from 1% to 6% were used for the melamine adulteration study, and a calibration curve was created using these samples. The NN model provided exceptional results, with a regression coefficient of 0.999, indicating almost perfect accuracy in predicting the melamine adulterated samples [48].

Table 10. Fraud detection and adulteration identification according to the utilized ML method.

ML	Tools	No and type of milk samples	Application	R ²	RMSE	Se (%)	Sp (%)	Accuracy	Ref.
NN	LIBS	22 b, c, o	melamine in toddler milk powder	0.999	-	-	-	Acc: 100%	[48]
	UV, Vis, IR	ND	adulterants in milk	-	-	-	-	Acc: 100%	[154]
CNN	LIBS	25 r	protein adulteration in milk powder	-	-	-	-	Acc: 97.8%	[55]
PLS-DA	NIRS	600 b, c	fraud in goat milk: water	-	-	100 in all cases	100 in all cases	-	[152]
			urea						
			bovine whey milk authentic						
PLSR	Fluorescence	40 b	adulteration in milk	0.99	⁽¹⁾ 1.16 ⁽²⁾ 6.24	-	-	-	[153]
NB	UV, Vis, IR	ND	adulterants in milk	-	-	-	-	90%	[154]
DT	UV, Vis, IR	ND	adulterants in milk	-	-	-	-	91.7%	[154]
LDA	UV, Vis, IR	ND	adulterants in milk	-	-	-	-	88.1%	[154]
	FTIR	ND	heat treatment to milk	-	-	-	-	0.84	[156]
RF	FTIR	ND	heat treatment to milk	-	-	-	-	0.92	[156]
	LIBS	25 r	protein adulteration in milk powder	-	-			0.886 (train) 0.871 (test)	[55]
k-NN	NIRS	600 b, c	fraud in goat milk: water	-	-	76.0	96.6	-	[152]
			urea			80.0	95.4		
			bovine whey			96.0	100		
			milk			80.0	100		
			authentic			99.0	88.0		
	FTIR	ND	heat treatment to milk	-	-	-	-	0.86	[156]
	LIBS	25 r	protein adulteration in milk powder	-	-	-	-	0.884 (train) 0.867 (test)	[55]

	UV, Vis, IR	ND	adulterants in milk	-	-	-	-	90%	[154]
SVM	LIBS	25 r	protein adulteration in milk powder	-	-	-	-	0.961 (train) 0.938 (test)	[55]
	FTIR	ND	heat treatment to milk	-	-	-	-	0.90	[156]
CART	FTIR	520 b	fraud of cheese whey to milk	-	-	-	-	96.2% (train), 97.2% (test)	[155]
MLP	FTIR	520 b	fraud of cheese whey to milk	-	-	-	-	97.8%	[155]

⁽¹⁾ RMSECV: root mean square error in cross validation, ⁽²⁾ RMSECP: root mean square error in prediction r: retail, c: caprine, o: ovine, b: bovine, ND: not defined, NN: Neural Networks, LIBS: laser-induced breakdown spectroscopy, UV: Ultraviolet, Vis: visible, IR: Infrared, CNN: convolutional neural network, PLS-DA: Partial least squares discriminant analysis, NIRS: Near-Infrared spectroscopy, PLSR: Partial Least Square Regression, RMSE: root mean square error, NB: Naive Bayes, DT: Decision Tree, FTIR: Fourier transform infrared, LDA: Linear Discriminant Analysis, RF: random forest, k-NN: k-Nearest Neighbors, SVM: Support Vector Machines, CART: classification and regression trees, MLP: multilayer perceptron, acc: accuracy, sp: specificity, se: sensitivity.

6.3. Milk source and Origin Classification

In their study, Nanou et al. [47], utilized ML on the spectra obtained by LIBS to identify the species the milk originated from (cow, goat, or sheep). For this reason, 683 lyophilized milk powder samples and 1296 raw liquid milk samples were examined by LIBS. The accompanying spectra were then examined using various ML algorithms to classify milk samples according to the species of origin. The trained models' parameters were adjusted to achieve reliable results when applying the NN algorithm. The key parameters selected for evaluation in a single layer MLP NN included the activation function of the hidden layer and the number of neurons it contains. For the LIBS spectra of liquid milk, the best results were obtained with 500 neurons in the hidden layer and the logistic sigmoid activation function, achieving a training (classification) accuracy of 97.2% ($\pm 0.6\%$) and a testing (predictive) accuracy of 86.3%. For the LIBS spectra of powder milk, the optimal parameters were 300 neurons in the hidden layer with the hyperbolic tangent (tanh) activation function, resulting in a training accuracy of 97.5% ($\pm 0.4\%$) and a testing accuracy of 94.5%. Using only the spectral lines of Mg(II), Ca(II), Ca(I), Na(I), and K(I) for liquid milk, the NN achieved a training accuracy of 98.0% ($\pm 0.1\%$) and a testing accuracy of 87.4%. Similarly, for milk powder, the training accuracy was 98.6% ($\pm 0.1\%$) and the testing accuracy was 92.7% when the same spectral lines were considered. For liquid milk, the best performance for mineral spectra was achieved with 60 neurons and the sigmoid logistic activation function, while for milk powder, the optimum parameters were 90 neurons and the ReLU activation function. In addition, the key results for SVM, LR, and Gradient Boosting (GB) are summarized as follows; the SVM method with a linear kernel, accomplished a training accuracy of approximately 96.6% ($\pm 0.2\%$) concerning liquid milk and 96.2% ($\pm 0.2\%$) for milk powder, with relevant test accuracy of 91.3% and 93.1%, respectively. Support Vector Machine carried out satisfying results when using specific spectral lines (Mg, Ca, Na, K) instead of the entire LIBS spectra, showing its efficiency in classifying milk samples. Similarly, LR showed excellent accuracy in samples of both liquid and powdered milk; it exhibited a training accuracy of 95.2% ($\pm 0.3\%$) and a testing accuracy of 92.8% for liquid milk, while the respective values for milk powder were 95.4% ($\pm 0.2\%$) and 93.5%, respectively. Furthermore, GB demonstrated satisfying performance, achieving a training accuracy of 96.7% ($\pm 0.2\%$) and a testing accuracy of 83.0% for liquid milk samples, and a training accuracy of 97.4% ($\pm 0.3\%$) with a testing accuracy of 91.4% for milk powder samples, surpassing the performance of other models. Finally, comparable classification outcomes were achieved by utilizing the Mg, Ca, Na, and K spectral lines rather than the complete LIBS spectra in the SVM, LR, and GB algorithms [47].

In their study, Amjad et al. [40], focused on the development of a ML-based classification system using Raman spectroscopy to distinguish human, cow, buffalo, and goat milk samples by investigating the spectral differences. A total of 602 milk samples were used in the study, 210 from humans, 152 from cows, 120 from buffaloes and 120 from goats. Principal Component Analysis was applied for the reduction of data dimensionality. Based on the transformed data from PCA, RF was employed to classify the milk samples into one of the four species. Moreover, several performance measures were estimated, including true positive, true negative, false positive, false negative, accuracy, sensitivity, and specificity. The average classification accuracy during the training phase was 94.3%, compared to 94.0% in the testing phase. The overall accuracy on the testing data was approximately 93.6%. The classification performance was the highest for human milk, achieving sensitivity and specificity values of 1.00 and 0.99, respectively. Compared to human milk, cow, buffalo, and goat milk showed some overlapping features, resulting in a higher misclassification rate. Indeed, sensitivity values were 0.95, 0.88 and 0.90 for cow, buffalo, and goat milk, respectively, whereas, the relevant values regarding specificity ranged from 0.96 to 0.99 [40].

