# An Introduction to Artificial Intelligence and Machine Learning for Veterinary Professionals

Christopher Pinard, DVM, DVSc[a,b,c,d,*]

**KEYWORDS**

• Artificial intelligence • Machine learning • Deep learning • Veterinary medicine
• Computer vision • Natural language processing • Clinical decision support
• Neural networks

**KEY POINTS**

- Artificial intelligence (AI) encompasses a family of computational methods that enable machines to perform tasks requiring human intelligence; machine learning (ML) is a subfield of AI.
- Supervised learning algorithms, trained on labeled datasets, underpin most current veterinary AI applications, including radiograph interpretation, cytology classification, and predictive diagnostics.
- Deep learning, a subset of ML using multilayered neural networks, has demonstrated performance comparable to board-certified specialists for specific image-based diagnostic tasks in veterinary patients.
- Adoption of AI in practice requires teams to understand core concepts such as data quality, validation, overfitting, and the distinction between classification and regression tasks.
- Ethical, regulatory, and practical considerations—including data availability, algorithmic bias, and interpretability—remain important challenges as AI tools are increasingly integrated into small animal practice.

## INTRODUCTION

Artificial intelligence (AI) has transitioned from a theoretic curiosity to a practical tool that is reshaping how medicine is practiced across species. In human health care,

[a] Department of Oncology, Toronto Animal Cancer Centre, Toronto, Ontario, Canada; [b] ANI.ML Research, ANI.ML Health Inc., Toronto, Ontario, Canada; [c] Department of Small Animal Clinical Sciences, University of Saskatoon, Saskatoon, Saskatchewan, Canada; [d] Centre for Advancing Responsible and Ethical Artificial Intelligence (CARE-AI), University of Guelph, Guelph, Ontario, Canada
* Department of Oncology, Toronto Animal Cancer Centre, 11 Maitland Street, Toronto, Ontario, M4Y 0J4.
*E-mail address:* christopher.pinard@torontoacc.com

Vet Clin Small Anim 56 (2026) 1023–1038
https://doi.org/10.1016/j.cvsm.2026.04.001 **vetsmall.theclinics.com**
0195-5616/26/© 2026 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

| Abbreviations | |
|---|---|
| AI | artificial intelligence |
| ANNs | artificial neural networks |
| AUC-ROC | area under the receiver operating characteristic curve |
| CNNs | convolutional neural networks |
| DL | deep learning |
| LLMs | large language models |
| ML | machine learning |
| NLP | natural language processing |
| RL | reinforcement learning |
| SVMs | support vector machines |

AI applications now assist with tasks ranging from interpreting medical images to predicting patient outcomes and streamlining clinical workflows.[1–3] Veterinary medicine, while somewhat earlier in its adoption trajectory, is following a similar path.[4–9] Commercial AI products for veterinary radiograph interpretation are already in clinical use,[4,8,10,11] machine learning (ML) algorithms are being applied to cytology and histopathology,[12] and large language models (LLMs) are being explored as tools for clinical documentation and client communication.[13–15]

Despite this growing presence, many veterinary professionals have had limited exposure to the foundational concepts that underpin AI and ML. Without this conceptual foundation, it can be difficult to critically evaluate the claims made about AI-based products, understand the strengths and limitations of these tools, or contribute meaningfully to their development and validation. This challenge is compounded by the rapid pace of innovation in the field and the often-opaque nature of how AI systems reach their conclusions.

The purpose of this article is to provide veterinary professionals with an accessible, clinically oriented introduction to AI and ML. Rather than a comprehensive technical review, the goal is to equip the reader with the vocabulary, conceptual frameworks, and practical knowledge needed to engage with the AI tools and research that are increasingly relevant to small animal practice. Subsequent articles in this issue will address specific applications in greater detail, including radiograph interpretation, cytology, LLMs, clinical decision support, and the ethical dimensions of AI in veterinary medicine.

## DEFINING ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, AND DEEP LEARNING

The terms AI, ML, and deep learning (DL) are frequently used interchangeably in popular discourse, but they refer to distinct, hierarchically nested concepts. Understanding these distinctions is essential for navigating the veterinary AI literature.

### *Artificial Intelligence*

AI is the broadest of these terms and refers to any computational system designed to perform tasks that would typically require human intelligence. These tasks include visual perception, speech recognition, decision-making, language translation, and problem-solving. AI encompasses a wide range of approaches, from simple rule-based expert systems (eg, a decision tree algorithm that recommends diagnostic tests based on clinical signs) to the most sophisticated neural networks capable of interpreting complex imaging data.

