A Predictive Analytics Architecture Integrating Machine Learning and Collaborative Filtering for Accurate Detection of Mastitis in Dairy Cows

Dr. Michael Anderson
Department of Agricultural Engineering
Faculty of Engineering and Information Technologies
The University of Melbourne
Melbourne, Victoria, Australia

Dr. Emily Thompson
Department of Agricultural and Biosystems Engineering
Faculty of Science and Engineering
The University of Western Australia
Perth, Western Australia, Australia

Received: 01 June 2026 | Received Revised Version: 15 June 2026 | Accepted: 25 June 2026 | Published: 01 July 2026

Volume 08 Issue 07 2026 |

Abstract

Mastitis remains one of the most prevalent and economically significant diseases affecting dairy cattle worldwide, resulting in substantial reductions in milk production, deterioration of milk quality, increased veterinary expenditures, premature culling, and diminished animal welfare. Although conventional diagnostic techniques such as somatic cell count analysis, bacterial culture, and clinical examination remain widely adopted, these approaches often identify the disease after physiological damage has already occurred. The increasing availability of precision livestock farming technologies, automatic milking systems, wearable sensors, and continuous physiological monitoring has created unprecedented opportunities for developing intelligent predictive systems capable of identifying mastitis before the appearance of obvious clinical symptoms. Recent advances in artificial intelligence have demonstrated considerable success in disease prediction through machine learning algorithms; however, existing systems frequently rely on isolated predictive models that inadequately exploit multidimensional relationships among cows, environmental variables, production characteristics, and historical disease patterns. This study proposes a predictive analytics architecture integrating machine learning and collaborative filtering to improve the accuracy and reliability of mastitis detection in dairy cows. The proposed conceptual architecture combines heterogeneous sensor data, historical herd records, behavioral monitoring, milk quality parameters, and collaborative similarity learning to generate individualized disease-risk assessments. Machine learning algorithms perform nonlinear pattern recognition, while collaborative filtering captures latent similarities among animals exhibiting comparable physiological and production characteristics. Together, these complementary mechanisms enhance prediction robustness and reduce false-positive and false-negative classifications. The paper synthesizes recent developments in machine learning, automated milking systems, risk assessment methodologies, ensemble learning, deep learning, and sensor-based monitoring exclusively from the provided literature. A comprehensive methodological framework is developed to explain data acquisition, preprocessing, feature engineering, collaborative similarity modeling, predictive analytics, validation strategies, and deployment considerations. The study further evaluates the theoretical advantages, implementation challenges, and practical implications of integrating collaborative intelligence into livestock disease prediction. The proposed architecture contributes a generalized decision-support framework capable of supporting veterinarians, dairy producers, and precision livestock management systems in achieving earlier intervention, improved herd health, optimized resource utilization, and enhanced sustainability within modern dairy farming.

Keywords: Mastitis prediction; Machine learning; Collaborative filtering; Precision livestock farming; Predictive analytics; Dairy cattle; Automated milking systems; Disease detection; Artificial intelligence; Sensor-based monitoring.

Cite This Article: Anderson D. M., & Thompson D. E. (2026). A Predictive Analytics Architecture Integrating Machine Learning and Collaborative Filtering for Accurate Detection of Mastitis in Dairy Cows. The American Journal of Agriculture and Biomedical Engineering, 8(07), 1–12. Retrieved from https://theamericanjournals.com/index.php/tajabe/article/view/8201

1. Introduction

1.1 Background

Mastitis represents one of the most persistent infectious diseases affecting dairy cattle and continues to impose severe economic, veterinary, and production-related consequences on dairy industries worldwide. The disease is characterized by inflammation of mammary gland tissue caused primarily by bacterial pathogens, resulting in deterioration of milk quality, reduction in milk yield, increased treatment costs, and compromised animal welfare. In severe cases, mastitis contributes to premature culling and long-term productivity losses that significantly reduce farm profitability.

The growing adoption of precision livestock farming technologies has fundamentally transformed dairy herd management by enabling continuous observation of animal health through automated monitoring systems. Modern dairy farms increasingly deploy automatic milking systems, wearable activity sensors, rumination monitors, conductivity sensors, milk analyzers, and environmental monitoring devices to collect extensive physiological and behavioral information. These technologies generate large-scale longitudinal datasets capable of supporting intelligent disease surveillance beyond traditional manual inspection methods.

Despite these technological advances, effective utilization of high-dimensional livestock data remains challenging. Conventional statistical techniques frequently assume linear relationships between variables and therefore struggle to identify the complex interactions governing disease progression. Mastitis development depends upon multiple interacting physiological, environmental, microbial, nutritional, and management-related factors whose nonlinear relationships cannot be adequately represented using traditional analytical methods.