The determination of milk origin was also studied by Behkami et al. [157], who utilized ANN to classify the geographical origin of raw cow milk based on spectral data obtained from two instruments: UV-Vis/NIR, with three detectors, and FT-NIR, with a single detector. Principal component analysis and ANN were applied to reduce the dimensionality of spectral data, making it suitable for the ANN classification process. A total of 63 raw cow milk samples, were separated into

training (60%), testing (25%), and validation (15%) sets. The ANNs architecture involved input, hidden, and output layers. Boosting was used to integrate models' performance by combining smaller models into a larger additive model, where 10 boosting models, 2 base models, and learning rates of 0.1 and 1 were the boosting parameters. One hidden layer with ten neurons, one output node, and seven input nodes made up the best model architecture, managing to achieve a learning rate of 1. This model reached 100% classification accuracy and using only two principal components (PCs), with a learning rate of 1, resulted in a 100% classification rate, explaining 95% of the training data and 92% of the validation data variance. Also, the best model with 10 input nodes, one hidden layer of 10 neurons, and one output node achieved a 100% prediction rate. Reducing the input variables (PCs) to four, still gave 100% classification accuracy, especially with a higher learning rate of 1. Additionally, model performance was not adversely affected by the reduction in the number of PCs, indicating that the approach is strong even in the presence of fewer inputs [157].

Table 11. Milk source and regional origin classification according to the utilized ML method.

ML	Tools	No and type of milk samples	Application	Accuracy (%)	Ref.
NN	LIBS	683 lyophilized 1296 liquid b, c, o	animal origin: liquid milk powdered milk Mg, Ca, Na, K	97.2 (train), 86.3 (test) 97.5 (train), 94.5 (test), 98.6 (train), 92.7 (test)	[47]
ANN	UV- Vis/NIR, FT- NIR	63 b	geographic origin of cow milk	100 classification 95 train 92 validation	[157]
SVM	LIBS	683 lyophilized 1296 liquid b, c, o	animal origin: liquid milk powdered milk	96.6 (train), 91.3 (test) 96.2 (train), 93.1 (test)	[47]
GBM	LIBS	683 lyophilized 1296 liquid b, c, o	animal origin: liquid milk powdered milk	96.7 (train), 83.0 (test) 97.4 (train), 91.4 (test)	[47]
RF	Raman	602 b, c, o, h	classify milk (cow, human, buffalo, goat)	93.63	[40]

h: human, c: caprine, o: ovine, b: bovine, NN: Neural Networks, ANN: artificial neural network, LIBS: laser-induced breakdown spectroscopy, UV: Ultraviolet, Vis: visible, NIR: Near-Infrared, SVM: Support Vector Machines, FT-NIR: Fourie Transform Near-Infrared, GBM: gradient boosting machines, RF: random forest.

7. Future Research

Future studies could aim to prove that the systems developed for milk analysis are operational in real-time applications in the dairy industry. One of the biggest challenges for long-term monitoring systems is how to develop flexible systems that can adapt to different sets of conditions (i.e., different conditions, animals, terrain, dirt) which might have a significant effect on the performance of the monitoring systems [105]. Even though a great deal of computational and analytic techniques have been developed for dairy farms, relatively few of them have been field tested.

Even though spectroscopy has long been a research focus, recent developments in hardware and software offer a plethora of opportunities for future research. These advancements, including the wide use of AI and ML techniques, have opened a wide range of applications in milk analysis, while they simultaneously improve the precision and efficiency of spectroscopic methods. With knowledge of the successful results of techniques fusion (ex. NIRS with LIBS [53]), another interesting suggestion is to combine different spectroscopy methods in order to fully exploit their strengths, for instance, to combine NIR, MIR or UV spectroscopy with fluorescence spectroscopy. This initiates a wide range of research opportunities where extraordinary results can be achieved by combining complementary spectroscopy methods with the right ML techniques. Furthermore, since some of the methods (like Raman spectroscopy [104]) haven't been extensively tested in conjunction with ML techniques, there is also room for more research.

Future research should combine data obtained directly from the source (farm, animals etc.) with data analyzed from related datasets to optimize the use of all available information. Additionally, the integration of blockchain technology can ensure data integrity and transparency across the milk supply chain, providing sealed records for milk quality, efficient safety management systems and safety parameters [158,159]. Another emerging tool is the digital twin, which can simulate real-world dairy production environments, allowing for real-time virtual testing of system adaptations and sensor behaviors before physical implementation, addressing the challenges of different conditions, animals, and terrains [159]. Also, there are certain difficulties with dairy productions data management, such as, the volume and complexity of these massive datasets, real-time data processing especially for farms with limitation on internet access, data quality and accuracy based on the conditions of the sensors and data integration which involves data from different systems [117]. Currently, farmers can only invest in multiple systems for maximum benefit, despite manufacturers claims to provide complete solutions. No single sensor system can accomplish everything that could be accomplished by utilizing all of the systems working together [160].

8. Conclusions

It is essential to comprehend the complex composition of milk and measure its key components accurately to optimize dairy production and ensure high product quality. As dairy producers and researchers continue to explore novel technologies and refine existing methods, the integration of sophisticated analytical tools and a deep understanding of milk's biochemical and optical properties will drive improvements in both dairy farming practices and product development. Precision livestock farming tools assist farmers in their decision-making by optimizing the management of daily tasks and herd supervision. This review underlines the importance of the application of innovative technologies in dairy ruminants. The use of spectroscopy-based methods for milk analysis combined with a variety of ML methods for real-time data analysis has proven the optimal approach in numerous cases.

This review was a comparative study of multiple spectroscopy and ML techniques. Different methods might be the best option depending on the application. In general, IR methods are more popular, with wide use in milk analysis. Particularly NIRS gaining popularity due to its speed, reliability, environmental friendliness and applicability in real-time applications. As seen though, in the comparison of the various methods, in most cases NIRS and MIRS have approximately the same predictive accuracy. On the other hand, LIBS has been proven suitable for applications like animal origin identification. Another interesting observation was that the fusion of spectroscopy techniques

that are complementary to each other such as NIRS and LIBS, in some cases results in very successful results. It was established that no specific classifier has the best fit for all problems and no classifier is always better than another one. However, ML techniques have proven to be powerful methods for advancing and understanding of dairy production, giving us, among other benefits, efficient solutions for milk traits, metabolic status, and dairy cow durability, to identify milk origins and detect rigging. Among studies, NN had most of the time, in both regression and classification, the best performance, as much in predicting milk composition, technological properties, and blood metabolites, as in fraud detection and in species identification. Additionally, SVM also excelled in specific applications such as milk quality classification. Concerning the regression-based methods, NN and RR were the ones that demonstrated high accuracy for predicting milk characteristics. Regarding classification tools, SVM excelled, particularly in predicting binary traits. Also, PLSDA showed excellent performance in milk classification, and adulterants classification. Artificial neural networks, especially when combined with dimensionality reduction techniques, proved robust for regional origin classification and for other health metrics of dairy ruminants related to milk analysis. In fact, algorithms like RF, GB, and k-NN also appeared to have competitive results in certain contexts, but in most cases outperformed by SVM, and NN. Taking everything into account, major factors in higher yield or production are better identification and management of animal health problems as well as adherence to medical guidelines. Finally, this review's underlying conclusion is that combining ML techniques with spectroscopy-based methods plays an increasingly important role in real-time milk analysis.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used “Conceptualization, A-A.A., M.P.N., A.I.G. and T.B.; methodology, A-A.A., M.P.N., T.B., N.C., K.D., A.I.G.; investigation, A-A.A. and M.P.N.; writing—original draft preparation, A-A.A. and M.P.N.; writing—review and editing, A.I.G., T.B., N.C. and K.D.; supervision, A.I.G. and T.B., funding acquisition: A.I.G. and T.B. All authors have read and agreed to the published version of the manuscript.” Please turn to the [CRediT taxonomy](#) for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: This research is being implemented within the framework of the National Recovery and Resilience Plan «Greece 2.0» funded by European Union – NextGenerationEU.



