SEBASTIEN - A smarter livestock breeding through advanced services tailoring innovative and multi-source data to users' needs

A. D'Anca[a,*], M. Milanesi[b,d], G. Trotta[c], P. Ajmone Marsan[b], R. Negrini[b], A. Reder[a], R. Valentini[d,f], E. Rubinacci[e], C. Caroli[c], F. Renzi[f], S. Noce[a], G. Tramonte[a], V. Scardigno[a], F. Gabbianelli[d], V. Aloisi[g], D. Pietrucci[d], D. Del Buono[d], M. Santini[a], L. Conte[a], A. Aloisio[a], C. Dellacasa[c], G. Scipione[c], M. Passamonti[b], A. Rullo[e], G. Chillemi[h,i], V. Piccolo[e], M. Barbato[j], P. Nassisi[a]

[a] *CMCC Foundation - Euro-Mediterranean Center on Climate Change, via Marco Biagi 5, 73100, Lecce, Italy*
[b] *Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122, Piacenza, Italy*
[c] *CINECA National SupercomputingCenter, via Magnanelli 6/3, 40033, Casalecchio di Reno (BO), Italy*
[d] *University of Tuscia, via S.M. in Gradi 4, 01100, Viterbo, Italy*
[e] *Neatec Spa, via Antiniana 2, 80078, Pozzuoli (NA), Italy*
[f] *Nature 4.0 SB SRL, Via della Chimica 7, 01100, Viterbo, Italy*
[g] *Department of Engineering for Innovation - University of Salento, Centro Ecotekne Pal. O - S.P. 6, 73047, Monteroni (LE), Italy*
[h] *National Institute for Infectious Diseases "Lazzaro Spallanzani" – IRCCS, Via Portuense 292, 00149, Roma, Italy*
[i] *University of Rome "Tor Vergata", Via Cracovia 50, 00133, Roma, Italy*
[j] *Department of Veterinary Sciences, University of Messina, Piazza Pugliatti 1, 98122, Messina, Italy*

Abstract

This study presents SEBASTIEN, a data-driven Decision Support System (DSS) designed to support smart livestock management by combining satellite observations, IoT sensor data, climate reanalysis and projections within a unified and scalable data platform. The system integrates multi-source data streams into a Data Lake architecture and implements Machine Learning and statistical models, including Gradient Boosting Machines and linear mixed models, developed through an AutoML workflow. SEBASTIEN delivers four main operational services: (i) short- and long-term prediction of the Temperature–Humidity Index (THI) for animal welfare assessment; (ii) estimation of milk yield and quality variations under heat stress; (iii) pasture

*Corresponding author. E-mail address: alessandro.danca@cmcc.it

biomass evaluation using satellite data; and (iv) disease risk mapping based on climatic and environmental drivers. The models are trained on large-scale datasets, suggesting robustness and applicability across real farming conditions. Predictive performance indicates high accuracy (e.g., THI prediction RMSE = 2.59, R^2 = 0.95), supporting reliable decision-making. Outputs are provided through interactive dashboards, geospatial maps, and interoperable APIs, enabling both farm-level management and regional-scale monitoring. The DSS supports key applications such as early warning of heat stress, optimization of feeding and grazing strategies, breed selection under climate scenarios, and proactive disease risk mitigation. These results indicate that SEBASTIEN represents a promising operational DSS for enhancing livestock resilience, improving animal welfare, and supporting climate adaptation strategies through integrated, data-driven decision-making.

Keywords: Smart Farming, Machine Learning, IoT, Decision Support System, Livestock, Sustainability, Predictive Analytics

1. Introduction

In the European Union (EU), livestock production is a significant economic sector, as animal products represent 45% of the overall value of agricultural production in the EU, amounting to EUR 168 billion per year [1].

European farms are generally managed by a limited workforce, often requiring increased efficiency and optimized resource management. The livestock sector plays a key economic role, supporting rural livelihoods and employment (almost 30 million people in the EU [2]), while promoting a more efficient circular bio-economy [3].

In Italy, the activities related to the livestock sector are concentrated in the North of the country. Compared with the total agricultural production, the value of the livestock sector is around 36% [4], with meat and dairy products that cover over 60% and 30% respectively (other products are mainly eggs, honey and non-food products) [5]. Considering the species, cattle cover 45% of the production, 48% considering also small ruminants (sheep and goats).

In the EU, out of 161 million hectares of agricultural land, approximately 110 million hectares, are used for animal production ($\approx$ 68%)[6] (71 million hectares of grassland and 39 million hectares of cereals and oilseeds). Beyond land use, agriculture is responsible for more than 10% of the EU's greenhouse gas (GHG) emissions, and 60%-70% of these are produced by the livestock sector[7] (especially methane (CH_4) and nitrous oxide (N_2O)). In Italy, 73% of the total CO_2eq release from agriculture is covered by direct and indirect components of livestock emissions (i.e., enteric fermentation, manure management, manure applied and left on soils)[8].

In the last 50 years the effort of the livestock industry has been mainly directed towards increasing productivity, by improving genetics, modifying the farming environment (e.g., stable structures), and improving nutritional management. In the last decade increasing attention has been paid on production efficiency and animal functionality and welfare. Only very recently stress resilience has been included in selection objectives and not in all species and breeds. The net result is that little has been done to improve stress resistance. While this approach has increased the productivity of livestock, it has also increased animal sensitivity (reduced thermal plasticity) to anomalous (generally warmer) environmental conditions[9].

The processes used by domestic animals to respond to changes in their environment can be classified into three categories: acclimatization, acclimation, and adaptation to external stress[10]. Whilst crucial for survival these processes adversely affect productivity and profitability of farming systems. For instance, cattle subjected to heat stress, respond by reducing rumination time and feed intake, while increasing basal metabolic energy requirements. The reduced ingestion due to heat stress has been quantified to approximately 35% of the lower milk synthesis [11].

In addition, high temperatures combined with excessive or low humidity increase the risks for animals, exacerbating the perceived temperature or drought conditions. Moreover, extreme cold, extraordinary wind conditions, and altered radiation regimes are also harmful for both animals and forages[12, 13, 14, 15, 16].

The environmental changes can impact on[17]:

1. animals' health, growth, and reproduction;
2. diseases occurrence and their epidemiology;
3. feed availability due to the production and quality of cereals, pastures, and fodder crops.

The services developed in this work are designed to explicitly address these three domains: animal welfare is monitored through the Temperature–Humidity Index (THI) and related indicators; disease occurrence is assessed through predictive models of livestock health risks; and feed availability is evaluated through pasture biomass estimation. These integrated services aim to support improved farm management under changing environmental conditions.

In this context, finding new solutions to improve farm management while considering the ongoing climate change is necessary in order to:

- promote mitigation and adaptation actions to reduce or cancel the possible loss of livestock production;
- improve animal well-being and reduce the impact and spread of diseases;
- promptly react to situations of stress or discomfort;
- have a medium-long term forecast available on the evolution of environmental conditions and their impact on different breeds of livestock.

Several Decision Support Systems (DSS) platforms have been proposed in the literature. However, they are often characterized by fragmented implementations, limited data integration, and a focus on specific applications rather than holistic farm management [18], [19]. In particular, existing DSS platforms typically lack the ability to integrate heterogeneous data streams, such as real-time IoT data combined with environmental and climate information, and often do not provide unified multi-domain services [18]. Furthermore, many existing systems are developed as stand-alone tools, with limited interoperability, reduced scalability, and insufficient support for both short-term operational decisions and long-term climate adaptation strategies. In this context, SEBASTIEN distinguishes itself by combining (i) a unified multi-source Data Lake architecture integrating IoT, satellite, and climate data, (ii) the use of very high-resolution climate datasets and projections, and (iii) a comprehensive suite of integrated services for animal welfare, production, pasture, and disease risk within a single operational platform.

The main objective of this work is to develop and demonstrate an integrated DSS for livestock farming that addresses the limitations of existing DSS platforms in terms of data fragmentation, limited interoperability, and lack of multi-domain analysis.

Specifically, this study aims to: (i) integrate heterogeneous data sources into a unified and scalable architecture; (ii) develop predictive models to support decision-making across multiple domains, including animal welfare, production, pasture management, and disease risk; and (iii) provide operational tools for both short-term management and long-term climate adaptation.

With the aim of adopting the best solutions for healthy and profitable breeding, the platform has been developed as an advanced decision support system for smart livestock management, based on the integration of climatic, environmental, satellite, IoT, and production data. Leveraging Machine Learning and data-driven analytics, the system provides services for monitoring animal welfare through the Temperature–Humidity Index (THI), predicting potential losses in milk production and quality under heat-stress conditions, assessing pasture biomass, and estimating livestock disease risk through predictive models. By combining satellite observations, sensor data, climate projections, and advanced modelling, the platform aims to enhance the resilience of livestock systems, optimize resource use, and mitigate the impacts of climate change on European farms.

The remainder of the paper is organized as follows: Section 2 presents the state of the art on the use of IoT technologies, sensor systems, and Machine Learning in livestock farming. Section 3 introduces the platform solution, detailing its architecture and its role as a decision support system for smarter livestock breeding. Section 4 illustrates the advanced services developed for intelligent breeding and data-driven farm management. Finally, Section 5 concludes the work and outlines future research directions.

2. The Role of IoT, Sensors, and Machine Learning in Livestock Farming

The adoption of software technologies, IoT, and Machine Learning in the livestock sector has accelerated in recent years, driven by the increasing need to monitor animal welfare, optimize resource use, and improve the resilience of farming systems to climate variability. These technologies offer significant potential to improve and simplify the management of livestock and its production, revolutionizing the sector by enhancing productivity, animal welfare, and sustainability. Smart farming technologies allow farmers, for example, to monitor livestock production and welfare conditions in real-time [20], make decisions using data-driven approaches [21], and generally optimize all the fundamental operations necessary to meet the growing global demand for animal products [22].

Various software and frameworks are available for data organization and analysis. Data analysis and fusion techniques allow for the integration of data from different sources, such as IoT sources (e.g., wearable sensors), satellite data (e.g., Sentinel data), climate and weather data (e.g., evaluating environmental conditions), etc. These platforms facilitate real-time livestock monitoring, predictive analytics, and efficient data storage and maintenance.

In the livestock sector, IoT-based wearable sensors have been increasingly adopted to enable continuous, real-time monitoring of individual animal health and welfare [23, 20, 24]. Sensors can be placed on animals or ingested (e.g., ruminal bolus) to monitor vital parameters such as temperature, heart rate, and movement patterns, allowing prompt intervention in case of abnormal conditions. This approach directly underpins the IoT component of SEBASTIEN (Section 3.5), where animal collars and environmental sensors collect real-time data on movement, heart rate, ambient temperature, relative humidity, GNSS position, and gas concentrations to support continuous welfare monitoring. Applications of ML algorithms further support predictive modelling, such as forecasting trends in animal productivity to improve resource allocation [23].