Machine learning has consequently emerged as a powerful computational paradigm capable of discovering hidden patterns within multidimensional agricultural datasets. Supervised learning algorithms, ensemble models, recurrent neural networks, gradient boosting methods, and deep learning architectures have demonstrated promising capabilities for predicting mastitis using milk production characteristics, electrical conductivity, rumination behavior, activity measurements, and historical health records (Ebrahimi et al., 2019; Naqvi et al., 2022; Tian et al., 2024).

Nevertheless, most current prediction models analyze each cow independently and often ignore valuable similarities that naturally exist among animals within comparable production environments. Dairy cows sharing similar lactation stages, behavioral characteristics, production histories, genetic backgrounds, or environmental exposures frequently exhibit comparable disease susceptibility. Collaborative filtering techniques, extensively utilized within recommendation systems, offer an opportunity to exploit these hidden similarities for livestock disease prediction. Integrating collaborative filtering with machine learning enables prediction models to simultaneously learn individual physiological patterns and collective herd-level behavioral relationships, thereby improving predictive robustness and personalization.

This article proposes a comprehensive predictive analytics architecture that integrates machine learning algorithms with collaborative filtering principles for accurate mastitis detection. The proposed framework aims to bridge existing methodological gaps by combining individualized prediction with similarity-based knowledge transfer across comparable animals, thereby enhancing disease prediction within precision dairy farming environments.

1.2 Research Problem

Although substantial progress has been achieved in automated mastitis prediction, current predictive systems continue to encounter several important limitations.

First, many prediction models rely upon a restricted number of physiological indicators such as somatic cell counts or milk electrical conductivity. While these variables possess diagnostic value, mastitis is a multifactorial disease influenced simultaneously by milk composition, behavioral changes, environmental conditions, lactation status, feeding practices, pathogen characteristics, and herd management strategies.

Second, numerous machine learning models emphasize algorithmic optimization without adequately considering relationships among cows exhibiting similar disease progression. Consequently, valuable latent knowledge embedded within herd-wide production histories remains underutilized.

Third, practical dairy farms continuously generate heterogeneous datasets characterized by missing values, sensor noise, inconsistent sampling frequencies, and variable data quality. Robust predictive architectures must therefore incorporate preprocessing, feature engineering, similarity analysis, uncertainty management, and adaptive learning mechanisms before reliable prediction becomes feasible.

Finally, prediction accuracy alone is insufficient for operational decision-making. Dairy producers require interpretable risk scores capable of supporting timely veterinary intervention while minimizing unnecessary treatments.

These limitations demonstrate the need for an integrated predictive architecture capable of combining machine learning intelligence with collaborative behavioral similarity analysis.

1.3 Research Objectives

This review aims to accomplish several interconnected objectives.

The first objective is to examine recent developments in machine learning-based mastitis prediction using automated livestock monitoring systems.

The second objective is to synthesize existing research concerning sensor technologies, predictive algorithms, and risk assessment frameworks applied within dairy health management.

The third objective is to develop a conceptual predictive analytics architecture integrating collaborative filtering with machine learning for early mastitis detection.

The fourth objective is to evaluate the potential advantages, implementation challenges, and practical implications of deploying integrated intelligent prediction systems in commercial dairy farms.

Finally, the study seeks to identify future research opportunities supporting precision livestock farming through explainable and adaptive artificial intelligence.

1.4 Research Significance

Early detection represents one of the most effective strategies for minimizing mastitis-associated economic losses. Disease identification during subclinical stages enables prompt treatment before irreversible tissue damage occurs, thereby improving therapeutic success while reducing antimicrobial usage.

The significance of predictive analytics extends beyond disease diagnosis. Intelligent monitoring systems facilitate precision livestock farming by continuously assessing animal health, optimizing veterinary resource allocation, improving milk quality, enhancing production efficiency, and promoting sustainable agricultural practices.

The integration of collaborative filtering introduces an additional dimension of intelligence. Rather than considering each cow as an isolated observation, collaborative learning recognizes hidden similarities among animals, enabling predictive models to benefit from collective herd knowledge. Such architectures are particularly valuable in farms where historical disease occurrences remain relatively infrequent yet highly informative.

Furthermore, this study contributes conceptually by integrating two computational paradigms that have largely evolved independently within agricultural analytics. Machine learning provides sophisticated nonlinear classification capabilities, whereas collaborative filtering contributes latent similarity modeling. Their integration offers a promising direction for next-generation veterinary decision-support systems.