Acknowledgments: The authors would like to thank TCB Avgidis Automations S.A. and Telefarm S.A. for their invaluable support in the preparation of this review. The resources, administrative assistance, and access to relevant materials provided by TCB Avgidis Automations S.A. and Telefarm S.A. were essential in enabling the authors to thoroughly analyze and compile the information presented in this manuscript. This support is gratefully acknowledged.

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

References

1. McLeod, A. *World Livestock 2011: Livestock in Food Security*; FAO: Rome, 2011; ISBN 978-92-5-107013-0.
2. Population Prospects 2019: Highlights.
3. Neethirajan, S.; Kemp, B. Digital Livestock Farming. *Sensing and Bio-Sensing Research* **2021**, *32*, 100408, doi:10.1016/j.sbsr.2021.100408.
4. Halachmi, I.; Guarino, M.; Bewley, J.; Pastell, M. Smart Animal Agriculture: Application of Real-Time Sensors to Improve Animal Well-Being and Production. *Annu. Rev. Anim. Biosci.* **2019**, *7*, 403–425, doi:10.1146/annurev-animal-020518-114851.
5. Ochs, D.S.; Wolf, C.A.; Widmar, N.J.O.; Bir, C. Consumer Perceptions of Egg-Laying Hen Housing Systems. *Poultry Science* **2018**, *97*, 3390–3396, doi:10.3382/ps/pey205.
6. Anestis, V.; Bartzanas, T.; Kittas, C. Life Cycle Inventory Analysis for the Milk Produced in a Greek Commercial Dairy Farm - The Link to Precision Livestock Farming. *7th European Conference on Precision Livestock Farming. ECPLF*. **2015**, 670–680.
7. Pereira, P.C. Milk Nutritional Composition and Its Role in Human Health. *Nutrition* **2014**, *30*, 619–627, doi:10.1016/j.nut.2013.10.011.

8. Evangelista, C.; Basiricò, L.; Bernabucci, U. An Overview on the Use of Near Infrared Spectroscopy (NIRS) on Farms for the Management of Dairy Cows. *Agriculture* **2021**, *11*, 296, doi:10.3390/agriculture11040296.
9. Samad, A.; Taze, S.; Kürsad Uçar, M. Enhancing Milk Quality Detection with Machine Learning: A Comparative Analysis of KNN and Distance-Weighted KNN Algorithms. *International Journal of Innovative Science and Research Technology (IJISRT)* **2024**, 2021–2029, doi:10.38124/ijisrt/IJISRT24MAR2123.
10. Tullo, E.; Finzi, A.; Guarino, M. Review: Environmental Impact of Livestock Farming and Precision Livestock Farming as a Mitigation Strategy. *Science of The Total Environment* **2019**, *650*, 2751–2760, doi:10.1016/j.scitotenv.2018.10.018.
11. Helwatkar, A.; Riordan, D.; Walsh, J. Sensor Technology For Animal Health Monitoring. *International Journal on Smart Sensing and Intelligent Systems* **2014**, *7*, 1–6, doi:10.21307/ijssis-2019-057.
12. Neethirajan, S. Recent Advances in Wearable Sensors for Animal Health Management. *Sensing and Bio-Sensing Research* **2017**, *12*, 15–29, doi:10.1016/j.sbsr.2016.11.004.
13. Spectroscopy | Definition, Types, & Facts | Britannica Available online: <https://www.britannica.com/science/spectroscopy> (accessed on 12 September 2024).
14. Pu, Y.-Y.; O'Donnell, C.; Tobin, J.T.; O'Shea, N. Review of Near-Infrared Spectroscopy as a Process Analytical Technology for Real-Time Product Monitoring in Dairy Processing. *International Dairy Journal* **2020**, *103*, 104623, doi:10.1016/j.idairyj.2019.104623.
15. Agelet, L.E.; Hurburgh, C.R. A Tutorial on Near Infrared Spectroscopy and Its Calibration. *Critical Reviews in Analytical Chemistry* **2010**, *40*, 246–260, doi:10.1080/10408347.2010.515468.
16. Herschel, W. Investigation of the Powers of the Prismatic Colours to Heat and Illuminate Objects; With Remarks, That Prove the Different Refrangibility of Radiant Heat. To Which Is Added, an Inquiry into the Method of Viewing the Sun Advantageously, with Telescopes of Large Apertures and High Magnifying Powers. *Philosophical Transactions of the Royal Society of London* **1800**, *90*, 255 - 283.
17. Gastélum-Barrios, A.; Soto-Zarazúa, G.M.; Escamilla-García, A.; Toledano-Ayala, M.; Macías-Bobadilla, G.; Jauregui-Vazquez, D. Optical Methods Based on Ultraviolet, Visible, and Near-Infrared Spectra to Estimate Fat and Protein in Raw Milk: A Review. *Sensors* **2020**, *20*, 3356, doi:10.3390/s20123356.
18. Dispersion (Optics). *Wikipedia* 2024.
19. Ma, W.; Ji, X.; Ding, L.; Yang, S.X.; Guo, K.; Li, Q. Automatic Monitoring Methods for Greenhouse and Hazardous Gases Emitted from Ruminant Production Systems: A Review. *Sensors* **2024**, *24*, 4423, doi:10.3390/s24134423.
20. Fazio, E.; Spadaro, S.; Corsaro, C.; Neri, G.; Leonardi, S.G.; Neri, F.; Lavanya, N.; Sekar, C.; Donato, N.; Neri, G. Metal-Oxide Based Nanomaterials: Synthesis, Characterization and Their Applications in Electrical and Electrochemical Sensors. *Sensors* **2021**, *21*, 2494, doi:10.3390/s21072494.
21. Hulanicki', A.; Geab, S.; Ingman, F. O. L. K. E. Chemical sensors: definitions and classification. *Pure and applied chemistry* **1991**, *63*(9), 1247-1250.
22. Kunes, R.; Bartos, P.; Iwasaka, G.K.; Lang, A.; Hankovec, T.; Smutny, L.; Cerny, P.; Poborska, A.; Smetana, P.; Kriz, P.; et al. In-Line Technologies for the Analysis of Important Milk Parameters during the Milking Process: A Review. *Agriculture* **2021**, *11*, 239, doi:10.3390/agriculture11030239.
23. He, H.; Sun, D.-W.; Pu, H.; Chen, L.; Lin, L. Applications of Raman Spectroscopic Techniques for Quality and Safety Evaluation of Milk: A Review of Recent Developments. *Critical Reviews in Food Science and Nutrition* **2019**, *59*, 770–793, doi:10.1080/10408398.2018.1528436.
24. Pellegrino, L.; Cattaneo, S.; De Noni, I. Nutrition and Health: Effects of Processing on Protein Quality of Milk and Milk Products. *Encyclopedia of dairy sciences Elsevier* **2016**, 1067-1074, ISBN 978-0-08-100596-5.
25. Reference Material for Somatic Cell Counting - European Commission Available online: https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/reference-material-somatic-cell-counting-2020-02-11_en (accessed on 1 October 2024).
26. Gelasakis, A.I.; Mavrogianni, V.S.; Petridis, I.G.; Vasileiou, N.G.C.; Fthenakis, G.C. Mastitis in Sheep – The Last 10 Years and the Future of Research. *Veterinary Microbiology* **2015**, *181*, 136–146, doi:10.1016/j.vetmic.2015.07.009.
27. Melfsen, A.; Hartung, E.; Haeussermann, A. Accuracy of In-Line Milk Composition Analysis with Diffuse Reflectance near-Infrared Spectroscopy. *Journal of Dairy Science* **2012**, *95*, 6465–6476, doi:10.3168/jds.2012-5388.
28. Numthuam, S.; Hongpathong, J.; Charoensook, R.; Rungchang, S. Method Development for the Analysis of Total Bacterial Count in Raw Milk Using Near-infrared Spectroscopy. *Journal of Food Safety* **2017**, *37*, e12335, doi:10.1111/jfs.12335.
29. Nicolaou, N.; Goodacre, R. Rapid and Quantitative Detection of the Microbial Spoilage in Milk Using Fourier Transform Infrared Spectroscopy and Chemometrics. *Analyst* **2008**, *133*, 1424, doi:10.1039/b804439b.
30. Pampoukis, G.; Lytou, A.E.; Argyri, A.A.; Panagou, E.Z.; Nychas, G.-J.E. Recent Advances and Applications of Rapid Microbial Assessment from a Food Safety Perspective. *Sensors* **2022**, *22*, 2800, doi:10.3390/s22072800.

