

Review

Applications of Machine Learning in Livestock Production and Genetic Improvement

Vishakha Uttam

ICAR-National Bureau of Animal Genetic Resources, Karnal - 132001, Haryana, India

ABSTRACT

Machine learning (ML) is increasingly being explored in livestock research because of its ability to analyse large and complex genomic, phenotypic, environmental and image-derived datasets. This review provides a concise overview of recent applications of ML in livestock, with emphasis on genomic prediction, genetic improvement, production prediction, high-throughput phenotyping and precision livestock farming. Recent studies indicate that approaches such as random forest, support vector machines, neural networks and deep learning can improve prediction or automate phenotypic measurements for selected traits and applications. However, ML does not consistently outperform established genomic prediction methods such as GBLUP and Bayesian approaches, with performance varying across traits, populations and datasets. Major challenges include limited and heterogeneous datasets, model generalizability, computational requirements and the interpretability of complex models. Future developments should focus on standardized datasets, explainable AI, independent validation and integration of genomic, phenotypic, environmental and multi-omics information. Overall, ML offers promising opportunities to complement conventional livestock breeding and support more efficient, data-driven genetic improvement and farm management.

Keywords: Genomic prediction, Livestock genomics, Machine learning**Corresponding author:** vishakhauttam13@gmail.com

Received: August 22, 2026

Accepted: August 28, 2026

INTRODUCTION

Over the past decade, biological research has generated increasingly large and complex datasets containing molecular and genetic information, including gene, protein and metabolite abundance, microbiome composition and population-level genetic variation. These datasets provide valuable opportunities for understanding biological systems and complex traits; however, their effective utilization requires analytical approaches capable of handling large and multidimensional datasets. Machine learning (ML), a branch of computer science, enables computers to identify patterns and relationships within data using mathematical and statistical approaches. Its primary objective is to develop predictive models by learning relationships among variables within a dataset (Camacho *et al.*, 2018). Depending on the nature of the data, trained models can be used to predict binary, categorical or continuous outcomes. For example, in gene identification and annotation, ML algorithms can learn characteristic features from previously annotated genomes, including transcription start sites, GC content and codon usage, and subsequently use these patterns to identify genes in newly analysed genomic sequences. The increasing availability of large-scale biological datasets has also created new opportunities for

livestock research. Conventional livestock breeding has traditionally relied on pedigree information, phenotypic records and statistical approaches such as best linear unbiased prediction (BLUP). Although these methods have provided the foundation for genetic evaluation, their application can be challenging for traits with low heritability or complex genetic architecture and may require extensive phenotypic recording over several generations (Koivula *et al.*, 2012). The development of genomic selection subsequently improved the efficiency of livestock breeding by incorporating genome-wide marker information into genetic evaluation. However, the increasing volume and complexity of genomic, phenotypic and environmental data have created a further need for flexible analytical approaches capable of extracting complex patterns from these datasets.

In this context, artificial intelligence (AI) and ML are being increasingly explored in livestock research. These approaches can analyse large-scale genetic, phenotypic and environmental datasets and may capture relationships that are difficult to model using conventional statistical approaches. Their applications include genomic prediction, identification of genomic regions associated with economically important traits, production prediction, automated phenotyping and animal monitoring (Jeste *et al.*, 2020; Fuentes

et al., 2022; Hua *et al.*, 2023). The integration of AI with imaging, sensing and other high-throughput technologies is further expanding its applications in animal breeding and precision livestock farming (Wang *et al.*, 2023).

Therefore, this review aims to provide a concise overview of recent applications of machine learning in livestock, with particular emphasis on genomic prediction, production traits, high-throughput phenotyping and precision livestock farming, while highlighting current challenges and future prospects for its application in livestock genetic improvement.

Machine learning approaches used in livestock

Machine learning (ML) provides a range of computational approaches for analysing large and complex biological datasets, identifying patterns and making predictions. In a typical ML workflow, the available data are divided into training and testing datasets. The training dataset is used to develop the model through steps such as data preprocessing, feature selection or extraction and model fitting, whereas the test dataset is used to evaluate how well the trained model performs on previously unseen observations. Model performance is then assessed by comparing predicted and observed outcomes using appropriate evaluation measures.

Based on the type of information available and the way the model learns, ML approaches are broadly classified into supervised, unsupervised and reinforcement learning (Sarker *et al.*, 2021; Shabtay *et al.*, 2021). Supervised learning uses labelled data to learn the relationship between input variables and a predefined outcome. It is mainly used for regression, when the outcome is continuous, and classification, when observations are assigned to predefined categories.

Common approaches include linear and logistic regression, decision trees, support vector machines (SVM), random forest (RF), boosting methods and artificial neural networks (ANNs). In livestock research, these methods can be used for prediction and classification of production, reproduction, health and genomic traits. Model performance is generally evaluated using a loss or error function that quantifies the difference between predicted and observed values (Crisci *et al.*, 2012).