2. Literature Review

The rapid expansion of precision livestock farming has accelerated research into intelligent disease detection technologies, particularly for mastitis prediction. The existing literature demonstrates a progressive evolution from conventional statistical diagnosis toward advanced machine learning systems incorporating sensor fusion, automated monitoring, ensemble learning, and deep neural networks.

Early investigations primarily emphasized understanding disease prevalence and clinical management. Chen (2023), Wang and Yang (2023), Ma et al. (2023), and Ye et al. (2023) comprehensively discussed mastitis epidemiology, prevention strategies, clinical manifestations, and therapeutic interventions within Chinese dairy production systems. These studies emphasized that early diagnosis remains the most effective approach for minimizing productivity losses while improving animal welfare.

Similarly, Zhang et al. (2020) investigated bacterial isolation, pathogen identification, antimicrobial sensitivity, and pathogenic mechanisms associated with mastitis infections. Their findings demonstrated that accurate identification of causative pathogens constitutes an essential component of effective disease control. However, microbiological diagnosis generally occurs after infection has progressed, highlighting the importance of predictive methodologies capable of identifying disease risk before clinical manifestation.

The development of automatic detection technologies represents another major research direction. Chu et al. (2023) reviewed advances in automated mastitis detection technologies and emphasized the increasing importance of intelligent sensing systems capable of continuously monitoring dairy cows. Their review demonstrated that integrating multiple sensing modalities substantially improves disease surveillance compared with isolated diagnostic measurements.

The transition from traditional diagnostic procedures to intelligent predictive systems has been accelerated by the widespread adoption of precision livestock farming technologies. Modern dairy farms routinely collect large volumes of data from automatic milking systems (AMS), wearable sensors, milk analyzers, activity monitors, rumination sensors, and environmental monitoring devices. These heterogeneous data sources provide continuous observations of animal behavior and physiology, allowing predictive algorithms to detect subtle abnormalities before clinical symptoms become evident.

Ozella et al. (2023) reviewed modeling approaches developed for automatic milking systems and concluded that the increasing complexity of AMS-generated datasets requires advanced computational methods capable of handling temporal variation, nonlinear interactions, and multidimensional feature spaces. Their review further emphasized that integrating multiple sensor modalities generally improves prediction performance compared with models relying on single-variable indicators.

Several studies have demonstrated that milk electrical conductivity, milk yield, rumination behavior, feeding activity, and locomotion are among the most informative indicators for mastitis prediction. Tian et al. (2024) developed machine learning models using milk yield, rumination time, and milk electrical conductivity to predict clinical mastitis and reported that combining these variables substantially improved predictive capability compared with traditional threshold-based approaches. Likewise, Zhai et al. (2024) examined relationships among milk yield, conductivity, and activity levels, confirming that physiological and behavioral indicators jointly contribute to disease discrimination.

Machine learning algorithms have become the dominant analytical tools for transforming these complex datasets into clinically meaningful predictions. Ebrahimi et al. (2019) performed a comprehensive comparison of machine learning methods for subclinical mastitis prediction and demonstrated that deep learning architectures together with gradient-boosted tree models consistently outperformed conventional machine learning algorithms. Their work highlighted the capacity of deep neural models to learn complex nonlinear relationships from heterogeneous livestock datasets.

Similarly, Bobbo et al. (2021) compared multiple machine learning techniques using somatic cell count information to predict udder health status. Their findings suggested that ensemble learning methods improve classification robustness while reducing prediction variability. The study also emphasized the importance of selecting informative features rather than relying solely on algorithmic complexity.

Shi et al. (2021) proposed a mastitis risk assessment model for Chinese Holstein cattle, demonstrating the value of structured risk evaluation frameworks in identifying susceptible animals before disease onset. Subsequently, Li et al. (2021) optimized this risk assessment system for practical dairy farm implementation, illustrating how computational models can be translated into operational decision-support tools.

Time-series analysis has emerged as another important research direction because mastitis develops progressively rather than instantaneously. Fan et al.

(2023) developed multivariable time-series classification methods capable of analyzing continuous sensor streams generated by automatic milking systems. Their results demonstrated that longitudinal behavioral patterns provide richer predictive information than isolated observations collected at single time points.

Naqvi et al. (2022) further extended temporal modeling by employing recurrent neural networks (RNNs) for mastitis detection within automated milking systems. The recurrent architecture effectively captured sequential dependencies among physiological variables, thereby improving early disease recognition. These findings indicate that temporal learning architectures are particularly well suited for monitoring continuously evolving livestock health conditions.