HerdDogg, for example, represents a smart-tagging solution that tracks the location, health, and behavior of animals. It provides farmers with alerts when modifications in normal conditions are detected[25]. Such a software improves animal welfare, optimizes feed efficiency, reduces costs, and increases productivity.

IoT-based solutions are crucial for implementing smart farming architectures where interconnected devices provide real-time data to optimize livestock management. For instance, IoT technologies automate routine tasks, such as milking or feeding, or control environmental conditions in stables (e.g., temperature or ventilation adjustments to improve animal comfort). IoT sensors can also localize and track animals in stables or pastures. Furthermore, combining IoT technologies with cloud computing allows for the storage and analysis of large datasets, offering insights into animal health, growth patterns, or optimal breeding times[21, 26].

Despite these advances, existing DSS in livestock farming are often characterized by limited interoperability among heterogeneous data sources, lack of standardized evaluation frameworks, and barriers to large-scale adoption in real farm conditions [18],[19]. In addition, many current systems operate as stand-alone solutions and provide limited explainability of model outputs, which can reduce user trust and hinder decision-making processes [27]. These limitations highlight the need for more integrated, transparent, and operational DSS architectures capable of combining multi-source data and delivering actionable insights to end users.

3. A Decision Support System for Smarter Livestock Breeding

The developed platform, enables seamless integration and coordination of diverse data sources. From Copernicus Services and DIASs to OpenData Portals for livestock and environmental information, as well as IoT systems for animal welfare monitoring, this platform consolidates diverse datasets and formats.

The design and specifications of the Data Platform are meticulously crafted to streamline access and processing of vast data volumes from heterogeneous origins. It addresses challenges related to data access and interoperability without requiring familiarity with the structures, formats, or backend systems of the integrated data sources. Key functional requirements include:

- disentangling information from its sources,
- furnishing generic data structures for analysis,
- preventing duplication of datasets from external sources,
- provisioning a unified Application Programming Interface (API) for accessing, processing, and storing datasets from diverse origins and multi-thematic portals.

The platform is deployed on an on-premises cloud cluster, ensuring scalability and near real-time data processing. All services are containerized using Docker and orchestrated via Kubernetes, enabling reproducible and portable deployments. The software components employed are continuously updated to the latest available stable release to ensure the highest level of security

The core of the platform is the Data Lake that stores, organizes and gives access to all the input and output data. The principal components of the Data Lake include:

1. **Backend:** Manages data requests from clients (Services and Data Portal).
2. **Catalog:** Gathers information concerning connections to data sources and access methods. This includes dataset details, metadata (e.g., dataset identifier), and parameters for data retrieval. Metadata stored in the Catalog are refreshed via event-driven triggers (e.g., a new dataset has been ingested).

3. **Connectors:** A set of custom adapters utilized by the Backend to execute retrieval queries accurately from the appropriate data source when requested.
4. **Data Storage:** Stores developed products like indicators and real-time data from IoT sensors. It also caches data retrieved from external sources.

Datasets are replaced with the latest version when available; to enforce the no-duplication requirement, the platform implements a cache validation mechanism based on timestamp comparison when needed (e.g., for Sentinel L2 data). Before caching any externally retrieved dataset, the Backend verifies whether an identical entry already exists in the Data Storage. If a match is found, the existing cached copy is reused; otherwise, the new dataset is stored and the old version is archived or discarded according to a configurable retention policy (i.e., 30-day rolling window).

The platform handles heterogeneous data from multiple sources, including IoT sensors collecting sensitive farm-level information such as animal GNSS positions, heart rate, and movement patterns. Farm-level data collected via IoT sensors are owned by the respective farmers so access to these data is granted exclusively to authorized users. Raw IoT data are retained for a maximum period of 1 month, after which they are permanently deleted. Data collected for model training purposes are used only upon explicit informed consent of the data owner. Aggregated outputs (e.g., indicators, maps) derived from farm-level data do not allow re-identification of individual farms or animals.

The platform exposes its functionalities through unified APIs, consumed internally by the services implemented within the Data Portal (e.g., THI monitoring, milk production forecasting, pasture biomass estimation, and disease risk assessment). Access to the APIs is regulated through an API key-based authentication mechanism, ensuring that only authorized services and users can retrieve or submit data. This approach guarantees controlled and secure access to sensitive farm-level information, such as IoT sensor data and animal production records.

The developed architecture ensures a robust, scalable, and efficient data management, enabling smarter decision-making in livestock breeding by integrating environmental, IoT, and satellite data (See Fig. 1).

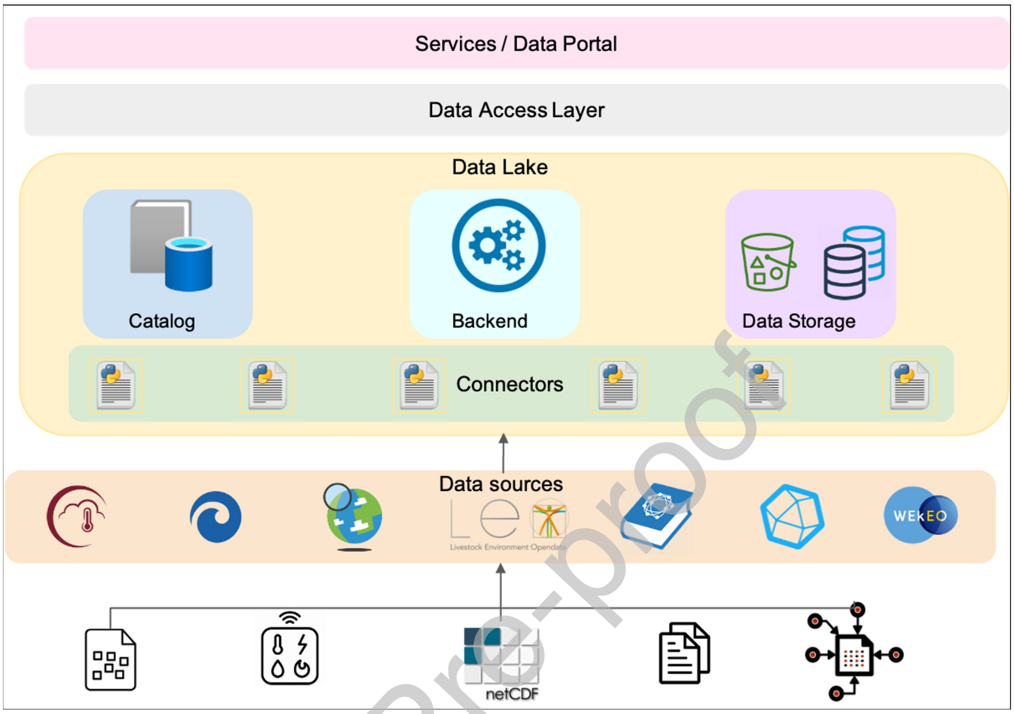

Figure 1: **Platform Architecture**

3.1. Data Storage

The Data Storage component is tasked with the temporary or permanent storage of data sourced from various origins. Specifically, it is utilized for storing developed products such as indicators and real-time data from IoT Systems for monitoring animal welfare. Additionally, it caches data pertaining to dataset queries from external sources like Copernicus Services and DIASs.

The implementation of the Data Storage employs different technologies depending on the type of data to be stored. Notably:

- An **Object Storage** based on MinIO is utilized for storing files in various formats (e.g., `netCDF`, `csv`, `zip`) associated with indicators, as well as for caching results of data queries linked to external data sources.

- A **Parallel High-performance storage file system** based on GPFS:

used for storing raw files (e.g., NetCDF) in order to guarantee near real time execution of complex and concurrent data queries. GPFS allows high-throughput I/O performance and native support for POSIX-compliant parallel access, which is essential for handling large-scale, multi-dimensional climate and satellite datasets. MinIO and GPFS operate as independent storage backends: GPFS handles raw, large datasets requiring fast parallel read/write operations, while MinIO manages derived products and cached query results. Data exchange between the two layers is mediated by the Backend, which routes read and write operations to the appropriate storage backend depending on the request to be fulfilled and the operations to be performed.

3.2. Data Source Connectors

To harmonize the integration of diverse data sources (e.g., external data portals and services like Copernicus Services and DIASs, files stored in Object Storages, and source raw data), purpose-built connectors are developed for data access and retrieval. These connectors employ a variety of techniques and optimizations, including:

- Caching,
- Multi-dimensional subsetting on the data source side,
- Efficient in-memory access.

Table 1 outlines the main datasets available for each data source, the technologies used to construct the connector, the format of the retrieved result, and the related services. All data are open access.

3.3. Backend

The primary responsibility of the *Backend* is to receive data requests originating from the *Data Access Layer* and to carry them out by interacting with the other components of the *Data Lake*. The retrieval process for accessing requested data can be implemented either synchronously or asynchronously. To illustrate the communication between the various components, from the initial user request to the provision of data, *Unified Modeling Language* (UML) sequence diagrams are provided for both types of processes.

The sequence diagram depicted in Fig. 2 illustrates the operations and communication among various components of the *Data Lake* during an asynchronous process, particularly when a data request from the *Data Access*

Table 1: **Overview of Data Sources, Connector Technologies, Retrieved Formats and Related Services**

Data Source	**Main Datasets**	**Connector Technology**	**Retrieved Format**	**Service**
Copernicus C3S	ERA5-Land	CDS API Python, client (cdsapi)	`NetCDF`, `GRIB`	Service 1
CMCC DDS	VHR-REA_IT (VHR-PRO_IT) very high resolution reanalysis (projection)	CMCC DDS Python client (ddsapi)	`NetCDF`	Service 1, Service 2, Service 4
Copernicus Open Access Hub	Sentinel 1 and 2	Python based requests	`SAFE`	Service 3
HIGHLANDER Data Portal	VHR-REA_IT and VHR-PRO_IT	Python based requests	`NetCDF`	Service 1, Service 2, Service 4
Mistral portal	COSMO-2I forecasts	Python library	`NetCDF`	Service 1
IoT AnimalTalker	IoT real-time data	`MQTT`	`JSON`	Service 1, Service 2, Service 4

Layer pertains to an external *Data Source*. The *Data Access Layer* initiates the request and forwards it to the *Backend*. Upon receiving the request, the *Backend* generates a request identifier and sends it back to the *Data Access Layer*. The *Data Access Layer* periodically polls the *Backend*, typically every two seconds, to monitor the status of the request until it is completed.