AI as a formal academic discipline originated in the 1950s, but practical, clinically useful applications have only become feasible in recent decades owing to increases in computational power, the availability of large digital datasets, and advances in

algorithmic design. In veterinary medicine, AI interest has accelerated substantially since 2020, with increasing numbers of publications, commercial products, and research initiatives dedicated to the field.[4–9,16]

### *Machine Learning*

ML is a subfield of AI in which algorithms learn patterns from data rather than following explicitly programmed rules. Instead of a programmer specifying every decision criterion, an ML algorithm is provided with a dataset and uses statistical methods to identify relationships between inputs (features) and outputs (labels or targets). Once trained, the algorithm can apply these learned relationships to new, previously unseen data to make predictions or classifications.

A simple veterinary analogy may be helpful. Consider how a veterinary student learns to identify heart murmurs. Rather than being given a single rigid set of rules, the student listens to many hearts—some normal, some abnormal—and gradually learns to distinguish patterns in the sounds they hear. Similarly, an ML algorithm is exposed to many examples (training data), adjusts its internal parameters to minimize errors, and over time develops the ability to classify new cases. The critical difference is that the ML algorithm performs this learning process mathematically, often identifying patterns that are too subtle or complex for human perception.

### *DeepLearning*

DL is a specialized subset of ML that uses artificial neural networks (computational systems modeled loosely on biological neurons, in which layers of interconnected nodes process and transform input data) with multiple layers (hence "deep") to learn hierarchical representations of data.[17,18] Each layer in a deep neural network transforms the data at increasing levels of abstraction. In an image classification task, for example, the first layer might detect edges and simple textures, intermediate layers might combine these into shapes and patterns, and the final layers might represent high-level concepts such as "tumor" or "normal tissue."

DL has been particularly transformative for tasks involving unstructured data. Unstructured data are data that are lacking a predefined organizational structure—in contrast, to structured data where it is organized properly into predefined columns and rows (though this can be any form of data such as images, recordings, and text). Convolutional neural networks (CNNs), a type of DL architecture, have become the dominant approach for medical image analysis[19–21] and now form the basis of most AI tools for radiograph and histopathology interpretation in both human and veterinary medicine.[10,17,22,23] The performance gains offered by DL come at the cost of increased computational requirements and reduced interpretability—a tradeoff that has important implications for clinical adoption, as discussed later in this article.

### *Generative Artificial Intelligence*

Generative AI refers to a class of DL models capable of synthesizing new content—such as text, images, or audio—that resembles the patterns learned from training data. Unlike discriminative models, which learn to classify or predict based on existing inputs, generative models produce novel outputs. The most widely recognized generative architectures include transformer-based LLMs such as generative pre-trained transformer (GPT), generative adversarial networks, and diffusion models. In veterinary medicine, generative AI applications include automated clinical documentation and discharge summary generation, client communication support, medical image synthesis for augmenting training datasets, and literature review assistance. The hierarchical relationship among AI, ML, DL, and generative AI is illustrated in **Fig. 1**.

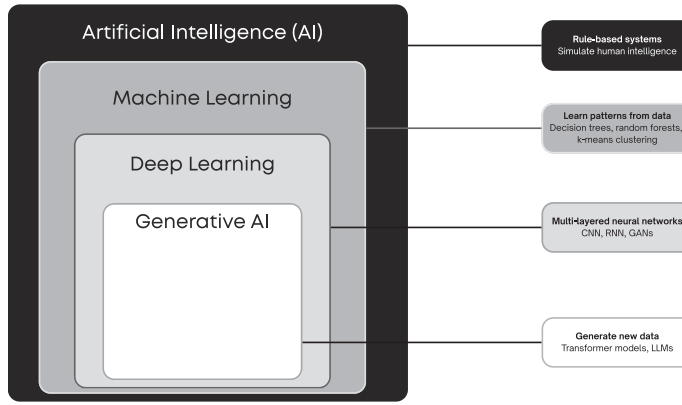



**Fig. 1.** The hierarchical relationship between artificial intelligence (AI), machine learning (ML), deep learning (DL), and generative AI. AI is the broadest category, encompassing all computational approaches that simulate human intelligence. ML is a subset of AI in which algorithms learn patterns from data. DL is a subset of ML that uses multilayered neural networks for automated feature extraction and complex pattern recognition. Generative AI is a subset of deep learning that uses models such as transformers and generative adversarial networks to synthesize new data (eg, images, text, or audio) that resemble the training distribution.

## LEARNING PARADIGMS IN MACHINE LEARNING

ML algorithms can be broadly categorized by how they learn from data. The 3 major paradigms are supervised learning, unsupervised learning, and reinforcement learning (RL). Each has distinct strengths, limitations, and applications in veterinary medicine (Table 1).