Forecasting studies have similarly demonstrated the importance of proactive disease management. Bonestroo et al. (2022) developed gradient-boosting classifiers using automatic milking system sensor data to forecast chronic mastitis. Their work illustrated that predictive models trained on continuously collected sensor information can successfully identify animals at elevated disease risk before severe clinical manifestations occur. The study further emphasized the practical value of integrating predictive analytics into everyday dairy management operations, providing strong evidence for intelligent decision-support systems (Bonestroo et al., 2022).

Complementary evidence was presented by Luo et al. (2023), who constructed a clinical mastitis risk prediction model using machine learning algorithms and demonstrated that predictive systems can support earlier intervention compared with traditional observation-based management. Pakrashi et al. (2023) similarly employed cow-level features to detect subclinical mastitis, reinforcing the importance of individualized prediction rather than herd-level generalization.

Performance comparisons among modern machine learning algorithms have received increasing attention. Satola and Satola (2024) compared bagging, boosting, stacking, super-learner ensembles, and individual machine learning models. Their analysis showed that ensemble strategies consistently achieved superior predictive performance because they combined complementary strengths from multiple learning algorithms while reducing overfitting. These findings suggest that integrated predictive architectures should prioritize ensemble intelligence over isolated classifiers.

Liebe et al. (2022) addressed an equally important practical challenge: predicting relatively low-incidence diseases using individual cow data. Their study highlighted issues related to class imbalance, limited positive observations, and real-world deployment. These challenges remain highly relevant because mastitis prediction systems must maintain high sensitivity without generating excessive false-positive alerts.

Recent developments in explainable artificial intelligence have also become increasingly relevant. Zhang et al. (2023) reviewed deep contrastive learning methods capable of learning discriminative feature representations from complex datasets. Although their survey was not specific to veterinary medicine, representation learning techniques offer promising opportunities for extracting robust latent features from multidimensional dairy datasets.

Feature optimization remains another critical component of predictive modeling. Zhou X. Z. et al. (2021) demonstrated how hybrid random forest methods integrated with GeoDetector and recursive feature elimination effectively identify influential predictive variables. While their work focused on landslide susceptibility mapping, the methodological principles are directly transferable to agricultural prediction problems involving high-dimensional feature selection.

Collectively, the reviewed literature demonstrates remarkable progress in automated mastitis prediction using machine learning. Nevertheless, several important research gaps remain.

First, most predictive models analyze cows independently despite the existence of meaningful similarities among animals sharing production characteristics, physiological profiles, or environmental conditions.

Second, relatively few studies integrate collaborative intelligence capable of transferring predictive knowledge between comparable animals.

Third, although ensemble learning, recurrent neural networks, and deep learning have significantly improved classification performance, limited attention has been devoted to architectures capable of simultaneously exploiting individualized physiological patterns and herd-level relational knowledge.

Finally, most existing studies optimize algorithmic performance rather than designing comprehensive predictive architectures suitable for operational deployment across diverse dairy production environments.

These observations motivate the development of an integrated predictive analytics architecture combining machine learning with collaborative filtering. Such an approach extends beyond traditional classification by incorporating similarity-based learning into predictive decision-making, thereby improving personalization, robustness, and practical applicability. The forecasting framework proposed by Bonestroo et al. (2022) further supports the feasibility of sensor-driven predictive architectures and provides important evidence that continuous monitoring systems can substantially improve proactive mastitis management (Bonestroo et al., 2022).

3. Methodology

3.1 Research Design

This study adopts a research and review methodology that synthesizes existing literature to develop a conceptual predictive analytics architecture for mastitis detection. Rather than introducing a new experimental dataset, the methodology integrates evidence from previously published machine learning, sensor-based monitoring, automated milking systems, and mastitis risk assessment studies into a unified analytical framework.

The methodological objective is to establish an architecture capable of supporting intelligent disease prediction by combining two complementary computational paradigms:

- Machine learning for nonlinear pattern recognition.
- Collaborative filtering for similarity-based knowledge transfer among dairy cows.

The proposed framework consists of six interconnected stages that transform raw farm data into actionable disease-risk predictions.

3.2 Overall Predictive Analytics Architecture

The proposed architecture comprises the following components:

1. Data Acquisition Layer
2. Data Preprocessing Layer
3. Feature Engineering Layer
4. Collaborative Filtering Layer
5. Machine Learning Prediction Layer
6. Decision Support Layer

Each layer performs specialized computational tasks while interacting with adjacent components to establish an end-to-end intelligent prediction system.

3.3 Data Acquisition Layer

The first stage collects heterogeneous information from multiple farm information systems.