The *Backend* collaborates with the *Catalog* to gather information regarding the data source and the connector required to access the relevant dataset.

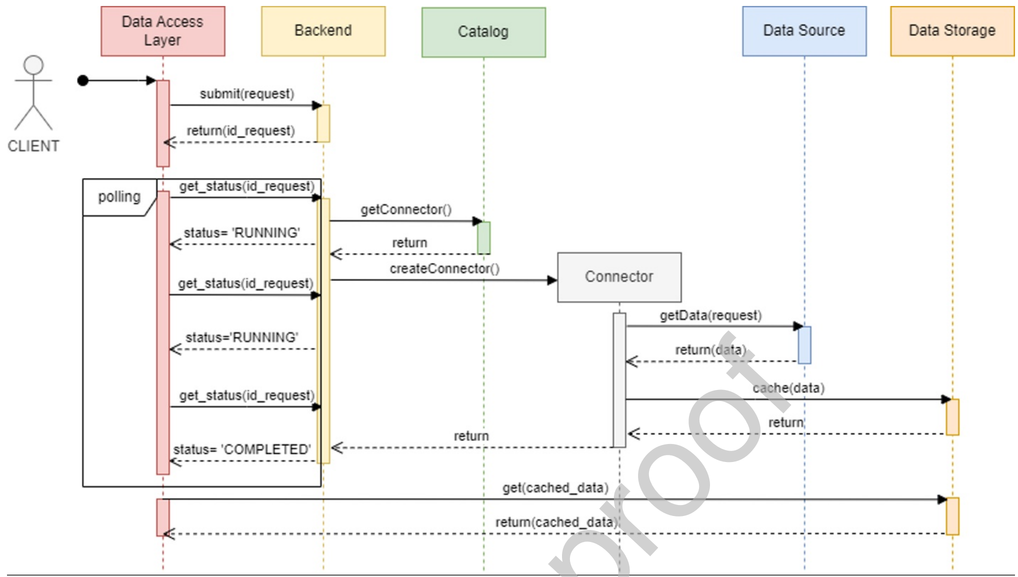

Figure 2: **Sequence Diagram for retrieving data from the Data Lake using the asynchronous mode**

Subsequently, the *Backend* instantiates the *Connector* object responsible for executing the query towards the corresponding dataset's datasource and caching the result in the *Data Storage*. Once the request is fulfilled, the *Backend* informs the *Data Access Layer* about the location of the cached data within the *Data Storage* and the *Data Access Layer* retrieves the data from the specified location in the *Data Storage*.

Similarly, during a synchronous process (Fig. 3), for instance, for accessing and analyzing streamed IoT real-time data, the *Data Access Layer* initiates the request and sends it to the *Backend*, entering a blocked state until the request is fulfilled. Upon receiving the request, the *Backend* communicates with the *Catalog* to obtain information regarding the data source and the connector required for accessing the relevant dataset. Based on the retrieved information, the *Backend* instantiates the *Connector* object responsible for executing the query directly towards the *Data Storage*.

The *Connector* executes the query to retrieve the required real-time data. Once the query is processed, the result is returned directly to the *Data Access Layer*, fulfilling the synchronous request. The *Data Access Layer* unblocks and continues with further processing or analysis based on the received real-

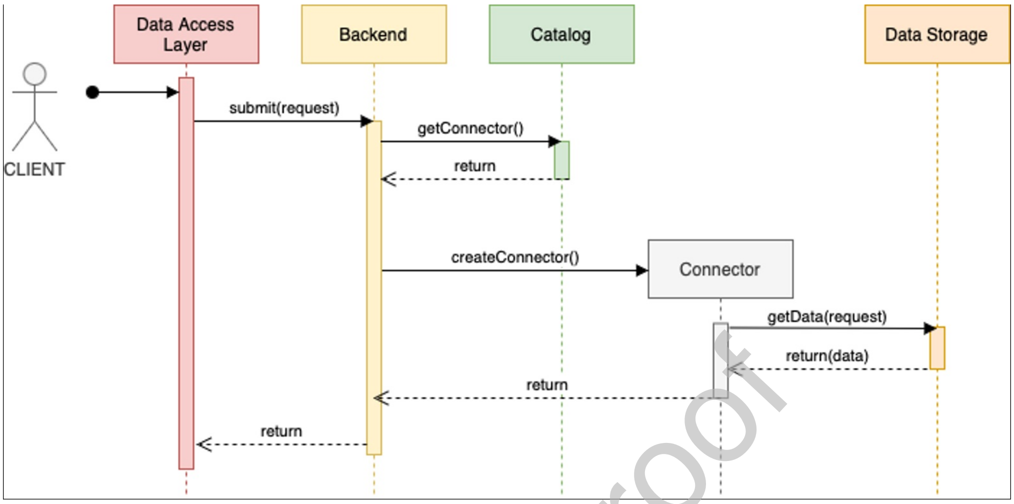

Figure 3: **Sequence Diagram for retrieving data from the Data Lake using the synchronous mode**

time data.

3.4. Data Catalog

The *Catalog* element within the *Data Lake* is designed to house comprehensive information about the diverse data sources. This encompasses details such as connection and access methods, available datasets, dataset metadata (including dataset identifiers within the data source and parameters required for data retrieval), and the associated data source connector. Information within the *Catalog* is stored in a dedicated database termed *MetaDB*. Through an API interface, the *Catalog* exposes various operations that the *Backend* can utilize to access information necessary for executing data requests.

Based on the requirements analysis of *MetaDB*, the primary entities that need representation are the data sources and the datasets available within each data source. Considering the potential variability in metadata schemas across datasets, a NoSQL database emerged as a suitable choice for implementing *MetaDB*.

MongoDB was selected as the underlying database for *MetaDB* for several reasons. First, unlike relational databases, MongoDB's document-oriented NoSQL architecture does not require a fixed schema, making it well-suited for managing heterogeneous metadata structures associated with datasets from diverse sources (e.g., Copernicus, CMCC DDS, IoT systems), which may differ significantly in their attributes and formats. Second, MongoDB natively supports nested and variable-length documents, allowing complex metadata structures to be stored and retrieved efficiently without the need for costly JOIN operations. Third, its horizontal scalability and support for high-throughput read/write operations make it appropriate for a data-intensive platform such as SEBASTIEN, where metadata entries are continuously updated via event-driven triggers. By modeling data sources and datasets as separate collections, *MetaDB* can effectively manage the dynamic nature of metadata schemas across the integrated data sources.

The *MetaDB* model includes two collections: one for data sources and another for datasets. To ingest datasets metadata into *MetaDB*, two potential methods are employed:

- **Utilizing Data Source API:** If a data source provides an API for dataset metadata, this API can be leveraged for indexing purposes. For instance, platforms like Wekeo (https://www.wekeo.eu/), CMCC DDS (https://dds.cmcc.it/), and HIGHLANDER provide endpoints for metadata extraction.
- **Customized Web Crawler:** For data sources lacking an API, a web crawler can be implemented to autonomously index metadata from the source's website.

The *Catalog* component provides the following API operations:

- `get_datasets(id)`: Retrieves the list of available datasets for a specific data source identified by `id`.
- `get_datasource(id)`: Retrieves the data source related to the dataset identified by `id`.
- `get_connector(id)`: Retrieves the connector corresponding to the data source identified by `id`.
- `get_metadata(id)`: Retrieves the list of parameters for a given dataset identified by `id`.

These operations enable the *Backend* to seamlessly obtain essential information for processing datasets within the *Data Lake*.

3.5. IoT Sensors Data Management

Environmental and animal sensors were developed during the project and employed to gather real-time data related to animals (both in barns and pastures) and their environments. These sensors provide crucial information to identify stressors such as heat stress, poor pasture quality, or health issues. Continuous and detailed phenotyping of animals using IoT is essential for evaluating animal welfare and preventing stress conditions. Additionally, these sensors can send real-time alerts based on the conditions of the animals and their environment. The collected data is published on a Message Queuing Telemetry Transport (MQTT) broker as JSON formatted messages.

Animal Sensors. The animal sensor consists of a collar that gathers data needed to assess animal welfare and makes it available remotely. The data collected includes movements (using an accelerometer), ambient temperature and relative humidity, GNSS position, and heart beat rate (experimental). Data are acquired at different frequencies and aggregated on-device: accelerometer, temperature, and humidity are sampled in 5 s windows every 5 min (12 times per hour) and averaged, while heart rate is computed over a 30 s window per hour. GNSS position is recorded hourly, and all data are transmitted once per hour. The device includes a tri-axial accelerometer with a ± 2 g full scale and 12-bit resolution, and a multi-constellation GNSS module with 1.5 m horizontal positioning accuracy. The heart rate is estimated via an optical sensor (red/infrared) based on blood volume variations during the cardiac cycle. Additional details are reported in Supplementary Materials (Text S1).

The collected data are used to calculate various indices and indicators related to animal health and well-being, which can be influenced by climatic events, for example.

Most parameters from the sensors independently represent animal welfare. However, some analysis is needed to provide clearer and more useful information. For instance, movement data can reveal the activity level of the animal, and variations may indicate anomalies or, for female animals, estrus periods. Heart rate evaluation can detect abnormal behavior in individual animals or the herd; for example, if all animals exhibit abnormal heart beat rates, the cause may be environmental.

Environmental Sensors. The environmental sensor platform measures the concentration of certain gases in the air, specifically CO_2, H_2S, NH_3, and

CH_4 (measured in parts per million); and particulate matter PM_1, $PM_{2.5}$, and PM_{10} (measured in $\mu g/m^3$). The sensors are factory-calibrated and may require periodic recalibration to maintain measurement accuracy. They are also modular and can be easily replaced in case of failure.

The importance of the gases and particulate matter in the livestock sector, as well as the characteristics of each sensor, are described in the Supplementary Materials (Text S2, Table S1).

The environmental sensor platform also measures environmental temperature and relative humidity. These measurements are used to compute the THI, which stakeholders can use to monitor the health and stress levels of animals in the barn, and provides a general evaluation of air quality inside barns that can actively impact human and livestock health.

4. Advanced Services for Intelligent Breeding

Platform's goals involve harmonising existing data resources effectively by employing models ranging from classical statistical inference to Machine Learning approaches, including Gradient Boosting Machines and linear mixed models. The aim is to derive quantitative and qualitative indicators to monitor and identify the impact of climate and environmental stresses on livestock systems.