31. Blanco, M.; Villarroya, I. NIR Spectroscopy: A Rapid-Response Analytical Tool. *TrAC Trends in Analytical Chemistry* **2002**, *21*, 240–250, doi:10.1016/S0165-9936(02)00404-1.
32. User:Jhausauer English: *Spectral Lines. Adapted from the Version in Italian.*; 2007;
33. Raman, C.V.; Krishnan, K.S. A new type of secondary radiation. *Nature* **1928**, *121*(3048), 501–502.
34. Vaskova, H.; Buckova, M. Measuring the Lactose Content in Milk. *MATEC Web Conf.* **2016**, *76*, 05011, doi:10.1051/mateconf/20167605011.
35. Mazurek, S.; Szostak, R.; Czaja, T.; Zachwieja, A. Analysis of Milk by FT-Raman Spectroscopy. *Talanta* **2015**, *138*, 285–289, doi:10.1016/j.talanta.2015.03.024.
36. El-Abassy, R.M.; Eravuchira, P.J.; Donfack, P.; Von Der Kammer, B.; Materny, A. Fast Determination of Milk Fat Content Using Raman Spectroscopy. *Vibrational Spectroscopy* **2011**, *56*, 3–8, doi:10.1016/j.vibspec.2010.07.001.
37. Rodrigues Júnior, P.H.; De Sá Oliveira, K.; Almeida, C.E.R.D.; De Oliveira, L.F.C.; Stephani, R.; Pinto, M.D.S.; Carvalho, A.F.D.; Perrone, Í.T. FT-Raman and Chemometric Tools for Rapid Determination of Quality Parameters in Milk Powder: Classification of Samples for the Presence of Lactose and Fraud Detection by Addition of Maltodextrin. *Food Chemistry* **2016**, *196*, 584–588, doi:10.1016/j.foodchem.2015.09.055.
38. Khan, K.M.; Krishna, H.; Majumder, S.K.; Gupta, P.K. Detection of Urea Adulteration in Milk Using Near-Infrared Raman Spectroscopy. *Food Anal. Methods* **2015**, *8*, 93–102, doi:10.1007/s12161-014-9873-z.
39. McGoverin, C.M.; Clark, A.S.S.; Holroyd, S.E.; Gordon, K.C. Raman Spectroscopic Quantification of Milk Powder Constituents. *Analytica Chimica Acta* **2010**, *673*, 26–32, doi:10.1016/j.aca.2010.05.014.
40. Amjad, A.; Ullah, R.; Khan, S.; Bilal, M.; Khan, A. Raman Spectroscopy Based Analysis of Milk Using Random Forest Classification. *Vibrational Spectroscopy* **2018**, *99*, 124–129, doi:10.1016/j.vibspec.2018.09.003.
41. Noll, R. *Laser-Induced Breakdown Spectroscopy: Fundamentals and Applications*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; ISBN 978-3-642-20667-2.
42. What Is LIBS? | SciAps Available online: <https://www.sciaps.com/products/lib/what-is-lib> (accessed on 12 September 2024).
43. Musazzi, S.; Perini, U. *Laser-Induced Breakdown Spectroscopy: Theory and Applications*; Springer, 2014; ISBN 978-3-642-45085-3.
44. Dos Santos Augusto, A.; Barsanelli, P.L.; Pereira, F.M.V.; Pereira-Filho, E.R. Calibration Strategies for the Direct Determination of Ca, K, and Mg in Commercial Samples of Powdered Milk and Solid Dietary Supplements Using Laser-Induced Breakdown Spectroscopy (LIBS). *Food Research International* **2017**, *94*, 72–78, doi:10.1016/j.foodres.2017.01.027.
45. Markiewicz-Keszycka, M.; Cama-Moncunill, X.; Casado-Gavaldà, M.P.; Dixit, Y.; Cama-Moncunill, R.; Cullen, P.J.; Sullivan, C. Laser-Induced Breakdown Spectroscopy (LIBS) for Food Analysis: A Review. *Trends in Food Science & Technology* **2017**, *65*, 80–93, doi:10.1016/j.tifs.2017.05.005.
46. Nanou, E.; Stefan, D.; Couris, S. Milk's Inorganic Content Analysis via Laser Induced Breakdown Spectroscopy. *Food Chemistry* **2023**, *407*, 135169, doi:10.1016/j.foodchem.2022.135169.
47. Nanou, E.; Pliatsika, N.; Stefan, D.; Couris, S. Identification of the Animal Origin of Milk via Laser-Induced Breakdown Spectroscopy. *Food Control* **2023**, *154*, 110007, doi:10.1016/j.foodcont.2023.110007.
48. Moncayo, S.; Manzoor, S.; Rosales, J.D.; Anzano, J.; Caceres, J.O. Qualitative and Quantitative Analysis of Milk for the Detection of Adulteration by Laser Induced Breakdown Spectroscopy (LIBS). *Food Chemistry* **2017**, *232*, 322–328, doi:10.1016/j.foodchem.2017.04.017.
49. Bilge, G.; Sezer, B.; Eseller, K.E.; Berberoglu, H.; Topcu, A.; Boyaci, I.H. Determination of Whey Adulteration in Milk Powder by Using Laser Induced Breakdown Spectroscopy. *Food Chemistry* **2016**, *212*, 183–188, doi:10.1016/j.foodchem.2016.05.169.
50. Abdel-Salam, Z.; Al Sharnoubi, J.; Harith, M.A. Qualitative Evaluation of Maternal Milk and Commercial Infant Formulas via LIBS. *Talanta* **2013**, *115*, 422–426, doi:10.1016/j.talanta.2013.06.003.
51. Abdel-Salam, Z.; El-Saeid, R.; Abdelghany, S.; Abdel-Salam, S.; Radwan, M. Assessment of Milk Quality at Farm Level Using Laser Techniques. *Egypt. J. Chem.* **2022**, *0*, 0–0, doi:10.21608/ejchem.2022.151233.6548.
52. Cama-Moncunill, X.; Markiewicz-Keszycka, M.; Dixit, Y.; Cama-Moncunill, R.; Casado-Gavaldà, M.P.; Cullen, P.J.; Sullivan, C. Feasibility of Laser-Induced Breakdown Spectroscopy (LIBS) as an at-Line Validation Tool for Calcium Determination in Infant Formula. *Food Control* **2017**, *78*, 304–310, doi:10.1016/j.foodcont.2017.03.005.
53. Eum, C.; Jang, D.; Kim, J.; Choi, S.; Cha, K.; Chung, H. Improving the Accuracy of Spectroscopic Identification of Geographical Origins of Agricultural Samples through Cooperative Combination of Near-Infrared and Laser-Induced Breakdown Spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2018**, *149*, 281–287, doi:10.1016/j.sab.2018.09.004.
54. Sezer, B.; Durna, S.; Bilge, G.; Berkkan, A.; Yetisemiyen, A.; Boyaci, I.H. Identification of Milk Fraud Using Laser-Induced Breakdown Spectroscopy (LIBS). *International Dairy Journal* **2018**, *81*, 1–7, doi:10.1016/j.idairyj.2017.12.005.