In contrast, unsupervised learning is applied when observations do not have a predefined target variable. Instead, the objective is to identify underlying structures, patterns or groups within the data. Clustering methods such as K-means and hierarchical clustering can be used to identify groups of genetically or phenotypically similar animals, while dimensionality-reduction approaches such as principal component analysis (PCA) can simplify high-dimensional datasets while retaining their major patterns (Sharma and Kumar, 2017). Such approaches are particularly relevant to livestock genomic datasets containing large numbers of genetic markers.

Reinforcement learning differs from both supervised and unsupervised learning because the model learns through interaction with an environment. An agent selects actions and receives rewards or penalties according to the resulting outcomes, allowing it to progressively learn strategies that maximize cumulative rewards (Kaelbling *et al.*, 1996; Pandey *et al.*, 2019). Although reinforcement learning has been less widely applied in livestock genomic prediction than supervised approaches, it has potential for decision-making and automation in areas such as animal management, navigation and precision livestock farming.

Table 1: Summary of commonly used ML approaches, their basic principles and their potential applications in livestock.

Learning type	Approach	Main principle	Major livestock application
Supervised	Linear regression	Predicts a continuous outcome from one or more explanatory variables	Prediction of quantitative traits
Supervised	Logistic regression	Estimates the probability of a categorical/binary outcome	Disease/trait classification
Supervised	Decision tree	Recursively divides data into groups based on feature values	Classification and regression
Supervised	Support Vector Machine (SVM)	Finds an optimal boundary separating different classes; kernels allow nonlinear relationships	Classification and regression
Supervised	K-Nearest Neighbors (KNN)	Predicts an outcome based on the most similar observations	Classification and regression
Supervised	Random Forest	Combines predictions from multiple decision trees to improve robustness	Classification and regression

Supervised	Bagging	Combines multiple models trained on bootstrap samples to improve prediction stability	Classification and regression
Supervised	Boosting / AdaBoost	Sequentially combines weak learners to produce a stronger predictor	Prediction and classification
Supervised	Artificial Neural Network (ANN)	Learns complex relationships through interconnected layers of neurons	Classification and regression
Unsupervised	K-means clustering	Groups observations according to similarity	Population/group clustering
Unsupervised	Principal Component Analysis (PCA)	Reduces dimensionality while retaining major patterns in the data	High-dimensional genomic data
Unsupervised	Hierarchical clustering	Builds groups based on similarity between observations	Pattern and population structure
Unsupervised	Hidden Markov Model (HMM)	Models relationships between sequential observations and hidden states	Sequence-related pattern detection
Deep learning	Multilayer Perceptron (MLP)	Uses multiple fully connected neural-network layers to learn nonlinear relationships	Prediction and classification
Deep learning	Recurrent Neural Network (RNN)	Uses previous information to learn patterns in sequential data	DNA sequences and temporal data
Deep learning	Convolutional Neural Network (CNN)	Automatically extracts relevant features using convolutional layers	Images, SNP/genomic and multi-dimensional data
Genomic prediction	RKHS	Uses kernel functions to model complex nonlinear relationships among genomic markers	Genomic selection/prediction

Among these approaches, supervised learning methods are particularly relevant to livestock breeding because many breeding objectives involve predicting continuous phenotypes or genetic merit from genomic and phenotypic information. Random forest, SVM, boosting algorithms and neural networks have therefore received increasing attention for genomic prediction and trait prediction.

Deep learning (DL) represents an advanced class of ML based primarily on artificial neural networks with multiple interconnected layers. Unlike many conventional ML approaches that require greater reliance on predefined features, deep neural networks can learn increasingly complex representations from the input data through successive layers (Angermueller *et al.*, 2016; Ching *et al.*, 2018). This characteristic makes DL particularly attractive for biological datasets that are large, high-dimensional or difficult to represent using manually selected features.

Different DL architectures are suited to different types of livestock data. Multilayer perceptrons (MLPs) can be used for prediction and classification of structured numerical data, whereas recurrent neural networks (RNNs) are designed to capture relationships in sequential or time-dependent data. Convolutional neural networks (CNNs) are particularly effective for image-based applications because they can automatically extract relevant spatial

features from images. Consequently, CNNs and related architectures are increasingly being explored for animal identification, body measurement, behaviour monitoring and automated phenotyping, while neural-network-based models are also being investigated for genomic prediction and multi-omics data analysis (Angermueller *et al.*, 2016; Ching *et al.*, 2018; Camacho *et al.*, 2018).