To achieve this, four primary services have been developed. Several approaches were explored to process, rectify, and integrate the collected climatic, territorial, and animal data. Pipelines were established with the aim of generating indicators and indices beneficial to stakeholders. A range of empirical approaches, as well as statistical and mathematical techniques (such as regression, clustering, ML, etc.), were experimented to construct prediction models.

Methods and Machine Learning Workflows

The search for the best ML algorithm family was conducted using the H2O.ai AutoML (https://h2o.ai/platform/h2o-automl/) and scikit-learn (https://scikit-learn.org/stable/) modules from Python. H2O.ai AutoML automatically trains and compares multiple Machine Learning algorithms. These algorithms were explored across the different services depending on the characteristics and size of the available datasets. In particular, GBM, XGBoost, and Distributed Random

Forest (DRF) were applied across multiple services, including milk yield prediction (Service 1a), indoor THI estimation (Service 2), Bluetongue modelling in Sardinia (Service 4a), and Somatic Cell Count prediction (Service 4b). Extremely Randomized Trees (XRT) were additionally tested for Services 1a and 2. Deep Learning models were considered for services characterized by larger datasets, such as Service 1a and Service 4b. Generalized Linear Models (GLM) were included as a baseline statistical approach within the AutoML framework. For Service 3, a linear regression model was adopted. The optimal algorithm hyperparameters are then determined through a grid search. Depending on the nature of the target variable, different metrics (e.g., MAE) are used to evaluate the best algorithm or the number of features to retain.

Using the best algorithm and parameters identified, the complete dataset is analyzed to select the most informative features associated with the target variable according to the 'feature importance' metric. Feature importance scores were computed using the training data only, and used to identify the most informative predictors associated with the target variable. Finally, the algorithm is trained using only the previously identified subset of features.

4.1. Service 1: Temperature-Humidity Index (THI) Evaluation

THI is a bioclimatic index that assesses livestock stress by combining temperature and relative humidity effects. THI is one of the most widely used metrics globally for determining comfort levels, stress levels, and life-threatening environmental conditions induced by heat stress in livestock [28]. Heat stress poses significant risks to animal health, subsequently impacting productivity. Since most research on heat stress in livestock has focused primarily on temperature and relative humidity, THI serves as a single value encapsulating the cumulative impact of these variables associated with thermal stress. This service is dedicated to evaluating the THI inside stables and the THI formula [29] used is:

$$THI = (1.8 \times T + 32) - (0.55 - 0.55 \times RH) \times [(1.8 \times T + 32) - 58] \quad (1)$$

where:

- T is the temperature in degrees Celsius ($^{\circ}C$).
- RH is the relative humidity (%).

Temperature		% Relative Humidity																		
°F	°C	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
72	22.0	64	65	65	65	66	66	67	67	67	68	68	69	69	69	70	70	70	71	71
73	23.0	65	65	66	66	66	67	67	68	68	68	69	69	70	70	71	71	71	72	72
74	23.5	65	66	66	67	67	67	68	68	69	69	70	70	70	71	71	72	72	73	73
75	24.0	66	66	67	67	68	68	68	69	69	70	70	71	71	72	72	73	73	74	74
76	24.5	66	67	67	68	68	69	69	70	70	71	71	72	72	73	73	74	74	75	75
77	25.0	67	67	68	68	69	69	70	70	71	71	72	72	73	73	74	74	75	75	76
78	25.5	67	68	68	69	69	70	70	71	71	72	73	73	74	74	75	75	76	76	77
79	26.0	67	68	69	69	70	70	71	71	72	73	73	74	74	75	76	76	77	77	78
80	26.5	68	69	69	70	70	71	72	72	73	73	74	75	75	76	76	77	78	78	79
81	27.0	68	69	70	70	71	72	72	73	73	74	75	75	76	77	77	78	78	79	80
82	28.0	69	69	70	71	71	72	73	73	74	75	75	76	77	77	78	79	79	80	81
83	28.5	69	70	71	71	72	73	73	74	75	75	76	77	78	78	79	80	80	81	82
84	29.0	70	70	71	72	73	73	74	75	75	76	77	78	78	79	80	80	81	82	83
85	29.5	70	71	72	72	73	74	75	75	76	77	78	78	79	80	81	81	82	83	84
86	30.0	71	71	72	73	74	74	75	76	77	78	78	79	80	81	81	82	83	84	84
87	30.5	71	72	73	73	74	75	76	77	77	78	79	80	81	81	82	83	84	85	85
88	31.0	72	72	73	74	75	76	76	77	78	79	80	81	81	82	83	84	85	86	86
89	31.5	72	73	74	75	75	76	77	78	79	80	80	81	82	83	84	85	86	86	87
90	32.0	72	73	74	75	76	77	78	79	79	80	81	82	83	84	85	86	86	87	88
91	33.0	73	74	75	76	76	77	78	79	80	81	82	83	84	85	86	86	87	88	89
92	33.5	73	74	75	76	77	78	79	80	81	82	83	84	85	85	86	87	88	89	90
93	34.0	74	75	76	77	78	79	80	80	81	82	83	85	85	86	87	88	89	90	91
94	34.5	74	75	76	77	78	79	80	81	82	83	84	86	86	87	88	89	90	91	92
95	35.0	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93
96	35.5	75	76	77	78	79	80	81	82	83	85	86	87	88	89	90	91	92	93	94
97	36.0	76	77	78	79	80	81	82	83	84	85	86	87	88	89	91	92	93	94	95
98	36.5	76	77	78	80	80	82	83	83	85	86	87	88	89	90	91	92	93	94	95
99	37.0	76	78	79	80	81	82	83	84	85	87	88	89	90	91	92	93	94	95	96
100	38.0	77	78	79	81	82	83	84	85	86	87	88	90	91	92	93	94	95	96	98
101	38.5	77	79	80	81	82	83	84	86	87	88	89	90	92	93	94	95	96	98	99
102	39.0	78	79	80	82	83	84	85	86	87	89	90	91	92	94	95	96	97	98	100
103	39.5	78	79	81	82	83	84	86	87	88	89	91	92	93	94	96	97	98	99	101
104	40.0	79	80	81	83	84	85	86	88	89	90	91	93	94	95	96	98	99	100	101
105	40.5		80	82	83	84	86	87	88	89	91	92	93	95	96	97	99	100	101	102
106	41.0	80	81	82	84	85	87	88	89	90	91	93	94	95	97	98	99	101	102	103
107	41.5	80	81	83	84	85	87	88	89	91	92	94	95	96	98	99	100	102	103	104

Figure 4: **THI table and relationship with temperature and relative humidity. Colors represent THI values ranging from comfort (white) to life threatening (purple) for dairy cattle [30]**

THI is expressed in degrees Fahrenheit ($^{\circ}F$).

Fig. 4 illustrates how THI correlates with temperature and relative humidity. THI denotes stress levels ranging from low to severe, starting at a value of 67 in dairy cows.

The evaluation approach involves a ML procedure that correlates various input parameters with the resulting THI value. Specifically, the input drivers include: i) the stable's latitude, ii) the stable's longitude, iii) the stable's altitude, iv) the month of the measurement, and v) the external THI at the nearest location to the stable.

4.1.1. Service 1a: Forecast of stables short-term environmental conditions

The first subservice of Service 1 (1a) aims to predict the variation of THI inside a stable over the next two days, with hourly resolution. To this end, we developed a ML approach that estimates the internal THI of a stable given its latitude, longitude, altitude, month (Jan, Feb, . . . , Dec), and external THI. In the training phase, data from 658 stables distributed across Italy were used, covering a monitoring period from November 2022 to September 2023. The distribution of stables within the Italian territory is reported as Supplementary Materials (Fig S1). Each stable was monitored hourly for internal temperature and relative humidity, which were used to calculate the internal THI as 'ground truth' for the ML model. It should be noted that information on stable types (e.g., open, closed, or semi-open) and the presence of ventilation or cooling systems was not available for the monitored stables, which may influence the internal-external THI relationship and represents a limitation of the current study. Future work will aim to incorporate such management-related variables to improve model generalizability. Regarding seasonal coverage, the monitoring period from November 2022 to September 2023 ensures a nearly uniform distribution of records across seasons, reducing potential seasonal bias in the training dataset. Additionally, data on external THI for these stables were collected from the ERA5-Land reanalysis, which provides hourly 2m temperature and 2m dewpoint temperature (converted to relative humidity). These data were downscaled to the COSMO-2I grid (2.2 km resolution) to compute the external THI index.

The ML model was set to learn the mapping between the external THI and the internal THI. For each stable, the external THI value was matched to the nearest COSMO-2I grid point based on geographical coordinates. Internal THI data preprocessing included outlier detection using the Z-score method and temporal alignment to fit the hourly resolution of the external data. The internal and external data for each stable were then combined, matching records by ID, date, and hour, to create a unified dataset.

The training phase used a subset of 450,000 records from the unified dataset, split into 80% training and 20% test sets. To avoid potential data leakage due to stable-specific temporal patterns, the split between training and test sets was performed based on stable ID, thus ensuring that all records from the same stable were allocated exclusively to either training or testing data. The best ML algorithm family was identified using H2O.ai AutoML; the GBM_4 algorithm achieved the best accuracy with an RMSE of 2.587 (Table 2).

Table 2: **Performance Metrics of Different Models**

Model ID	RMSE*	MAE*	MSE*	R^2*
GBM_4	2.587	1.907	6.693	0.950
GBM_1	2.598	1.915	6.752	0.949
GBM_3	2.640	1.952	6.972	0.948
GBM_2	2.659	1.967	7.069	0.947
GBM_5	2.695	1.997	7.263	0.945

* Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE) and R-squared (R^2) are used to predict the model accuracy. They provide an estimate of the typical magnitude of prediction errors. For RMSE, MAE and MSE, lower values indicate better model performance; for R^2, higher values (closer to 1) indicate better model performance.

To improve the interpretability of its prediction performance, the test-set errors were further stratified according to external THI severity ranges: comfort, moderate, and severe heat stress. As shown in Table 3, prediction errors progressively decreased with increasing THI severity. Specifically, MAE declined from 2.58 in the comfort range to 1.11 under severe heat stress. This finding indicates that the proposed model maintains, and even improves, predictive reliability under the most critical environmental conditions for farm-level decision support.