55. Huang, W.; Guo, L.; Kou, W.; Zhang, D.; Hu, Z.; Chen, F.; Chu, Y.; Cheng, W. Identification of Adulterated Milk Powder Based on Convolutional Neural Network and Laser-Induced Breakdown Spectroscopy. *Microchemical Journal* **2022**, *176*, 107190, doi:10.1016/j.microc.2022.107190.
56. Mota, L.F.M.; Giannuzzi, D.; Bisutti, V.; Pegolo, S.; Trevisi, E.; Schiavon, S.; Gallo, L.; Fineboym, D.; Katz, G.; Cecchinato, A. Real-Time Milk Analysis Integrated with Stacking Ensemble Learning as a Tool for the Daily Prediction of Cheese-Making Traits in Holstein Cattle. *Journal of Dairy Science* **2022**, *105*, 4237–4255, doi:10.3168/jds.2021-21426.
57. Brandão, M. C. M. P.; Carmo, A. P.; Bell, M. J. V.; Anjos, V. C. Characterization of milk by infrared spectroscopy. *Revista do Instituto de Laticínios Cândido Tostes* **2010**, *65*(373), 30–33.
58. Aernouts, B.; Van Beers, R.; Watté, R.; Huybrechts, T.; Lammertyn, J.; Saeys, W. Visible and Near-Infrared Bulk Optical Properties of Raw Milk. *Journal of Dairy Science* **2015**, *98*, 6727–6738, doi:10.3168/jds.2015-9630.
59. Aernouts, B.; Polshin, E.; Lammertyn, J.; Saeys, W. Visible and Near-Infrared Spectroscopic Analysis of Raw Milk for Cow Health Monitoring: Reflectance or Transmittance? *Journal of Dairy Science* **2011**, *94*, 5315–5329, doi:10.3168/jds.2011-4354.
60. Korelidou, V.; Simitzis, P.; Massouras, T.; Gelasakis, A.I. Infrared Thermography as a Diagnostic Tool for the Assessment of Mastitis in Dairy Ruminants. *Animals* **2024**, *14*(18), 2691.
61. Swinehart, D.F. The Beer-Lambert Law. *Journal of chemical education* **1962**, *39*(7), 333.
62. Givens, D.I.; De Boever, J.L.; Deaville, E.R. The Principles, Practices and Some Future Applications of near Infrared Spectroscopy for Predicting the Nutritive Value of Foods for Animals and Humans. *Nutr. Res. Rev.* **1997**, *10*, 83–114, doi:10.1079/NRR19970006.
63. Yakubu, H.G.; Kovacs, Z.; Toth, T.; Bazar, G. The Recent Advances of Near-Infrared Spectroscopy in Dairy Production—a Review. *Critical Reviews in Food Science and Nutrition* **2022**, *62*, 810–831, doi:10.1080/10408398.2020.1829540.
64. 14:00-17:00 ISO 21543:2006 Available online: <https://www.iso.org/standard/40318.html> (accessed on 12 September 2024).
65. 14:00-17:00 ISO 21543:2020 Available online: <https://www.iso.org/standard/77606.html> (accessed on 13 September 2024).
66. Albanell, E.; Caja, G.; Such, X.; Rovai, M.; Salama, A.A.K.; Casals, R. Determination of Fat, Protein, Casein, Total Solids, and Somatic Cell Count in Goat's Milk by Near-Infrared Reflectance Spectroscopy. *Journal of AOAC INTERNATIONAL* **2003**, *86*, 746–752, doi:10.1093/jaoac/86.4.746.
67. Revilla, I.; Escuredo, O.; González-Martín, M.I.; Palacios, C. Fatty Acids and Fat-Soluble Vitamins in Ewe's Milk Predicted by near Infrared Reflectance Spectroscopy. Determination of Seasonality. *Food Chemistry* **2017**, *214*, 468–477, doi:10.1016/j.foodchem.2016.07.078.
68. Holroyd, S.E. The Use of near Infrared Spectroscopy on Milk and Milk Products. *Journal of Near Infrared Spectroscopy* **2013**, *21*, 311–322, doi:10.1255/jnirs.1055.
69. Coppa, M.; Martin, B.; Agabriel, C.; Chassaing, C.; Sibra, C.; Constant, I.; Graulet, B.; Andueza, D. Authentication of Cow Feeding and Geographic Origin on Milk Using Visible and Near-Infrared Spectroscopy. *Journal of Dairy Science* **2012**, *95*, 5544–5551, doi:10.3168/jds.2011-5272.
70. Al-Qadiri, H.M.; Lin, M.; Al-Holy, M.A.; Cavinato, A.G.; Rasco, B.A. Monitoring Quality Loss of Pasteurized Skim Milk Using Visible and Short Wavelength Near-Infrared Spectroscopy and Multivariate Analysis. *Journal of Dairy Science* **2008**, *91*, 950–958, doi:10.3168/jds.2007-0618.
71. Cattaneo, T.M.P.; Cabassi, G.; Profaizer, M.; Giangiacomo, R. Contribution of Light Scattering to near Infrared Absorption in Milk. *Journal of Near Infrared Spectroscopy* **2009**, *17*, 337–343, doi:10.1255/jnirs.867.
72. Tsenkova, R.; Meilina, H.; Kuroki, S.; Burns, D.H. Near Infrared Spectroscopy Using Short Wavelengths and Leave-One-Cow-Out Cross-Validation for Quantification of Somatic Cells in Milk. *Journal of Near Infrared Spectroscopy* **2009**, *17*, 345–351, doi:10.1255/jnirs.868.
73. Coppa, M.; Ferlay, A.; Leroux, C.; Jestin, M.; Chilliard, Y.; Martin, B.; Andueza, D. Prediction of Milk Fatty Acid Composition by near Infrared Reflectance Spectroscopy. *International Dairy Journal* **2010**, *20*, 182–189, doi:10.1016/j.idairyj.2009.11.003.
74. Núñez-Sánchez, N.; Martínez-Marín, A.L.; Polvillo, O.; Fernández-Cabanás, V.M.; Carrizosa, J.; Urrutia, B.; Serradilla, J.M. Near Infrared Spectroscopy (NIRS) for the Determination of the Milk Fat Fatty Acid Profile of Goats. *Food Chemistry* **2016**, *190*, 244–252, doi:10.1016/j.foodchem.2015.05.083.
75. Wu, D.; He, Y.; Feng, S.; Sun, D.-W. Study on Infrared Spectroscopy Technique for Fast Measurement of Protein Content in Milk Powder Based on LS-SVM. *Journal of Food Engineering* **2008**, *84*, 124–131, doi:10.1016/j.jfoodeng.2007.04.031.
76. Soulat, J.; Andueza, D.; Graulet, B.; Girard, C.L.; Labonne, C.; Aït-Kaddour, A.; Martin, B.; Ferlay, A. Comparison of the Potential Abilities of Three Spectroscopy Methods: Near-Infrared, Mid-Infrared, and Molecular Fluorescence, to Predict Carotenoid, Vitamin and Fatty Acid Contents in Cow Milk. *Foods* **2020**, *9*, 592, doi:10.3390/foods9050592.