Overall, the choice of ML approach depends on the nature of the livestock data, the prediction objective and the complexity of the relationship being modelled. Supervised methods are particularly relevant for genomic and trait prediction, unsupervised approaches are useful for exploring population structure and high-dimensional datasets, while deep-learning approaches offer considerable potential for image-based and multimodal phenotyping. The application of these approaches to livestock genomics, production, phenotyping and precision livestock farming is discussed in the following section.

Applications in livestock genomics and breeding

Recent applications of ML in livestock have expanded from genomic prediction to production prediction, automated phenotyping and precision livestock farming. In genomic selection, ML approaches such as random forest (RF), support vector regression (SVR), gradient boosting and neural networks have been investigated

as alternatives or complements to established methods such as genomic best linear unbiased prediction (GBLUP) and Bayesian approaches. Studies conducted across beef cattle, broilers and Holstein cattle indicate that the performance of ML models varies considerably among traits and populations. While conventional genomic prediction methods remained superior for

several traits, selected ML models showed advantages for particular traits, suggesting that the usefulness of ML is context-dependent rather than universally superior to established approaches (Chafai *et al.*, 2023). Recent studies representing these applications are summarized in Table 2.

Table 2: Recent applications of machine learning and deep learning in livestock genomics, production prediction, phenotyping and precision livestock farming.

Study	Species	Data	ML approach	Application	Major finding
Hay <i>et al.</i> , 2024	Beef cattle	Genomic and phenotype	RF, SVR, CNN, MLP	Genomic prediction	GBLUP performed better for some traits; RF performed best for weaning weight
Liu <i>et al.</i> , 2025	Broilers	SNP and phenotype	SVR, RF, GBDT, XGBoost, etc.	Genomic prediction	ML advantageous for selected carcass traits, but conventional methods better for several traits
Vergani <i>et al.</i> , 2024	Holstein cattle	Genomic and production/ environment	Neural network	Milk yield prediction	Genomic breeding values + ML improved prediction
Zheng <i>et al.</i> , 2025	Holstein cattle	122,672 SNPs and EBVs	DPA-net, SVR, KRR	Genomic prediction	Some ML models improved accuracy for selected traits
Yang <i>et al.</i> , 2025	Simulated livestock	Genomic and causal variants	RF, SVR, GBLUP	Genomic prediction	Causal information can improve prediction, but performance varies by model
Wang <i>et al.</i> , 2025	Cattle	Images	Deep metric learning	Individual identification	Very high identification accuracy
Nilchuen <i>et al.</i> , 2025	Brahman cattle	12,660 images	YOLOv11/deep learning	Body measurement & weight prediction	Image-based phenotyping showed strong potential
Sharma <i>et al.</i> , 2025	Cattle	Depth images/ point clouds	Deep learning	Individual identification	Breed-independent automated identification

The evidence summarized in Table 2 indicates that genomic prediction remains a major application of ML in livestock breeding. However, recent studies do not support ML as a universal replacement for conventional approaches such as GBLUP and Bayesian methods. In beef cattle, broilers and Holstein cattle, the relative performance of ML varied among traits, with conventional methods remaining competitive or superior for several traits, while selected ML models performed better for specific traits (Hay *et al.*, 2024; Liu *et al.*, 2025; Zheng *et al.*, 2025). Thus, the usefulness of ML appears to depend on the trait, population, genomic information and model used, supporting its application as a complement to established genomic prediction approaches.

Beyond genomic prediction, ML is being increasingly applied to production prediction and high-throughput phenotyping. Integration of genomic breeding values

with production and environmental information has been used to improve individual milk-yield prediction, while deep-learning-based image analysis has enabled automated body measurement and weight estimation in cattle (Vergani *et al.*, 2024; Nilchuen *et al.*, 2025). Such approaches can reduce manual measurements and facilitate the collection of large-scale phenotypic records relevant to livestock evaluation and selection.

Another emerging application is precision livestock farming, particularly automated animal identification and monitoring. Deep-learning approaches based on coat patterns and depth images have demonstrated high accuracy for individual cattle identification (Wang *et al.*, 2025; Sharma *et al.*, 2025). Integration of these automated phenotypic records with genomic and production information could support more comprehensive animal evaluation and data-driven breeding and management decisions. Overall, ML

applications in livestock extend from genomic prediction to automated phenotyping and precision management. Their future value will depend on the ability to integrate diverse data sources while maintaining model accuracy, robustness, interpretability and validation across independent populations.

Challenges and future outlook

Despite the growing potential of AI and machine learning in livestock research, several challenges remain. Their performance depends heavily on the availability of large, well-annotated and high-quality datasets, which are often limited, costly, or affected by farm-specific variation. Differences among breeds, environments, sensors and management systems can further reduce model generalizability. In addition, complex deep-learning models are often difficult to interpret, creating a “black box” problem that can limit their biological relevance and practical acceptance. High computational requirements and difficulties in field deployment also remain important constraints (Camacho *et al.*, 2018; Chafai *et al.*, 2023).