Table 3: **Performance Metrics by THI Range**

THI Range	MAE	RMSE	R^2
Comfort	2.580	3.450	0.890
Moderate	1.320	1.790	0.430
Severe	1.110	1.550	0.400

4.1.2. Service 1b: Projection of stables long-term environmental conditions

Service 1b extends this by assessing THI variations inside stables for near- and long-term horizons under *IPCC-RCP4.5* and *RCP8.5* scenarios. Future climate projections in these scenarios are used to compute internal THI changes compared to a thirty-year baseline (1981-2010), providing insights into potential climate change impacts on livestock welfare. THI changes are defined as differences between projected and baseline values, computed from monthly mean THI derived from hourly data. Hourly THI values are first calculated and then aggregated into monthly means, and changes are obtained by comparing future and baseline monthly averages. In addition, results are also expressed as the increase in the number of heat stress days during summer, derived from hourly THI values exceeding a predefined threshold. The data will show expected changes between 30-year future periods and a baseline (1981-2010).

The climate projections for the second subservice come from VHR-PRO_IT (Very High-Resolution PROjections for Italy), an open-access hourly climate projection with a resolution of approximately 2.2 km from 1981 to 2070, covering Italy and neighboring areas. VHR-PRO_IT was produced within the Highlander project by dynamically downscaling the Italy 8 km CM climate projection (spatial resolution about 8 km; output frequency = 6 h; driven by the CMIP5 GCM = CMCC-CM) using the Regional Climate Model COSMO-CLM. Its global forcing includes the historical experiment for 1981–2005 and the RCP4.5 and RCP8.5 greenhouse gas concentration trajectories for 2006–2070 [31].

The model used in Service 1a is also applied in Service 1b. This model identifies the relationship between input variables (latitude, longitude, altitude, month of the measurement, and external THI) and the internal THI of the stable. Thus, the model can determine the internal THI using future climate projections (external THI) as input variables. Data are presented as expected changes between 30-year future periods and the baseline (1981-2010). This choice reduces the influence of systematic biases in the climate projections, as the analysis focuses on relative differences rather than absolute values. It is worth noting that bias correction was not explicitly applied here, considering the high computational cost associated with hourly resolution across numerous grid points, although it could further improve the robustness of absolute estimates.

4.2. Service 2: Percentage Variation of Milk Yield, Protein, and Fat Content

To help farmers minimize the consequences of heat stress, developing a model to predict the effects of climate variation on livestock in both the short and long term can be a useful tool for adapting current and future farm management practices. This Service investigates the impact of climate change on milk production and quality. A ML model was developed using milk production and milk quality data from the *Pezzata Rossa Italiana* breed in the Friuli-Venezia Giulia region (1990–2020). Data underwent analysis using a *Linear Mixed Model* (LMM), correlating production residues with climatic variables.

The main outputs are three indices reported as a color scale, from green to red, representing the expected variation in milk yield, fat content, and protein content for both long-term (static) and short-term (dynamic) predictions. When the color shifts to red, higher is the expected loss in milk quantity and quality. These indices aim to provide stakeholders with valuable insights to make informed decisions and implement strategies to mitigate potential impacts.

To properly address stakeholders' needs, the service was divided into three sub-services.

4.2.1. Service 2a: Production decline

The objective of Service 2a is to create a ML model that predicts the effects of heat stress on livestock, applicable to short-term weather forecasts (2 days) and long-term climate projections.

In this service, we implemented a ML prediction model using animal-based and bioclimatic data. For the animal data, we utilized:

- Production data, specifically milk yield, protein, and fat percentage from "Pezzata Rossa Italiana" (Italian Simmental) sourced through the LEO project (https://opendata.leo-italy.eu/portale/home) and ANAPRI (Associazione Nazionale Allevatori Bovini di Razza Pezzata Rossa Italiana) (https://www.anapri.eu/it/).

- Associated data, including days in milk, number of lactations, age, and number of functional controls, to correct observed values for environmental influences. Estimated breeding values (EBVs) were also considered, provided by ANAPRI.

For the climatic data, we used:

- Single climatic variables such as temperature, relative humidity from the VHR-REA dataset (Very High Resolution Dynamical Downscaling of ERA5 Reanalysis over Italy is a very high-resolution, 2.2 km hourly, climate dataset for Italy, produced by dynamically downscaling ERA5 data using the COSMO model from 1981) [32, 33].

- Climatic indices such as the THI for external conditions from the VHR-REA dataset, and internal conditions data from Service 1 results.

A pilot dataset was created using production data from 1990 to 2020 from Friuli-Venezia Giulia region, totaling 2,511,947 Functional Control (FC) records, from 1,115 farms and 101,595 animals. The dataset was cleaned of outliers and incomplete data, retaining records with days in milk (DIM) between 5 and 400, parity up to 9, animals older than 22 months, and number of FC per lactation between 5 and 14. The climatic effect on production was evaluated up to 30 days before the FC to assess both short and long-term impacts on milk production and quality. The dataset was split as follows: 80% for the training phase and the remaining 20% for testing phase.

Different models were developed to test single climatic variables and the THI for both external (pasture) and internal conditions.

The first analysis step involved applying a multiple Linear Mixed Model to each phenotype to correct for fixed and random effects. Fixed effects included DIM (in 15-day intervals), age in months, parity (grouped from 1 to 6, with 7+ as a single class), and IDAS EBV (Sustainable Double Purpose Index, from ANAPRI). Animal and farm identifications were included as random effects to account for data repetition and farm management differences. The LMM was implemented in R using the lme and lmerTest packages and it can be expressed as:

$$Y_{ijklmn} = \mu + DIM_i + Age_j + Parity_k + EBV_l + a_m + f_n + \epsilon_{ijklmn} \quad (2)$$

where Yijklmn is the observed phenotypic value, DIM represents days in milk (grouped in classes), Age is the age at recording, Parity is the number of lactations, and EBV is the estimated breeding value. Random effects include animal ID (am) and farm ID (fn), accounting for repeated measurements and farm-specific conditions, respectively.

The goal was to obtain residual values, computed as the difference between observed and predicted values from the LMM, that reflect the error of the model and the environmental effects, which were then evaluated using the ML model. Climatic variables were analyzed using a correlation matrix to check for autocorrelation. Strong correlations were found, particularly between the same variable on different days and between different climatic variables (e.g., maximum and average temperature). Consequently, some variables (e.g., average temperature) were removed and statistical transformations (e.g., mean or sum) were applied to the remaining variables to reduce high correlation and avoid bias. Data were regrouped according to specific ranges for each climatic variable: 5 days for temperature, 3 days for relative humidity, 2 days for wind speed, and 2 days for cloud coverage. Precipitation data were not regrouped. These aggregation windows were selected to balance temporal resolution and the reduction of autocorrelation, while also reflecting the different temporal dynamics of climatic variables and their effects on animal response.

Animal phenotype (target variable - residual from the linear model) and climatic (features) data were combined into a single dataset for subsequent ML analyses. The previously described ML workflow was applied using H2O.ai. The first step of the ML pipeline was to identify the best family algorithm (Table 4). A GBM resulted as the best model for all three phenotypes.

Table 4: **Comparison of Model Performance Metrics for milk yield**

Rank	**model ID**	**RMSE**	**MSE**	**MAE**	**Mean residual deviance**
1	GBM_ 15	3,84	14,7	2,91	14,7
2	GBM_ 5	3,84	14,7	2,91	14,7
3	GBM_10	3,84	14,7	2,91	14,7
4	GBM_4	3,84	14,7	2,91	14,7
5	XRT_0	3,84	14,8	2,91	14,8
6	GBM_8	3,84	14,8	2,91	14,8
7	DRF_0	3,84	14,8	2,91	14,8
8	GBM_26	3,85	14,8	2,92	14,8
9	GBM_1	3,85	14,8	2,92	14,8
10	GBM_0	3,85	14,8	2,92	14,8

Once the best algorithm for each phenotype was identified, a grid search was conducted to optimize its hyperparameters. Then, this optimized ML model was used to determine the importance of the features. Next, we identified the optimal number of features for the ML model by minimizing the MAE, with the first four features emerging as the most important. With the optimal features determined, the ML model was trained to create the prediction model. The following tables provide details on the ML analyses for milk yield, fat, and protein with the selected climatic variables, and list the selected characteristics (See Tables 5, 6).

Table 5: **Identification and evaluation of the best Machine Learning model using the climatic variable as values**

Feature	Algorithm	Proxy	RMSE	MAE	R-squared	Nr of features*
Milk yield	Gradient Boosting Machine	Production	28.963	26.797	0.1979	4
Fat	Gradient Boosting Machine	Milk quality	0,3988	0.3778	0.1836	6
Protein	Gradient Boosting Machine	Milk quality	0.2101	0.1517	0.2304	7

* Number of features selected for the model

After evaluating the algorithm's performance, a SHAP (SHapley Additive exPlanations)[34] analysis, using the H2O.ai functionality, was conducted on the test set to identify and explain the most important variables. The SHAP plot shows each feature's contribution to the prediction, such as high values of the minimum temperature "avg_T_min_1-5" days before functional control negatively impacting predictions, while low values have a positive impact. (See Fig. 5)

The dataset used in this experiment is extensive, combining nearly 100 features with over 2.5 million phenotypic data points, totaling almost 0.25 billion data points. ML approaches are more efficient for such large datasets compared to classical methods, as they are designed for big data and can reveal complex, nonlinear relationships that traditional linear models might

Table 6: **Features (i.e. climatic variables) selected for each phenotype analysed**

Variable	Feature selected
Milk yield	avg_T_MIN_1-5 avg_T_MAX_1-5 avg_T_MIN_26-30 avg_WS_KMH_5-6
Fat	avg_T_MAX_1-5 avg_T_MAX_6-10 avg_WS_KMH_3-4 avg_WS_KMH_15-16 avg_WS_KMH_21-22 avg_WS_KMH_1-2
Protein	avg_T_MAX_1-5 avg_T_MIN_1-5 avg_WS_KMH_27-28 avg_WS_KMH_7-8 avg_WS_KMH_1-2 avg_WS_KMH_29-30 avg_WS_KMH_19-20

miss.

4.2.2. Service 2b: Adaptability of species/breeds

Cattle are highly susceptible to heat stress, which occurs when their core body temperature rises beyond their ability to dissipate heat. This condition compromises their health, reduces feed intake, lowers milk production, decreases fertility, and can lead to mortality in severe cases. Different breeds have varying levels of heat tolerance, with some better adapted to hot climates than others. The THI is a crucial parameter for assessing the impact of temperature and humidity on cattle comfort and identifying heat stress [35].

In Service 2b, we developed a tool that integrates THI data predictions for both external (pasture) and internal conditions (barn), along with breed-specific heat tolerance information. This tool helps farmers make informed breed selection decisions in anticipation of rising THI values in the coming decades.