77. Coppa, M.; Revello-Chion, A.; Giaccone, D.; Ferlay, A.; Tabacco, E.; Borreani, G. Comparison of near and Medium Infrared Spectroscopy to Predict Fatty Acid Composition on Fresh and Thawed Milk. *Food Chemistry* **2014**, *150*, 49–57, doi:10.1016/j.foodchem.2013.10.087.
78. Balabin, R.M.; Smirnov, S.V. Melamine Detection by Mid- and near-Infrared (MIR/NIR) Spectroscopy: A Quick and Sensitive Method for Dairy Products Analysis Including Liquid Milk, Infant Formula, and Milk Powder. *Talanta* **2011**, *85*, 562–568, doi:10.1016/j.talanta.2011.04.026.
79. Henn, R.; Kirchler, C.G.; Grossgut, M.-E.; Huck, C.W. Comparison of Sensitivity to Artificial Spectral Errors and Multivariate LOD in NIR Spectroscopy – Determining the Performance of Miniaturizations on Melamine in Milk Powder. *Talanta* **2017**, *166*, 109–118, doi:10.1016/j.talanta.2017.01.035.
80. Llano Suárez, P.; Soldado, A.; González-Arrojo, A.; Vicente, F.; De La Roza-Delgado, B. Rapid On-Site Monitoring of Fatty Acid Profile in Raw Milk Using a Handheld near Infrared Sensor. *Journal of Food Composition and Analysis* **2018**, *70*, 1–8, doi:10.1016/j.jfca.2018.03.003.
81. Liu, N.; Parra, H.A.; Pustjens, A.; Hettinga, K.; Mongondry, P.; Van Ruth, S.M. Evaluation of Portable Near-Infrared Spectroscopy for Organic Milk Authentication. *Talanta* **2018**, *184*, 128–135, doi:10.1016/j.talanta.2018.02.097.
82. De La Roza-Delgado, B.; Garrido-Varo, A.; Soldado, A.; González Arrojo, A.; Cuevas Valdés, M.; Maroto, F.; Pérez-Marín, D. Matching Portable NIRS Instruments for in Situ Monitoring Indicators of Milk Composition. *Food Control* **2017**, *76*, 74–81, doi:10.1016/j.foodcont.2017.01.004.
83. Diaz-Olivares, J.A.; Adriaens, I.; Stevens, E.; Saeys, W.; Aernouts, B. Online Milk Composition Analysis with an On-Farm near-Infrared Sensor. *Computers and Electronics in Agriculture* **2020**, *178*, 105734, doi:10.1016/j.compag.2020.105734.
84. Kalinin, A.; Krashennnikov, V.; Sadovskiy, S.; Yurova, E. Determining the Composition of Proteins in Milk Using a Portable near Infrared Spectrometer. *Journal of Near Infrared Spectroscopy* **2013**, *21*, 409–415, doi:10.1255/jnirs.1054.
85. Santos, P.M.; Pereira-Filho, E.R.; Rodriguez-Saona, L.E. Rapid Detection and Quantification of Milk Adulteration Using Infrared Microspectroscopy and Chemometrics Analysis. *Food Chemistry* **2013**, *138*, 19–24, doi:10.1016/j.foodchem.2012.10.024.
86. Etzion, Y.; Linker, R.; Cogan, U.; Shmulevich, I. Determination of Protein Concentration in Raw Milk by Mid-Infrared Fourier Transform Infrared/Attenuated Total Reflectance Spectroscopy. *Journal of Dairy Science* **2004**, *87*, 2779–2788, doi:10.3168/jds.S0022-0302(04)73405-0.
87. Dabrowska, A.; David, M.; Freitag, S.; Andrews, A.M.; Strasser, G.; Hinkov, B.; Schwaighofer, A.; Lendl, B. Broadband Laser-Based Mid-Infrared Spectroscopy Employing a Quantum Cascade Detector for Milk Protein Analysis. *Sensors and Actuators B: Chemical* **2022**, *350*, 130873, doi:10.1016/j.snb.2021.130873.
88. Frizzarin, M.; Gormley, I.C.; Berry, D.P.; Murphy, T.B.; Casa, A.; Lynch, A.; McParland, S. Predicting Cow Milk Quality Traits from Routinely Available Milk Spectra Using Statistical Machine Learning Methods. *Journal of Dairy Science* **2021**, *104*, 7438–7447, doi:10.3168/jds.2020-19576.
89. De Marchi, M.; Fagan, C.C.; O'Donnell, C.P.; Cecchinato, A.; Dal Zotto, R.; Cassandro, M.; Penasa, M.; Bittante, G. Prediction of Coagulation Properties, Titratable Acidity, and pH of Bovine Milk Using Mid-Infrared Spectroscopy. *Journal of Dairy Science* **2009**, *92*, 423–432, doi:10.3168/jds.2008-1163.
90. De Marchi, M.; Toffanin, V.; Cassandro, M.; Penasa, M. Invited Review: Mid-Infrared Spectroscopy as Phenotyping Tool for Milk Traits. *Journal of Dairy Science* **2014**, *97*, 1171–1186, doi:10.3168/jds.2013-6799.
91. Ceni, C.; Spina, A.A.; Piras, C.; Oppedisano, F.; Tilocca, B.; Roncada, P.; Britti, D.; Morittu, V.M. Recent Advances in the Determination of Milk Adulterants and Contaminants by Mid-Infrared Spectroscopy. *Foods* **2023**, *12*, 2917, doi:10.3390/foods12152917.
92. Fox, P.F.; Uniacke-Lowe, T.; McSweeney, P.L.H.; O'Mahony, J.A. *Dairy Chemistry and Biochemistry*; Springer International Publishing: Cham, 2015; ISBN 978-3-319-14891-5.
93. Nicolaou, N.; Xu, Y.; Goodacre, R. Fourier Transform Infrared Spectroscopy and Multivariate Analysis for the Detection and Quantification of Different Milk Species. *Journal of Dairy Science* **2010**, *93*, 5651–5660, doi:10.3168/jds.2010-3619.
94. Loudiyi, M.; Temiz, H.T.; Sahar, A.; Haseeb Ahmad, M.; Boukria, O.; Hassoun, A.; Ait-Kaddour, A. Spectroscopic Techniques for Monitoring Changes in the Quality of Milk and Other Dairy Products during Processing and Storage. *Critical Reviews in Food Science and Nutrition* **2022**, *62*, 3063–3087, doi:10.1080/10408398.2020.1862754.
95. Fragkouli, N.; Samartzis, P.C.; Velegrakis, M. Commercial Milk Discrimination by Fat Content and Animal Origin Using Optical Absorption and Fluorescence Spectroscopy. *International Dairy Journal* **2021**, *123*, 105181, doi:10.1016/j.idairyj.2021.105181.
96. Andersen, C.M.; Mortensen, G. Fluorescence Spectroscopy: A Rapid Tool for Analyzing Dairy Products. *J. Agric. Food Chem.* **2008**, *56*, 720–729, doi:10.1021/jf072025o.
97. Shaikh, S.; O'Donnell, C. Applications of Fluorescence Spectroscopy in Dairy Processing: A Review. *Current Opinion in Food Science* **2017**, *17*, 16–24, doi:10.1016/j.cofs.2017.08.004.