Future efforts should focus on developing standardized and diverse datasets, integrating multi-omics and phenotypic information, and improving model transferability across breeds and production environments. Explainable AI approaches can make model predictions more interpretable, while lightweight models and edge-computing solutions may facilitate on-farm applications. Combining AI with genomics, transcriptomics and other biological data could further improve understanding of gene–environment interactions and support more accurate, sustainable and efficient livestock breeding.

REFERENCES

- Angermueller C, Pärnamaa T, Parts L, and Stegle O. 2016. Deep learning for computational biology. *Mol Syst Biol*, 12(7).
- Camacho DM, Collins KM, Powers RK, Costello JC, and Collins JJ. 2018. Next-generation machine learning for biological networks. *Cell*, 173(7): 1581–1592.
- Chafai N, Hayah I, Houaga I, and Badaoui B. 2023. A review of machine learning models applied to genomic prediction in animal breeding. *Front Genet*, 14: 1150596.
- Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, and Greene CS. 2018. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*, 15(141): 20170387.
- Crisci C, Ghattas B, and Perera G. 2012. A review of supervised machine learning algorithms and their applications to ecological data. *Ecol Modell*, 240: 113–122.
- Fuentes S, Viejo CG, Tongson E, and Dunshea FR. 2022. The livestock farming digital transformation: implementation of new and emerging technologies using artificial intelligence. *Anim Health Res Rev*, 23(1): 59–71.
- Hay EH. 2024. Machine learning for the genomic prediction of growth traits in a composite beef cattle population. *Animals*, 14(20): 3014.
- Hua H, Zou S, Ma Z, Guo W, Fong CY, and Khoo BL. 2023. A deformability-based biochip for precise label-free stratification of metastatic subtypes using deep learning. *Microsyst Nanoeng*, 9(1): 120.
- Jeste DV, Graham SA, Nguyen TT, Depp CA, Lee EE, and Kim HC. 2020. Beyond artificial intelligence: exploring artificial wisdom. *Int Psychogeriatr*, 32(8): 993–1001.
- Kaelbling LP, Littman ML, and Moore AW. 1996. Reinforcement learning: A survey. *J Artif Intell Res*, 4: 237–285.
- Koivula M, Strandén I, Su G, and Mäntysaari EA. 2012. Different methods to calculate genomic predictions—Comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). *J Dairy Sci*, 95(7): 4065–4073.
- Libbrecht MW and Noble WS. 2015. Machine learning applications in genetics and genomics. *Nat Rev Genet*, 16(6): 321–332.
- Liu B, Liu H, Tu J, Xiao J, Yang J, He X, and Zhang H. 2025. An investigation of machine learning methods applied to genomic prediction in yellow-feathered broilers. *Poult Sci*, 104(1): 104489.
- McCulloch WS and Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys*, 5(4): 115–133.
- Nilchuen P, Suwanasopee T, and Koonawootrittriron S. 2025. Integrating deep learning and mobile imaging for assessment of automated conformational indices and weight prediction in Brahman cattle. *Smart Agric Technol*, 12: 101079.
- Pandey D, Niwaria K, and Chourasia B. 2019. Machine learning algorithms: a review. *Mach Learn*, 6(2): 916–922.
- Sarker IH. 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput Sci*, 2: 160.
- Shabtay L, Fournier-Viger P, Yaari R, and Dattner I. 2021. A guided FP-Growth algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data. *Inf Sci*, 553: 353–375.
- Sharma A, Randewich L, Andrew W, Hannuna S, Campbell N, Mullan S, and Burghardt T. 2025. Universal bovine identification via depth data and deep metric learning. *Comput Electron Agric*, 229: 109657.

- Sharma D and Kumar N. 2017. A review on machine learning algorithms, tasks and applications. *Int J Adv Res Comput Eng Technol*, 6(10): 2278–1323.
- Vergani AM, Bagnato A, and Masseroli M. 2024. Predicting bovine daily milk yield by leveraging genomic breeding values. *Comput Electron Agric*, 219: 108777.
- Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, and Zitnik M. 2023. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972): 47–60.
- Wang Y, Mùcher S, Wang K, Wang W, and Kooistra L. 2025. Deep metric learning for individual cattle identification using coat patterns: Proposal for a best practice. *Comput Electron Agric*, 238: 110754.
- Yang J, Calus MP, Wientjes YC, Meuwissen TH, and Duenk P. 2025. Incorporating information of causal variants in genomic prediction using GBLUP or machine learning models in a simulated livestock population. *J Anim Sci Biotechnol*, 16(1): 118.
- Zheng W, Zhang Q, He J, Han B, Zhang Q, and Sun D. 2025. Comparative evaluation of SNP-weighted, Bayesian, and machine learning models for genomic prediction in Holstein cattle. *BMC Genomics*, 26(1): 1037.