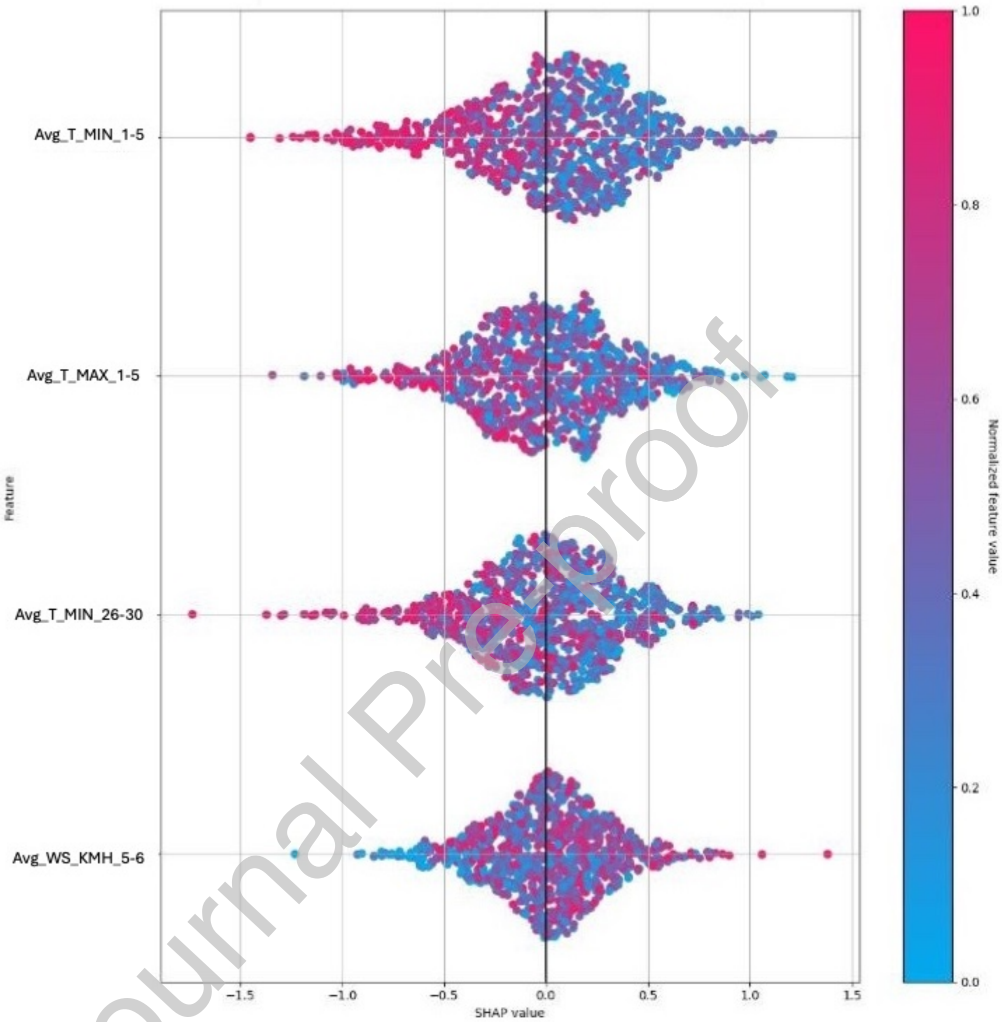

Figure 5: **Identification of the most important features involved in the prediction of the target variable "milk yield production". The variables are reported on the y-axis and are sorted from the most important (at the top of the graph) to the less important (on the lower part of the graph). On the x-axis, the SHAP value is reported. Each dot represents a sample used in the test set. Each sample is colored according to the corresponding normalized feature value**

To identify THI tolerance thresholds for dairy and beef cattle and specific breeds, a comprehensive literature review was conducted, encompassing papers and publications on THI tolerance in cattle (See Table 7).

Table 7: **THI thresholds for bovine, according to literature**

Name	**Group**	**No stress**	**Moderate stress**	**High stress**	**Extreme stress**	**Reference**
Beef	purpose	THI ¡72	72 ¿= THI ¡ 82	82 ¿= THI ¡ 94	THI ¿=94	[13]
Dairy	purpose	THI ¡72	72 ¿= THI ¡ 79	79 ¿= THI ¡ 89	THI ¿=89	[13]
Holstein	breed	THI ¡72	72 ¿= THI ¡ 79		THI ¿=79	[14]
Jersey	breed	THI ¡72	72 ¿= THI ¡ 90		THI ¿=90	[14]
Brown Swiss	breed	THI ¡72	72 ¿= THI ¡ 83	83¿= THI ¡ 89	THI ¿=89	[15]

The no-stress threshold of THI ¡ 72 was adopted as the reference value in Table 7 as it represents the most widely used and consolidated threshold in the livestock literature for classifying heat stress in dairy and beef cattle. However, as noted in Section 4.1, a mild but non-negligible level of stress may already occur in dairy cows from THI = 67, with the range 67–72 representing a transitional zone of low-level thermal discomfort. For each geographical location and breed with available THI tolerance information, we created a color-coded system ranging from green to red to indicate no to severe heat stress, respectively. This system can be indicative of a breed suitability at regional scale for future farming based on projected THI values in specific locations.

4.2.3. Service 2c: EBVs corrected for heat stress

The goal of this service is to identify animals that are genetically more resilient to adverse environmental conditions and can pass this trait to future generations. To achieve this, we collaborated with ANAPRI to estimate stress-resilience EBVs by applying ANAPRI models to phenotypic data collected under stressful environmental conditions, particularly in terms of THI measured inside the barn. This allows animals to be ranked according to the new EBV, which may differ from current EBV rankings. Farmers and breeding centers will thus have complementary information on the genetic potential of sires and dams under average conditions (current EBVs) and stress conditions (stress-resilience EBVs). This information will help farmers who want to breed animals for robustness and resilience in anticipation of climate change.

ANAPRI EBVs are referred to as IDAS ("Indice Doppia Attitudine Sostenibile" – Sustainable Dual-Purpose Index). The routine model includes, as fixed effects, the farm–FC effect (contemporary group), the combined effect of calving season nested within calving year, and the combined effect of lactation stage nested within age class at calving nested within calving order. Additionally, days in lactation effect is fitted nested within the lactation stage, itself nested within age class at calving and parity. Regarding the random effects, the model includes the additive genetic effect of the animal, the permanent environment, and the random error. The average THI inside the barn over the 5 days preceding the FC was also included as a covariate. Models were estimated both with and without the THI effect. As the calculation of this index is legally reserved for breeders' associations, extra model details cannot be disclosed.

4.3. Service 3: Pasture Biomass Evaluation

Managing an extensive system farm is challenging due to the difficulty of constantly monitoring animals and feed availability in terms of quantity and quality. Satellite data can be utilized to detect vegetation status, enabling the farmer to schedule and evaluate grazing availability and identify potential overgrazing. Service 3 is designed to predict the quantity of fresh and dry matter biomass within a user-defined area, such as a pasture: satellite data are integrated with pasture field data in a statistical model to evaluate pasture productivity and characteristics.

Pasture data was collected from two farms in the Lazio region, central Italy, with 17 and 16 sampling days for the first and second farm, respectively. The pasture management system employed is known as "rational pasture" where animals are moved through different sub-areas based on grass status. This system allows identifying simultaneous areas with minimum grass cover (just grazed) and areas with optimal grass cover. On each sampling day for each farm, field data was collected from three areas (5 square meters) corresponding to low, medium, and high levels of the NDVI (Normalized Difference Vegetation Index) Sentinel 2 index, which serves as a proxy for pasture status. This approach aims to achieve a distribution of values over time. Biomass, both fresh and dry matter, was collected from 297 sampling points, and laboratory analysis performed to evaluate fiber characteristics, lignin, protein, and fat. Climate and topographic data, which are important for correcting shading, radiation, and background effects, were also considered. However, these variables are not typically included in the models and, in future steps, we plan to include this information as fixed effects in the regression model or features in the ML models.

The Sentinel 2 satellite data used included bands (B2, B3, B4, B5, B6, B7, B8, B8A) and indexes (NDVI, NDWI - Normalized Difference Water Index, EVI - Enhanced Vegetation Index, GLI - Green leaf index, SAVI - Soil Adjusted Vegetation Index, GCI - Green Chlorophyll Vegetation Index, RGR - Simple Ratio Red/Green Red-Green Ratio, SIPI - Structure Insensitive Pigment Index, ARVI - Atmospherically Resistant Vegetation Index, NBRI - Normalized Burned Ratio Index).

Predictions were tested using well-documented approaches such as linear regression [36]. ML approaches, like random forest [37], were also considered; however, we opted not to apply them due to an insufficient amount of field data.

For each phenotype (fresh and dry matter), we tested single bands/indexes, all bands, all indexes, and combinations of bands and indexes in multiple linear regression models. For models with multiple fixed effects, only the significant ones were retained in the final model. For each model, we recorded the MAE, R-squared, AIC (Akaike information criterion), and overall p-value, computed on the training data. The model with the highest R-squared (i.e., indicating more efficient prediction) was selected for prediction purposes.

For fresh matter, the best model (bands: B2, B3 and B8; indexes: NDVI, NDWI, GLI, GCI and RGR) reached 0.47 R-squared. For dry matter, the best model (bands: B2, B4, B6 and B8A; indexes: NDVI, GLI, GCI, SIPI and ARVI) reached 0.25 R-squared.

We obtained lower R-squared values from both models when compared with previous analyses, likely due to the limited number of field samples collected and the use of samples from different pastures and in different times of the year; including additional factors such as climatic and topographic data could enhance model accuracy.

The model outputs the total quantity of fresh and dry matter biomass in the selected area. Within the service, users can visualize the number of animals or days that can utilize the area. This involves dividing the estimated total quantity by the expected fresh and dry matter biomass consumption for young beef cattle. Users have the flexibility to adjust these parameters for more accurate results. This calculation results in a service index, dynamically obtained to provide short-term predictions. These predictions offer valuable information to breeders, aiding real-time decision-making in managing their animals.

Additionally, field data pertaining to biomass quantity and characteristics are available as dynamic indicators. These indicators complement predictions, providing breeders with comprehensive insights into biomass dynamics and facilitating informed management practices.

4.4. Service 4: Risk of parasites and diseases spread

Service 4 aims to assist farmers and decision-makers in monitoring the spread of diseases caused by parasites (such as bluetongue in sheep) and health conditions (such as mastitis in cattle). Prediction models were developed using data derived from a previously published study [38]. These data include farm-level information (e.g., number of animals, number of infected and vaccinated animals, and management practices) for the year 2013. Climate projections were employed to anticipate future shifts in conditions conducive to parasites and diseases. The outcome of this service are risk maps indicating the spread of these parasites and diseases.