98. Barreto, M.C.; Braga, R.G.; Lemos, S.G.; Fragoso, W.D. Determination of Melamine in Milk by Fluorescence Spectroscopy and Second-Order Calibration. *Food Chemistry* **2021**, *364*, 130407, doi:10.1016/j.foodchem.2021.130407.
99. Bogomolov, A.; Dietrich, S.; Boldrini, B.; Kessler, R.W. Quantitative Determination of Fat and Total Protein in Milk Based on Visible Light Scatter. *Food Chemistry* **2012**, *134*, 412–418, doi:10.1016/j.foodchem.2012.02.077.
100. Yang, B.; Guo, W.; Liang, W.; Zhou, Y.; Zhu, X. Design and Evaluation of a Miniature Milk Quality Detection System Based on UV/Vis Spectroscopy. *Journal of Food Composition and Analysis* **2022**, *106*, 104341, doi:10.1016/j.jfca.2021.104341.
101. Karoui, R.; Martin, B.; Dufour, É. Potentiality of Front-Face Fluorescence Spectroscopy to Determine the Geographic Origin of Milks from the Haute-Loire Department (France). *Lait* **2005**, *85*, 223–236, doi:10.1051/lait:2005008.
102. Birlouez-Aragon, I.; Sabat, P.; Gouti, N. A New Method for Discriminating Milk Heat Treatment. *International Dairy Journal* **2002**, *12*, 59–67, doi:10.1016/S0958-6946(01)00131-5.
103. Hougaard, A.B.; Lawaetz, A.J.; Ipsen, R.H. Front Face Fluorescence Spectroscopy and Multi-Way Data Analysis for Characterization of Milk Pasteurized Using Instant Infusion. *LWT - Food Science and Technology* **2013**, *53*, 331–337, doi:10.1016/j.lwt.2013.01.010.
104. Domingo, E.; Tirelli, A.A.; Nunes, C.A.; Guerreiro, M.C.; Pinto, S.M. Melamine Detection in Milk Using Vibrational Spectroscopy and Chemometrics Analysis: A Review. *Food Research International* **2014**, *60*, 131–139, doi:10.1016/j.foodres.2013.11.006.
105. Vázquez-Diosdado, J.A.; Paul, V.; Ellis, K.A.; Coates, D.; Loomba, R.; Kaler, J. A Combined Offline and Online Algorithm for Real-Time and Long-Term Classification of Sheep Behaviour: Novel Approach for Precision Livestock Farming. *Sensors* **2019**, *19*, 3201, doi:10.3390/s19143201.
106. Kaplan, A.; Haenlein, M. Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence. *Business Horizons* **2019**, *62*, 15–25, doi:10.1016/j.bushor.2018.08.004.
107. Niloofar, P.; Francis, D.P.; Lazarova-Molnar, S.; Vulpe, A.; Vochin, M.-C.; Suciu, G.; Balanescu, M.; Anestis, V.; Bartzanas, T. Data-Driven Decision Support in Livestock Farming for Improved Animal Health, Welfare and Greenhouse Gas Emissions: Overview and Challenges. *Computers and Electronics in Agriculture* **2021**, *190*, 106406, doi:10.1016/j.compag.2021.106406.
108. Norton, T.; Berckmans, D. Developing Precision Livestock Farming Tools for Precision Dairy Farming. *Animal Frontiers* **2017**, *7*, 18–23, doi:10.2527/af.2017.0104.
109. VanderWaal, K.; Morrison, R.B.; Neuhauser, C.; Vilalta, C.; Perez, A.M. Translating Big Data into Smart Data for Veterinary Epidemiology. *Front. Vet. Sci.* **2017**, *4*, 110, doi:10.3389/fvets.2017.00110.
110. Benjamin, M.; Yik, S. Precision Livestock Farming in Swine Welfare: A Review for Swine Practitioners. *Animals* **2019**, *9*, 133, doi:10.3390/ani9040133.
111. Morota, G.; Ventura, R.V.; Silva, F.F.; Koyama, M.; Fernando, S.C. Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *Journal of animal science* **2018**, *96*(4), 1540-1550, doi:10.1093/jas/sky014.
112. Southwest Jiaotong University, China; Muhammad, I.; Yan, Z.; Southwest Jiaotong University, China SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY. *ICTACT Journal on Soft Computing* **2015**, *05*, 946–952, doi:10.21917/ijsc.2015.0133.
113. Soofi, A.A.; & Awan, A. Classification techniques in machine learning: applications and issues. *Journal of Basic & Applied Sciences* **2017**, *13*, 459-465.112.
114. Montesinos López, O.A.; Montesinos López, A.; Crossa, J. *Multivariate Statistical Machine Learning Methods for Genomic Prediction*; Springer International Publishing: Cham, 2022; ISBN 978-3-030-89009-4.
115. Nasir, N.; Kansal, A.; Alshaltone, O.; Barneih, F.; Sameer, M.; Shanableh, A.; Al-Shamma'a, A. Water Quality Classification Using Machine Learning Algorithms. *Journal of Water Process Engineering* **2022**, *48*, 102920, doi:10.1016/j.jwpe.2022.102920.
116. Akalin, A. 5.13 Logistic Regression and Regularization | *Computational Genomics with R*;
117. García, R.; Aguilar, J.; Toro, M.; Pinto, A.; Rodríguez, P. A Systematic Literature Review on the Use of Machine Learning in Precision Livestock Farming. *Computers and Electronics in Agriculture* **2020**, *179*, 105826, doi:10.1016/j.compag.2020.105826.
118. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; Taylor & Francis, 1984; ISBN 978-0-412-04841-8.
119. Mu, F.; Gu, Y.; Zhang, J.; Zhang, L. Milk Source Identification and Milk Quality Estimation Using an Electronic Nose and Machine Learning Techniques. *Sensors* **2020**, *20*, 4238, doi:10.3390/s20154238.
120. Montesinos López, O.A.; Montesinos López, A.; Crossa, J. *Multivariate Statistical Machine Learning Methods for Genomic Prediction*; Springer International Publishing: Cham, 2022; ISBN 978-3-030-89009-4.
121. Sun, S.; Huang, R. An Adaptive K-Nearest Neighbor Algorithm. In *2010 seventh international conference on fuzzy systems and knowledge discovery*, IEEE: Yantai, China, August 2010, *1*, 91–94.