Typical data sources include:

- Automatic milking systems
- Milk yield records
- Milk electrical conductivity
- Somatic cell count
- Milk temperature
- Lactation stage
- Cow age
- Breed
- Body condition score
- Feeding behavior
- Rumination duration
- Walking activity
- Resting time
- Environmental temperature
- Humidity
- Historical mastitis records
- Veterinary treatment history

Unlike conventional prediction models relying upon only one or two indicators, the proposed architecture integrates physiological, behavioral, environmental, and historical information simultaneously.

Continuous monitoring enables temporal representation of disease progression instead of isolated observations.

3.4 Data Preprocessing

Raw livestock datasets frequently contain missing observations, sensor failures, duplicate records, and inconsistent sampling frequencies.

The preprocessing stage therefore performs:

- Missing value imputation
- Noise removal
- Outlier detection
- Timestamp synchronization
- Feature normalization
- Data standardization
- Duplicate elimination

Data quality directly influences predictive performance.

Improper preprocessing may introduce bias into machine learning algorithms and reduce generalization capability.

3.5 Feature Engineering

Feature engineering transforms raw sensor measurements into meaningful predictive variables.

Representative engineered features include:

Milk-related variables

- Daily milk yield
- Yield variation
- Electrical conductivity trend
- Milking duration
- Milk flow rate

Behavioral variables

- Rumination frequency
- Feeding duration
- Standing time
- Walking activity
- Lying behavior

Historical variables

- Previous mastitis occurrence
- Treatment frequency
- Recovery duration
- Lactation history

Environmental variables

- Temperature–humidity index
- Seasonal indicators
- Housing conditions

Feature engineering reduces dimensionality while preserving biologically meaningful information.

Recent studies consistently demonstrate that combining multiple complementary variables improves disease prediction compared with isolated measurements.

3.6 Collaborative Filtering Layer

The collaborative filtering module represents the principal innovation of the proposed architecture.

Instead of treating each cow independently, the model identifies animals exhibiting comparable production histories, physiological characteristics, behavioral patterns, and disease trajectories.

Similarity may be computed using variables such as:

- Milk production profile
- Lactation stage
- Conductivity evolution
- Activity behavior
- Rumination trend
- Previous mastitis episodes
- Recovery characteristics

Once similar cows are identified, disease information from neighboring animals contributes additional predictive evidence.

This collaborative mechanism allows the system to estimate disease probability even when direct evidence for an individual cow remains limited.

Consequently, prediction becomes more personalized while reducing uncertainty associated with sparse observations.

3.7 Machine Learning Prediction Layer

Following feature extraction and collaborative similarity analysis, the processed dataset is forwarded to the machine learning prediction layer. This layer serves as the core analytical engine of the proposed architecture by identifying complex nonlinear relationships between multidimensional variables and mastitis occurrence.

Rather than relying on a single predictive algorithm, the architecture is designed to support multiple machine learning paradigms depending on data characteristics and farm requirements. Based on the reviewed literature, the following algorithm categories are particularly suitable:

- Decision tree-based classifiers
- Random forest models
- Gradient boosting machines
- Extreme Gradient Boosting (XGBoost)
- Artificial neural networks
- Deep neural networks
- Recurrent neural networks
- Ensemble learning methods
- Stacking and bagging approaches

The literature consistently demonstrates that ensemble learning achieves greater predictive stability than individual classifiers. Satola and Satola (2024) reported that stacking, boosting, and super-learner ensembles generally outperform single machine learning models because they integrate complementary decision boundaries. Likewise, Ebrahimi et al. (2019) found that deep learning and gradient-boosted trees provide superior classification performance for subclinical mastitis detection.

Within the proposed architecture, collaborative filtering outputs are incorporated as additional predictive features rather than replacing conventional physiological variables. Consequently, each machine learning model simultaneously learns:

- Individual physiological behavior.
- Historical disease progression.
- Herd-level similarity patterns.
- Environmental influences.
- Temporal behavioral evolution.

This hybrid learning mechanism enables the predictive model to exploit information unavailable to conventional classifiers that analyze each animal independently.

3.8 Prediction Workflow

The operational workflow of the proposed architecture consists of the following sequential stages:

1. Continuous sensor data collection.
2. Data validation and preprocessing.
3. Feature engineering and dimensional transformation.
4. Collaborative similarity computation.
5. Machine learning model training.
6. Disease probability estimation.
7. Risk categorization.
8. Veterinary decision support.
9. Continuous model updating.

Unlike static prediction systems, this workflow continuously incorporates newly acquired sensor observations, allowing disease probabilities to evolve dynamically as animal conditions change.

3.9 Risk Classification Framework

The prediction output is translated into interpretable disease-risk categories that facilitate practical farm management.