4.4.1. Service 4a: Probability of Developing the Blue-Tongue

Service 4a focuses on predicting the risk probability of blue-tongue infection, specifically for sheep in the island of Sardinia, Italy.

ML techniques were employed to generate risk maps for the spread of parasites and diseases, by integrating abiotic and biotic factors. For the Sardinia case study, a ML pipeline was applied to obtain a Logistic multilevel Mixed Model of bluetongue, a vector-borne disease transmitted by Culicoides midges.

To create the dataset for developing the ML model, three types of data were collected: farm-related information, climatic data, and environmental data. The farm-related data included the farm's latitude, longitude, and unique ID, provided by the Experimental Zooprophylactic Institute (IZS) of Sardinia. Additionally, the number of animals per farm, dates of confirmed bluetongue cases, and vaccination dates were recorded. If the vaccination date preceded the clinical case, the animals were considered vaccinated. The target variable was defined as the within-farm prevalence of bluetongue infection, calculated as the ratio between the number of infected animals and the total number of animals on the farm. Data from 2013, a year with numerous bluetongue cases, were used to provide valuable input for the initial ML model. Data was collected from 5,600 farms and complemented with climatic and environmental information, both critical for understanding the life cycle of *Culicoides*. Climatic data, organized in NetCDF files and sourced from the Highlander DDS, included variables like mean, minimum, and maximum temperatures, relative humidity, cloud coverage, precipitation, wind speed, and solar radiation. These were collected up to 60 days before clinical cases, with data averaged over five day intervals. To avoid collinearity, Pearson correlation coefficients were calculated, and highly collinear variables were removed using the Variance Inflation Factor. Environmental data was downloaded from the Sardinian geoportal website (https://www.sardegnageoportale.it/). Each farm's environmental characteristics were associated with the nearest polygon or, if multiple polygons were within 500m, the most frequent characteristic was used.

Several ML algorithm families were tested using the AutoML function in the H2O.ai package. The GBM was chosen for its lower MAE value on the validation set (see Table S2). A Grid search for parameter tuning was conducted. The final model, trained with 5-fold cross-validation, was developed using the 75% of the data as training set, 15% as test set and 10% as validation set. It used 43 of the 67 initial variables, as this subset provided the same MAE as the full model (MAE on test set 14.8% and 15.3% on validation set).

The output of the model is a quantitative index ranging from 0 to 100. A value of 0 indicates no animals in the stable develop the disease, while 100 indicates 100% of animals present in the stable develop the disease. This index serves as a valuable indicator of animal health and can be utilized for both static (long-term) and dynamic (short-term) predictions, offering insights into the risk of blue-tongue infection among sheep in the Sardinia region.

4.4.2. Service 4b: Somatic Cell Count Variation

Service 4b aims to study somatic cells (somatic cell count - SCC), which could be used as a proxy of mammary gland health. SCC varies due to factors like animal health, lactation stage, and breed. Increased SCC indicates environmental and stress-related changes, with mastitis causing a significant rise. Mastitis is a major concern for the dairy industry, leading to production losses, increased costs, and antibiotic resistance issues [39]. This study focuses on assessing the impact of environmental stress, particularly heat stress, on SCC.

The same pilot dataset, cleaned of outliers and incomplete data, and pipelines from Service 2a were utilized here enabling predictions of short-term (dynamic) and long-term (static) effects, in order to ensure methodological consistency across services. SCC data underwent a base 10 logarithm transformation for normalization and modelled as a continuous variable. A GBM ML model was employed, and its performance is reported in Table 8. Similarly, this service provides three indices displayed on a color scale ranging from green to red, indicating the expected variation in SCC levels. Regarding the feature importance analysis, the wind speed was identified as the most significant climatic variable (Table 9). Short- and long-term effects were observed, indicating acute and chronic impacts of stressful conditions. Interestingly, temperature, commonly present in other ML models, was not a key feature here, possibly due to the phenotype's focus on cattle health rather than production.

Table 8: **Identification and evaluation of the best Machine Learning model using the climatic variable as values**

Feature	**Algorithm**	**Proxy**	**RMSE**	**MAE**	**R-squared**	**Nr of features***
SCC	Gradient Boosting Machine	Health	0.4533	0.3468	0.0689	4

* Number of features selected for the model

Table 9: **Features (i.e. climatic variables) selected for the phenotype SCC**

Variable	**Feature selected**
SCC	sum_WS_KMH_5-6 sum_WS_KMH_1-2 sum_WS_KMH_29-30 sum_WS_KMH_15-16

5. Conclusions and future directions

This work presented SEBASTIEN, an integrated Decision Support System for smart livestock breeding that combines IoT sensors, satellite data, climate projections, and Machine Learning to support farm management under changing environmental conditions. The performance metrics obtained across the developed services suggest that the proposed data-driven approach holds significant potential for livestock monitoring and forecasting. Service 1a demonstrated reliable short-term prediction of internal stable THI, with the best GBM model achieving an RMSE of 2.587 and R^2 of 0.950 on a dataset of over 450,000 records from 658 Italian stables. Notably, prediction accuracy improved under the most critical heat stress conditions (MAE of 1.11 under severe stress vs. 2.58 under comfort conditions), supporting its value for farm-level decision-making. Service 1b extended this capability to long-term climate horizons using IPCC RCP4.5 and RCP8.5 projections, providing breeders with insights into expected changes in thermal stress conditions over the coming decades. Service 2a produced GBM-based predictive models for milk yield, fat, and protein content, identifying key climatic drivers — particularly minimum and maximum temperature in the 1–5 days preceding functional controls — through SHAP analysis. These models were developed on an extensive dataset of over 2.5 million phenotypic records, highlighting the advantage of ML approaches over classical methods for large-scale, nonlinear data. Service 2b integrated breed-specific THI tolerance thresholds to support informed breed selection decisions, while Service 2c provided stress-resilience Estimated Breeding Values (EBVs) to help farmers identify genetically robust animals. Service 3 addressed pasture biomass estimation by combining Sentinel-2 satellite data with field measurements. While the approach proved feasible, the models achieved moderate predictive performance (R^2 of 0.47 for fresh matter and 0.25 for dry matter), likely due to the limited number of field samples available. These results represent a preliminary but promising contribution, and future work will focus on expanding the field dataset and incorporating climatic and topographic covariates to improve model accuracy. Service 4a produced a cross-validated GBM model for bluetongue risk prediction in Sardinian sheep farms, integrating climatic, environmental, and farm-level data from over 5,600 farms. Service 4b explored Somatic Cell Count as a proxy for mastitis risk, identifying wind speed as the most relevant climatic driver, a finding that warrants further investigation. Taken together, these results indicate that the integration of heterogeneous data sources within a unified and scalable architecture can support both short-term operational decisions and long-term climate adaptation strategies in livestock farming. However, several limitations should be acknowledged: some services rely on data from specific regions or breeds, which may limit generalizability; external validation across diverse farm contexts remains to be conducted; and the operational effectiveness of the platform at scale requires further assessment. To maximize the impact of these solutions, greater stakeholder involvement, tighter integration with farm decision-making workflows, and continuous technological updates are required. The future of sustainable livestock farming will increasingly depend on the ability to adopt data-driven solutions capable of addressing the challenges posed by climate change and evolving market demands.

CRediT authorship contribution statement

A. D'Anca: Conceptualization, Supervision, Funding acquisition, Methodology, Writing – original draft, Project administration, Writing – review & editing. **M. Milanesi:** Conceptualization, Formal analysis, Funding acquisition, Validation, Methodology, Writing – original draft, Writing – review & editing. **G. Trotta:** Data curation, Software, Visualization, Methodology, Writing – original draft. **P. Ajmone Marsan:** Conceptualization, Supervision, Funding acquisition, Investigation, Methodology, Writing – review & editing. **R. Negrini:** Conceptualization, Supervision, Funding acquisition. Investigation, Methodology, Writing – review & editing. **A. Reder:** Conceptualization, Formal analysis, Validation, Methodology, Writing – original draft. **R. Valentini:** Conceptualization, Supervision, Funding acquisition, Investigation, Methodology. **E. Rubinacci:** Data curation, Software, Methodology, Visualization. **C. Caroli:** Supervision, Funding acquisition, Methodology. **F. Renzi:** Software, Writing – original draft. **S. Noce:** Conceptualization, Software, Formal analysis. **G. Tramonte:** Data curation, Software. **V. Scardigno:** Data curation, Software. **F. Gabbianelli:** Funding acquisition, Methodology. **V. Aloisi:** Software, Formal analysis, Writing – original draft. **D. Pietrucci:** Software, Validation. **D. Del Buono:** Software, Methodology, Writing – review & editing. **M. Santini:** Conceptualization, Supervision, Funding acquisition. **L. Conte:** Funding acquisition, Project administration. **A. Aloisio:** Project administration. **C. Dellacasa:** Funding acquisition, Project administration. **G. Scipione:** Conceptualization, Supervision, Funding acquisition. **M. Passamonti:** Validation, Investigation. **A. Rullo:** Conceptualization, Funding acquisition. **G. Chillemi:** Conceptualization, Validation, Investigation, Methodology, **V. Piccolo:** Software, Visualization. **M. Barbato:** Methodology, Writing – review & editing. **P. Nassisi:** Conceptualization, Supervision, Funding acquisition, Investigation, Methodology, Writing – original draft, Project administration, Writing – review & editing

Funding
This work was supported by the SEBASTIEN Project, co-financed by the Connecting Europe Facility of the European Union under agreement No: INEA/CEF/ICT/A2020/2373580; Action No: 2020-IT-IA-0234

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge ANAPRI for providing data used in Services 2a, 2c and 4b, IZS Sardegna for data contribution under Service 4a, and the Agritech project ("National Research Centre for Agricultural Technologies" of the National Recovery and Resilience Plan, PNRR, financed by the European Union "Next Generation EU") for its support of this work.