122. Japkowicz, N. Learning from imbalanced data sets: a comparison of various strategies. In *AAAI workshop on learning from imbalanced data sets* **2000**, 68, 10-15.
123. Tan, S. Neighbor-Weighted K-Nearest Neighbor for Unbalanced Text Corpus. *Expert Systems with Applications* **2005**, 28, 667–671, doi:10.1016/j.eswa.2004.12.023.
124. Zeng, Y.; Yang, Y.; Zhao, L. Pseudo Nearest Neighbor Rule for Pattern Classification. *Expert Systems with Applications* **2009**, 36, 3587–3595, doi:10.1016/j.eswa.2008.02.003.
125. Shinde, T.A.; Prasad, J.R. IoT based animal health monitoring with naive Bayes classification. *IJETT* **2017**, 1(2).
126. Rish, I. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* **2001**, 3(22), 41-46.
127. Rong, S.; Bao-wen, Z. The Research of Regression Model in Machine Learning Field. *MATEC Web Conf.* **2018**, 176, 01033, doi:10.1051/mateconf/201817601033.
128. Sharma, A.; Paliwal, K.K. Linear Discriminant Analysis for the Small Sample Size Problem: An Overview. *Int. J. Mach. Learn. & Cyber.* **2015**, 6, 443–454, doi:10.1007/s13042-013-0226-9.
129. Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassanien, A.E. Linear Discriminant Analysis: A Detailed Tutorial. *AIC* **2017**, 30, 169–190, doi:10.3233/AIC-170729.
130. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*; 6th ed.; Pearson Prentice Hall: Upper Saddle River, N.J, 2007; ISBN 978-0-13-187715-3.
131. Ferreira, A.J.; Figueiredo, M.A.T. Boosting Algorithms: A Review of Methods, Theory, and Applications. In *Ensemble Machine Learning*; Zhang, C., Ma, Y., Eds.; Springer New York: New York, NY, 2012; pp. 35–85 ISBN 978-1-4419-9325-0.
132. Schapire, R.E. The Boosting Approach to Machine Learning: An Overview. In *Nonlinear Estimation and Classification*; Denison, D.D., Hansen, M.H., Holmes, C.C., Mallick, B., Yu, B., Eds.; Lecture Notes in Statistics; Springer New York: New York, NY, 2003; Vol. 171, pp. 149–171 ISBN 978-0-387-95471-4.
133. Ganaie, M.A.; Hu, M.; Malik, A.K.; Tanveer, M.; Suganthan, P.N. Ensemble Deep Learning: A Review. *Engineering Applications of Artificial Intelligence* **2022**, 115, 105151, doi:10.1016/j.engappai.2022.105151.
134. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences* **1997**, 55, 119–139, doi:10.1006/jcss.1997.1504.
135. Pence, I.; Kumaş, K.; Cemel, M.S.; Akyüz, A. Future Prediction of Biogas Potential and CH₄ Emission with Boosting Algorithms: The Case of Cattle, Small Ruminant, and Poultry Manure from Turkey. *Environ Sci Pollut Res* **2024**, 31, 24461–24479, doi:10.1007/s11356-024-32666-7.
136. Bai, J.; Xue, H.; Jiang, X.; Zhou, Y. Recognition of Bovine Milk Somatic Cells Based on Multi-Feature Extraction and a GBDT-AdaBoost Fusion Model. *MBE* **2022**, 19, 5850–5866, doi:10.3934/mbe.2022274.
137. Wang, F.; Li, Z.; He, F.; Wang, R.; Yu, W.; Nie, F. Feature Learning Viewpoint of Adaboost and a New Algorithm. *IEEE Access* **2019**, 7, 149890–149899, doi:10.1109/ACCESS.2019.2947359.
138. Sun, Y.; Kamel, M.; Wang, Y. Boosting for Learning Multiple Classes with Imbalanced Class Distribution. In *Proceedings of the Sixth International Conference on Data Mining (ICDM) 2006*, IEEE: Hong Kong, China, December 2006, 592–602.
139. Çelik, A. Using Machine Learning Algorithms to Detect Milk Quality. *Eurasian Journal of Food Science and Technology* **2022**, 6(2), 76-87.
140. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* **2001**, 29, 1189–1232.
141. Otchere, D.A.; Ganat, T.O.A.; Ojoro, J.O.; Tackie-Otoo, B.N.; Taki, M.Y. Application of Gradient Boosting Regression Model for the Evaluation of Feature Selection Techniques in Improving Reservoir Characterisation Predictions. *Journal of Petroleum Science and Engineering* **2022**, 208, 109244, doi:10.1016/j.petrol.2021.109244.
142. Pirouz, D.M. An Overview of Partial Least Squares. *SSRN Journal* **2006**, doi:10.2139/ssrn.1631359.
143. Garthwaite, P.H. An Interpretation of Partial Least Squares. *Journal of the American Statistical Association* **1994**, 89, 122–127, doi:10.1080/01621459.1994.10476452.
144. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems* **2001**, 58, 109–130, doi:10.1016/S0169-7439(01)00155-1.
145. Cheng, J.-H.; Sun, D.-W. Partial Least Squares Regression (PLSR) Applied to NIR and HSI Spectral Data Modeling to Predict Chemical Properties of Fish Muscle. *Food Eng Rev* **2017**, 9, 36–49, doi:10.1007/s12393-016-9147-1.
146. Meisel, S.; Stöckel, S.; Elschner, M.; Melzer, F.; Rösch, P.; Popp, J. Raman Spectroscopy as a Potential Tool for Detection of *Brucella* Spp. in Milk. *Appl Environ Microbiol* **2012**, 78, 5575–5583, doi:10.1128/AEM.00637-12.
147. Mota, L.F.M.; Pegolo, S.; Baba, T.; Peñagaricano, F.; Morota, G.; Bittante, G.; Cecchinato, A. Evaluating the Performance of Machine Learning Methods and Variable Selection Methods for Predicting Difficult-to-Measure Traits in Holstein Dairy Cattle Using Milk Infrared Spectral Data. *Journal of Dairy Science* **2021**, 104, 8107–8121, doi:10.3168/jds.2020-19861.

148. Frizzarin, M.; O'Callaghan, T.F.; Murphy, T.B.; Hennessy, D.; Casa, A. Application of Machine-Learning Methods to Milk Mid-Infrared Spectra for Discrimination of Cow Milk from Pasture or Total Mixed Ration Diets. *Journal of Dairy Science* **2021**, *104*, 12394–12402, doi:10.3168/jds.2021-20812.
149. Giannuzzi, D.; Mota, L.F.M.; Pegolo, S.; Gallo, L.; Schiavon, S.; Tagliapietra, F.; Katz, G.; Fainboym, D.; Minuti, A.; Trevisi, E.; et al. In-Line near-Infrared Analysis of Milk Coupled with Machine Learning Methods for the Daily Prediction of Blood Metabolic Profile in Dairy Cattle. *Sci Rep* **2022**, *12*, 8058, doi:10.1038/s41598-022-11799-0.
150. Giannuzzi, D.; Mota, L.F.M.; Pegolo, S.; Tagliapietra, F.; Schiavon, S.; Gallo, L.; Marsan, P.A.; Trevisi, E.; Cecchinato, A. Prediction of Detailed Blood Metabolic Profile Using Milk Infrared Spectra and Machine Learning Methods in Dairy Cattle. *Journal of Dairy Science* **2023**, *106*, 3321–3344, doi:10.3168/jds.2022-22454.
151. Soyeurt, H.; Grelet, C.; McParland, S.; Calmels, M.; Coffey, M.; Tedde, A.; Delhez, P.; Dehareng, F.; Gengler, N. A Comparison of 4 Different Machine Learning Algorithms to Predict Lactoferrin Content in Bovine Milk from Mid-Infrared Spectra. *Journal of Dairy Science* **2020**, *103*, 11585–11596, doi:10.3168/jds.2020-18870.
152. Teixeira, J.L.D.P.; Caramês, E.T.D.S.; Baptista, D.P.; Gigante, M.L.; Pallone, J.A.L. Vibrational Spectroscopy and Chemometrics Tools for Authenticity and Improvement the Safety Control in Goat Milk. *Food Control* **2020**, *112*, 107105, doi:10.1016/j.foodcont.2020.107105.
153. Ullah, R.; Khan, S.; Ali, H.; Bilal, M. Potentiality of Using Front Face Fluorescence Spectroscopy for Quantitative Analysis of Cow Milk Adulteration in Buffalo Milk. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2020**, *225*, 117518, doi:10.1016/j.saa.2019.117518.
154. Sowmya, N.; Ponnusamy, V. Development of Spectroscopic Sensor System for an IoT Application of Adulteration Identification on Milk Using Machine Learning. *IEEE Access* **2021**, *9*, 53979–53995, doi:10.1109/ACCESS.2021.3070558.
155. Lima, J.S.; Ribeiro, D.C.S.Z.; Neto, H.A.; Campos, S.V.A.; Leite, M.O.; Fortini, M.E.D.R.; De Carvalho, B.P.M.; Almeida, M.V.O.; Fonseca, L.M. A Machine Learning Proposal Method to Detect Milk Tainted with Cheese Whey. *Journal of Dairy Science* **2022**, *105*, 9496–9508, doi:10.3168/jds.2021-21380.
156. Wang, Y.-T.; Ren, H.-B.; Liang, W.-Y.; Jin, X.; Yuan, Q.; Liu, Z.-R.; Chen, D.-M.; Zhang, Y.-H. A Novel Approach to Temperature-Dependent Thermal Processing Authentication for Milk by Infrared Spectroscopy Coupled with Machine Learning. *Journal of Food Engineering* **2021**, *311*, 110740, doi:10.1016/j.jfoodeng.2021.110740.
157. Behkami, S.; Zain, S.M.; Gholami, M.; Khir, M.F.A. Classification of Cow Milk Using Artificial Neural Network Developed from the Spectral Data of Single- and Three-Detector Spectrophotometers. *Food Chemistry* **2019**, *294*, 309–315, doi:10.1016/j.foodchem.2019.05.060.
158. Nychas, G.-J.E.; Panagou, E.Z.; Mohareb, F. Novel Approaches for Food Safety Management and Communication. *Current Opinion in Food Science* **2016**, *12*, 13–20, doi:10.1016/j.cofs.2016.06.005.
159. Nychas, G.-J.; Sims, E.; Tsakanikas, P.; Mohareb, F. Data Science in the Food Industry. *Annu. Rev. Biomed. Data Sci.* **2021**, *4*, 341–367, doi:10.1146/annurev-biodatasci-020221-123602.
160. Knight, C.H. Review: Sensor Techniques in Ruminants: More than Fitness Trackers. *Animal* **2020**, *14*, s187–s195, doi:10.1017/S1751731119003276.

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.