An illustrative risk classification framework may include:

Predicted Risk Score Clinical Interpretation Recommended Action

0–20% Healthy Routine monitoring

21–40% Low riskIncreased observation

41–60% Moderate risk Preventive examination

61–80% High risk Veterinary inspection

81–100% Critical risk Immediate intervention

Such categorization improves interpretability and assists dairy producers in prioritizing veterinary resources.

3.10 Model Validation Strategy

Reliable prediction systems require rigorous evaluation before deployment.

The proposed architecture incorporates multiple validation procedures.

Internal Validation

Internal validation evaluates model performance using historical farm data through repeated training and testing cycles. Performance metrics include:

- Accuracy
- Precision
- Recall
- Sensitivity
- Specificity
- F1-score
- Area Under the ROC Curve (AUC)

These indicators collectively evaluate classification effectiveness under different disease prevalence conditions.

External Validation

External validation examines model generalizability across independent dairy farms.

Differences in herd genetics, management practices, feeding strategies, climate, and sensor infrastructure frequently influence predictive performance. Therefore, cross-farm validation is essential for ensuring practical applicability.

Temporal Validation

Because mastitis develops over time, temporal validation assesses predictive capability using future observations rather than randomly shuffled datasets.

This evaluation more accurately reflects real-world deployment conditions.

3.11 Decision Support Module

Prediction alone does not improve herd health unless translated into actionable recommendations.

The final architecture therefore incorporates a decision-support module that communicates prediction results through intuitive dashboards and automated alerts.

Potential outputs include:

- Individual cow risk scores.
- Daily herd health summaries.
- High-risk animal identification.
- Disease progression trends.
- Recommended veterinary inspections.
- Suggested monitoring intervals.
- Treatment prioritization.

Integration with farm management software enables automated notification systems that support rapid intervention before severe clinical mastitis develops.

Bonestroo et al. (2022) demonstrated the practical value of integrating predictive analytics with automatic milking systems, showing that continuous sensor monitoring can substantially improve chronic mastitis forecasting. Their findings reinforce the importance of embedding predictive intelligence directly within operational dairy management systems rather than treating prediction as an isolated analytical exercise (Bonestroo et al., 2022).

3.12 Advantages of the Proposed Architecture

The proposed hybrid framework offers several conceptual and operational advantages.

First, integrating collaborative filtering with machine learning enhances personalization by exploiting similarities among cows exhibiting comparable physiological profiles.

Second, multidimensional data integration improves predictive robustness because disease progression is influenced by numerous interacting variables rather than isolated measurements.

Third, continuous learning supports adaptive prediction as new farm data become available.

Fourth, ensemble-based prediction reduces classification variability and improves generalization.

Finally, decision-support integration facilitates practical implementation within precision livestock farming environments.

3.13 Potential Implementation Challenges

Despite its advantages, the proposed architecture also faces several implementation challenges.

Data quality remains a significant concern because missing sensor observations, communication failures, and inconsistent recording practices may reduce predictive reliability.

Another challenge involves computational scalability. Large commercial dairy farms generate millions of sensor observations annually, requiring efficient storage and real-time analytical capabilities.

Model interpretability also deserves consideration. While deep learning algorithms often achieve high predictive accuracy, their decision-making processes may be difficult for veterinarians and producers to interpret. Incorporating explainable artificial intelligence techniques may therefore improve user confidence.

Finally, prediction models require periodic retraining to accommodate evolving herd characteristics, management practices, and environmental conditions.

4. Results

The synthesis of the reviewed literature demonstrates a clear evolution from conventional diagnostic methods toward intelligent, data-driven mastitis prediction systems. Across the examined studies, machine learning consistently outperformed traditional statistical approaches by effectively modeling nonlinear interactions among physiological, behavioral, and environmental variables. Deep learning, gradient-boosting methods, recurrent neural networks, and ensemble classifiers emerged as the most reliable techniques for identifying subclinical and clinical mastitis before obvious clinical symptoms appeared.

A major finding of this review is that prediction accuracy increases substantially when multiple sensor-derived variables—including milk yield, electrical conductivity, rumination activity, feeding behavior, and historical disease records—are analyzed collectively rather than individually. Studies employing automated milking system data demonstrated particularly strong predictive performance because continuous monitoring captures temporal changes associated with disease progression. Forecasting approaches based on sensor data further confirmed the practical feasibility of proactive disease surveillance in commercial dairy farms (Bonestroo et al., 2022).

The proposed predictive analytics architecture extends these findings by introducing collaborative filtering as an additional analytical layer. Unlike conventional models that evaluate each cow independently, collaborative filtering leverages similarities among animals with comparable production characteristics and health histories. This integration has the potential to improve prediction in situations where individual-level observations are incomplete or insufficient.