References

[1] Animal Task Force (ATF), "Why is European animal production important today? Facts and figures", 2017. [Online]. Available: https://animaltaskforce.eu/wp-content/uploads/2024/11/Facts-and-figures-sustainable-and-competitive-livestock-sector-in-EU_FINAL.pdf [Accessed: 23 Apr. 2026]

[2] Animal Task Force (ATF), "Research and Innovation for a Sustainable Livestock Sector in Europe", White Paper, 2013. [Online]. Available: https://animaltaskforce.eu/wp-content/uploads/2024/11/ATF-white-paper-Research-priorities-for-a-sustainable-livestock-sector-in-Europe.pdf [Accessed: 23 Apr. 2026]

[3] Animal Task Force (ATF), "A strategic research and innovation agenda for a sustainable livestock sector in Europe", Second White Paper, Short Version, 2017. [Online]. Available: https://animaltaskforce.eu/wp-content/uploads/2024/11/ATF_White_Paper_2017_Brochure.pdf [Accessed: 23 Apr. 2026]

[4] FAOSTAT, "Food and Agriculture Organization Corporate Statistical Database", 2018. [Online]. Available: http://www.fao.org/faostat/en/#data/GT [Accessed: 23 Apr. 2026]

[5] CREA, "La zootecnia in Italia Produzioni, regolamentazione, ricerca, politiche per la qualità e la biodiversità". 2017 [Online]. Available: https://rica.crea.gov.it/download.php?id=963 [Accessed 23 Apr. 2026]

[6] Eurostat, "Agriculture statistics - family farming in the EU", 2019. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Agriculture_statistics_-_family_farming_in_the_EU [Accessed: 23 Apr. 2026]

[7] European Environment Agency (EEA), "Annual European Union greenhouse gas inventories 1990-2017/18 and Inventory reports 2019/20". [Online]. Available: https://www.eea.europa.eu/en/analysis/publications/european-union-greenhouse-gas-inventory-2019 [Accessed 23 Apr. 2026]

[8] FAOSTAT, "Food and Agriculture Organization Corporate Statistical Database", 2021. [Online]. Available: http://www.fao.org/faostat/en/#data/GT [Accessed 23 Apr. 2026]

[9] U. Bernabucci, "Climate change: impact on livestock and how can we adapt", Animal Frontiers, vol. 9, no. 1, 2019, pp. 3-5. https://doi.org/10.1093/af/vfy039

[10] R. J Collier, L. H Baumgard, R. B Zimbelman, Y. Xiao, "Heat stress: physiology of acclimation and adaptation", Animal Frontiers, vol. 9, no. 1, 2019, pp. 12-19. https://doi.org/10.1093/af/vfy031

[11] M.L. Rhoads, R.P. Rhoads, M.J. VanBaale, R.J. Collier, S.R. Sanders, W.J. Weber, B.A. Crooker, L.H. Baumgard, "Effects of heat stress and plane of nutrition on lactating Holstein cows: I. Production, metabolism,

and aspects of circulating somatotropin1", Journal of Dairy Science, vol. 92, no. 5, 2009, pp 1986-1997. https://doi.org/10.3168/jds.2008-1641.

[12] D. R. Ames, L. W. Insley, "Wind-Chill Effect for Cattle and Sheep", Journal of Animal Science, vol. 40, no. 1, 1975, pp. 161-165, https://doi.org/10.2527/jas1975.401161x

[13] P. Thornton, G. Nelson, D. Mayberry, & M. Herrero. (2021), "Increases in extreme heat stress in domesticated livestock species during the twenty-first century", Global Change Biology, 27, 5762–5772. https://doi.org/10.1111/gcb.15825

[14] D.L. Smith, T. Smith, B.J. Rude, S.H. Ward, "Short communication: Comparison of the effects of heat stress on milk and component yields and somatic cell score in Holstein and Jersey cows", J Dairy Sci. 2013;96: 3028–3033. https://doi.org/10.3168/jds.2012-5737

[15] J.S. Leles, I.C.S. Rodrigues, M.F.V. Neto, A.M.V. Neto, D.R. da Rocha, A.N.L. da Costa, et al, "Heat Stress and Body Temperature in Brown Swiss Cows Raised in Semi-Arid Climate of Ceará State, Brazil", Acta Scientiae Veterinariae. 2017;45: 1–8. https://doi.org/10.22456/1679-9216.80470

[16] A. M. Lees, V. Sejian,A.L. Wallage, C.C. Steel, T.L. Mader, J.C. Lees, J. C., J.B. Gaughan, 2019, "The Impact of Heat Load on Cattle", Animals, vol. 9, no. 6, 2019, p. 322. https://doi.org/10.3390/ani9060322

[17] R. Rötter, S. Van de Geijn, "Climate Change Effects on Plant Growth, Crop Yield and Livestock", Climatic Change, vol. 43, 1999, pp. 651-681. https://doi.org/10.1023/a:1005541132734

[18] P. Niloofar, S. Lazarova-Molnar, D. A. Thumba, K. I. Shahin, "A conceptual framework for holistic assessment of decision support systems for sustainable livestock farming, Ecological Indicators", vol. 155, p. 111029, 2023. https://doi.org/10.1016/j.ecolind.2023.111029

[19] M. C. Mura, O. Trimasse, V. Carcangiu, S. Luridiana, "Precision Livestock Farming for Dairy Sheep: A Literature Review of IoT and Decision-Support Systems for Enhanced Management and Welfare, AgriEngineering", vol. 8, 2026, p. 58. https://doi.org/10.3390/agriengineering8020058

[20] S. Neethirajan, 2020. ”Transforming the Adaptation Physiology of Farm Animals through Sensors”, Animals, vol. 10, no. 9, p. 1512, 2020. https://doi.org/10.3390/ani10091512

[21] S. Wolfert, L. Ge, C. Verdouw, M. Bogaardt, ”Big Data in Smart Farming – A review”, Agricultural Systems, vol. 153, 2017, pp. 69-80. https://doi.org/10.1016/j.agsy.2017.01.023

[22] D. Berckmans, ”General introduction to precision livestock farming”, Animal Frontiers, vol. 7, no. 1, 2017, pp. 6-11. https://doi.org/10.2527/af.2017.0102

[23] S. Neethirajan, ”Recent advances in wearable sensors for animal health management, Sensing and Bio-Sensing Research”, vol. 12, 2017, pp. 15-29. https://doi.org/10.1016/j.sbsr.2016.11.004.

[24] I. Akhigbe, K. Munir, O. Akinade, L. Akanbi, L.O. Oyedele, ”IoT Technologies for Livestock Management: A Review of Present Status, Opportunities, and Future Trends”, Big Data and Cognitive Computing, vol. 5, no. 1, 2021, p. 10. https://doi.org/10.3390/bdcc5010010

[25] HerdDogg, “Smart livestock management solutions”. [Online]. Available: https://herddogg.com/ [Accessed: 25 Apr. 2026]

[26] G. Idoje, T. Dagiuklas, M. Iqbal, ”Survey for smart farming technologies: Challenges and issues”, Computers & Electrical Engineering, vol. 92, 2021, p. 107104. https://doi.org/10.1016/j.compeleceng.2021.107104.

[27] J. L. Kleen, R. Guatteo, ”Precision Livestock Farming: What Does It Contain and What Are the Perspectives?, Animals”, vol. 13, no. 5, 2023, p. 779. https://doi.org/10.3390/ani13050779

[28] M.M. Passamonti, E. Somenzi, M. Barbato, G. Chillemi, L. Colli, S. Joost, M. Milanesi, R. Negrini, M. Santini, E. Vajana, et al. ”The Quest for Genes Involved in Adaptation to Climate Change in Ruminant Livestock”, Animals. vol. 11, no. 10, 2021, p. 2833. https://doi.org/10.3390/ani11102833

[29] C.F. Kelly, T.E. Bond, 1971. "Bioclimatic factors and their measurements", in: A guide to environmental research in animals. Natl. Acad. Sci. 1971, pp. 7-92.

[30] R. Collier, L. Hall, S. Rungruang, R. Zimbelman, "Quantifying Heat Stress and Its Impact on Metabolism and Performance", Proceedings of the 24th Florida Ruminant Nutrition Symposium, Gainesville, FL, USA, 2012

[31] M. Raffa, et al. "Very high resolution projections over Italy under different CMIP5 IPCC scenarios", Scientific Data, vol. 10, 2023, p. 238. https://doi.org/10.1038/s41597-023-02144-9.

[32] M. Raffa, A. Reder, G. Marras, M. Mancini, G. Scipione, M. Santini, et al. "VHR-REA_IT dataset: very high resolution dynamical downscaling of ERA5 reanalysis over Italy by COSMO-CLM", [dataset], 2021, 6, 88. https://doi.org/10.3390/data6080088

[33] M. Adinolfi, M. Raffa, A. Reder, P. Mercogliano, "Investigation on potential and limitations of ERA5 reanalysis downscaled on Italy by a convection–permitting model", Clim. Dyn., 2023, https://doi.org/10.1007/s00382-023-06803-w

[34] S.M. Lundberg, S.I. Lee, "A Unified Approach to Interpreting Model Predictions", Advances in Neural Information Processing Systems, vol. 30, MIT Press, Cambridge, MA, USA, 2017, pp. 4766-4775. https://doi.org/10.48550/arXiv.1705.07874

[35] S.L. Cartwright, J. Schmied, N. Karrow, B.A. Mallard, "Impact of heat stress on dairy cattle and selection strategies for thermotolerance: a review", Frontiers in Veterinary Science, vol. 10, 2023. https://doi.org/10.3389/fvets.2023.1198697

[36] R. Primi, G. Filibeck, A. Amici, C. Bückle, L. Cancellieri, A. Di Filippo, C. Gentile, A. Guglielmino, R. Latini, L.D. Mancini, S.A. Mensing, C.M. Rossi, F. Rossini, A. Scoppola, C. Sulli, R. Venanzi, B. Ronchi, G. Piovesan, "From Landsat to leafhoppers: A multidisciplinary approach for sustainable stocking assessment and ecological monitoring in mountain grasslands", Agriculture, Ecosystems & Environment, vol. 234, 2016, pp. 118-133. https://doi.org/10.1016/j.agee.2016.04.028

[37] I.L. Bretas, D.S.M. Valente, F.F. Silva, et al. "Prediction of aboveground biomass and dry-matter content in Brachiaria pastures by combining meteorological data and satellite imagery", Grass Forage Science, vol. 76, 2021, pp. 340-352. https://doi.org/10.1111/gfs.12517

[38] S. Cappai, F. Loi, A. Coccollone, M. Contu, P. Capece, M. Fiori, S. Canu, C. Foxi, S. Rolesu. "Retrospective analysis of Bluetongue farm risk profile definition, based on biology, farm management practices and climatic data", Preventive Veterinary Medicine, vol. 155, 2018, pp. 75-85 https://doi.org/10.1016/j.prevetmed.2018.04.004

[39] M.N. Alhussien, A.K. Dang, "Milk somatic cells, factors influencing their release, future prospects, and practical utility in dairy animals: An overview", Veterinary World, vol. 11, no. 5, 2018, p. 562. https://doi.org/10.14202/vetworld.2018.562-577