The review also indicates that ensemble learning strategies consistently provide greater robustness than single algorithms by combining complementary predictive strengths. Moreover, comprehensive preprocessing, feature engineering, and continuous model updating were identified as essential prerequisites for reliable real-world deployment. Collectively, these findings support the development of integrated, adaptive, and interpretable decision-support systems capable of enhancing mastitis detection while improving dairy herd management.

5. Discussion

The findings demonstrate that predictive analytics has become an indispensable component of precision livestock farming. The increasing availability of sensor-generated data enables dairy producers to shift from reactive disease treatment toward preventive herd health management. Machine learning algorithms have significantly enhanced the ability to identify complex disease patterns that remain undetectable using traditional statistical analyses. Nevertheless, prediction accuracy depends not only on algorithm selection but also on data quality, feature representation, and system integration.

The proposed architecture addresses an important limitation observed throughout the literature by combining machine learning with collaborative filtering. Existing studies predominantly focus on optimizing classification algorithms, whereas relatively little attention has been devoted to exploiting similarities among cows sharing comparable physiological or behavioral characteristics. Incorporating collaborative learning enables prediction models to utilize collective herd intelligence, potentially improving personalization and reducing uncertainty.

The reviewed literature also suggests that no single machine learning algorithm performs optimally under all conditions. Instead, prediction effectiveness varies according to dataset characteristics, sensor availability, herd management practices, and disease prevalence. Consequently, flexible architectures capable of integrating multiple algorithms are likely to achieve greater long-term applicability than rigid single-model systems.

Despite these advantages, several practical considerations remain. Large-scale deployment requires reliable sensor infrastructure, standardized data collection protocols, and computational resources capable of processing continuous data streams. Furthermore, model transparency remains essential because veterinary professionals require interpretable predictions to support treatment decisions. Future research should therefore prioritize explainable artificial intelligence, federated learning across multiple farms, and adaptive retraining strategies capable of maintaining predictive accuracy under changing production conditions. Overall, integrating collaborative filtering with machine learning represents a promising advancement toward intelligent, sustainable, and economically efficient dairy health management.

6. Conclusion

Mastitis continues to represent one of the most economically significant diseases affecting dairy cattle, emphasizing the need for accurate, timely, and scalable prediction systems. This review synthesized recent advances in machine learning, automated monitoring technologies, sensor-based analytics, and mastitis risk assessment to develop a comprehensive predictive analytics architecture integrating machine learning with collaborative filtering.

The proposed framework demonstrates how heterogeneous sensor data, historical herd records, behavioral observations, and similarity-based learning can be combined into a unified decision-support system. By incorporating collaborative intelligence alongside conventional machine learning, the architecture addresses limitations associated with isolated prediction models and promotes individualized disease-risk estimation.

The review further highlights that successful implementation depends upon comprehensive preprocessing, effective feature engineering, continuous model validation, and practical decision-support integration. Ensemble learning, temporal modeling, and automated milking system data collectively provide a strong foundation for future predictive livestock management.

The conceptual architecture contributes to precision livestock farming by offering a scalable framework capable of supporting earlier intervention, improving animal welfare, enhancing milk quality, reducing unnecessary treatments, and increasing farm profitability. Future investigations should focus on validating the proposed architecture using multi-farm datasets, incorporating explainable artificial intelligence techniques, and evaluating real-time deployment within commercial dairy production systems.

References

1. Bobbo T, Biffani S, Taccioli C, Penasa M, Cassandro M. Comparison of machine learning methods to predict udder health status based on somatic cell counts in dairy cows. Sci. Rep., 2021; 11: 13642.
2. Bonestroo J, van der Voort M, Hogeveen H, Emanuelson U, Klaas I C, Fall N. Forecasting chronic mastitis using automatic milking system sensor data and gradient-boosting classifiers. Comput. Electron. Agric., 2022; 198: 107002.
3. Chen L J. Cow mastitis and scientific prevention and control measures. China Anim. Health, 2023; 25: 40–41. (in Chinese)
4. Chu M Y, Liu X W, Zeng X T, Wang Y C, Liu G. Research advances in the automatic detection technology for mastitis of dairy cows. Trans. Chin. Soc. Agric. Eng., 2023; 39(11): 1–12. (in Chinese)
5. Ebrahimi M, Mohammadi-Dehcheshmeh M, Ebrahimie E, Petrovski K R. Comprehensive analysis of machine learning models for prediction of subclinical mastitis: Deep learning and gradient-boosted trees outperform other models. Comput. Biol. Med., 2019; 114: 103456.
6. Fan X, Watters R D, Nydam D V, Virkler P D, Wieland M, Reed K F. Multivariable time series classification for clinical mastitis detection and prediction in automated milking systems. J. Dairy Sci., 2023; 106(5): 3448–3464.
7. Li W L, Zhao T T, Da R, Shi Y L, Guo G, Wang Y C, et al. Application and optimization of dairy cow mastitis risk assessment system in Chinese Holstein. Chin. J. Anim. Sci., 2021; 57(10): 65–72.
8. Liebe D M, Steele N M, Petersson-Wolfe C S, De Vries A, White R R. Practical challenges and potential approaches to predicting low-incidence diseases on farm using individual cow data: A

clinical mastitis example. J. Dairy Sci., 2022; 105(3): 2369–2379.

9. Luo W K, Dong Q, Feng Y. Risk prediction model of clinical mastitis in lactating dairy cows based on machine learning algorithms. Prev. Vet. Med., 2023; 221: 106059.
10. Ma R J, Du L, Ma W D, Zhao J H, Li Q C, Lei C Z, et al. Study on the general situation and prevention and control measures of subclinical mastitis. China Cattle Sci., 2023; 49(4): 47–50. (in Chinese)
11. Naqvi S A, King M T M, Matson R D, Devries T J, Deardon R, Barkema H W. Mastitis detection with recurrent neural networks in farms using automated milking systems. Comput. Electron. Agric., 2022; 192: 106618.
12. Ozella L, Brotto R K, Forte C, Giacobini M. A literature review of modeling approaches applied to data collected in automatic milking systems. Animals, 2023; 13(12): 1916.
13. Pakrashi A, Ryan C, Guéret C, Berry D P, Corcoran M, Keane M T, et al. Early detection of subclinical mastitis in lactating dairy cows using cow-level features. J. Dairy Sci., 2023; 106(7): 4978–4990.
14. Satola A, Satola K. Performance comparison of machine learning models used for predicting subclinical mastitis in dairy cows: bagging, boosting, stacking, and super-learner ensembles versus single machine learning models. J. Dairy Sci., 2024; 107(6): 3959–3972.
15. Shi Y L, Li W L, Tang Y J, Mi S Y, Xiao W, Liu L, et al. Studies on risk-assessment-model establishment and prediction of mastitis in Chinese Holstein cattle. Chin. J. Anim. Sci., 2021; 57(3): 84–90. (in Chinese)
16. Sun Y, Zhou G Y, Wu T B, Li Y L, Ji S Q, Zhang T. Recent research progress of cow mastitis in China. China Dairy, 2022(4): 43–51. (in Chinese)
17. Tian H, Zhou X J, Wang H, Xu C, Zhao Z X, Xu W, et al. The prediction of clinical mastitis in dairy cows based on milk yield, rumination time, and milk electrical conductivity using machine learning algorithms. Animals, 2024; 14(3): 427.
18. Wang A H, Yang L F. Causes, clinical symptoms, diagnosis and treatment of cow mastitis. Mod. Anim. Husb. Sci. Technol., 2023(10): 94–96. (in Chinese)
19. Ye W, Ma Z, Yu Y, Han B. Incidence status of mastitis in dairy cows and its prevention and treatment measures in China. Chin. J. Anim. Sci., 2023; 59(9): 343–348. (in Chinese)
20. Zhai Y, Zhou B, Zhou F Z, Dai X, Liang Y, Zhang H R, et al. Analysis of factors affecting milk yield, conductivity, and activity level in Holstein cows. Chin. J. Anim. Sci., 2024; 60(6): 148–153. (in Chinese)
21. Zhang C S, Chen J, Li Q L, Deng B Q, Wang J, Chen C G. Deep contrastive learning: A survey. Acta Autom. Sin., 2023; 49(1): 15–39.
22. Zhang Y, Shi Q, Zhou Q M, Feng W Y, Xu X, Wu X. Isolation, identification, drug sensitivity and pathogenicity of pathogenic bacteria in dairy cow mastitis. Heilongjiang Anim. Sci. Vet. Med., 2020; (23): 85–88, 167–168. (in Chinese)
23. Zhou X J, Xu C, Wang H, Xu W, Zhao Z X, Chen M X, et al. The early prediction of common disorders in dairy cows monitored by automatic systems with machine learning algorithms. Animals, 2022; 12(10): 1251.
24. Zhou X Z, Wen H J, Zhang Y L, Xu J H, Zhang W G. Landslide susceptibility mapping using hybrid random forest with Geo Detector and RFE for factor optimization. Geosci. Front., 2021; 12(5): 101211.