### *Supervised Learning*

Supervised learning is the most common ML paradigm in veterinary applications. In supervised learning, the algorithm is trained on a labeled dataset—one in which each input example is paired with a known, correct output. The algorithm's objective

**Table 1**
**Overview of machine learning paradigms and their veterinary applications**

| Paradigm | Data Requirements | Common Tasks | Veterinary Examples |
|---|---|---|---|
| Supervised learning | Labeled dataset (input–output pairs) | Classification and regression | Radiograph interpretation, tumor grading, survival prediction, and metastasis prediction |
| Unsupervised learning | Unlabeled dataset | Clustering, dimensionality reduction, and association | Disease phenotyping, multidrug resistance pattern detection, and patient stratification |
| Reinforcement learning | Environment with reward/penalty signals | Sequential decision-making and policy optimization | Treatment optimization (theoretic), adaptive therapy planning, and anesthesia protocol optimization |

is to learn a mapping function from inputs to outputs so that it can accurately predict the output for new, unseen inputs (Fig. 2).

Supervised learning tasks fall into 2 main categories. In classification tasks, the algorithm assigns inputs to discrete categories. For example, classifying a radiograph as showing cardiogenic pulmonary edema or not, or classifying a cytology sample as neoplastic versus inflammatory, is classification problems.[23] In regression tasks, the algorithm predicts a continuous numerical value. Predicting a patient's survival time based on clinical and imaging features is a regression problem.

The quality and representativeness of the labeled training data are paramount in supervised learning. If the training data are biased (for instance, overrepresenting certain breeds or underrepresenting atypical disease presentations), the algorithm's performance on real-world cases will suffer. Obtaining large, accurately labeled veterinary datasets remains one of the major practical challenges in the field, particularly given the smaller patient volumes and less standardized data collection practices compared to human medicine.[7]

### *Unsupervised Learning*

In unsupervised learning, the algorithm receives input data without any corresponding labels and must discover structure or patterns on its own. The most common unsupervised learning task is clustering, in which the algorithm groups similar data points together. Other unsupervised tasks include dimensionality reduction (simplifying complex, high-dimensional data while retaining its essential structure) and association (identifying relationships between variables).

Veterinary applications of unsupervised learning include identifying phenotypic subgroups within disease populations, discovering patterns in multidrug resistance, and grouping patients by clinical presentation for treatment stratification. For example, clustering algorithms applied to electronic health record data could reveal previously unrecognized disease subtypes or comorbidity patterns.[6,24] While unsupervised

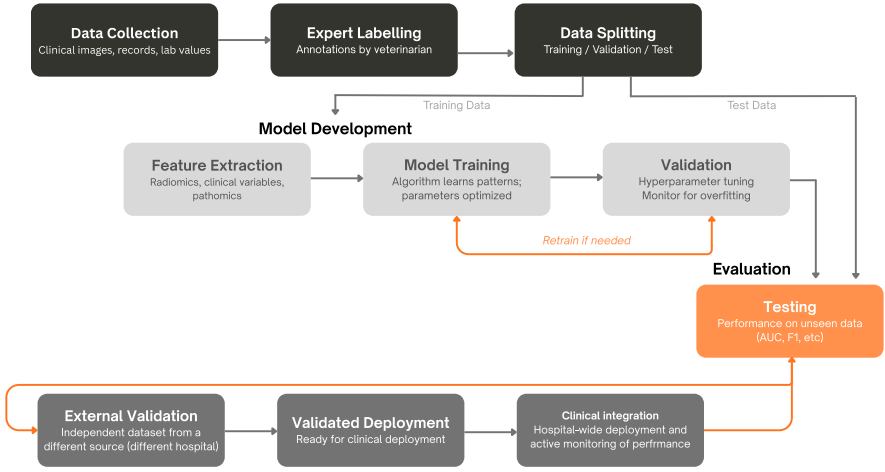



**Fig. 2.** Overview of the supervised machine learning workflow. Data are collected and labeled, divided into training and test sets, and used to train a model. The trained model is evaluated on held-out test data and, once validated, applied to new clinical cases for prediction or classification. This is then monitored over time and repeatedly tested as more data comes in.

methods receive less attention than supervised learning in the clinical literature, they play an important role in exploratory data analysis and hypothesis generation.

### *Reinforcement Learning*

RL is a paradigm in which an agent learns to make sequential decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. The agent's goal is to learn a strategy (or policy) that maximizes cumulative reward over time. Unlike supervised learning, RL does not require a labeled dataset; instead, the agent learns through trial and error.

Reinforcement learning has achieved remarkable success in domains such as game-playing and robotics, and it is being explored in human medicine for applications such as optimizing treatment regimens.[1,25] In veterinary medicine, RL applications are still largely theoretic, but potential use cases include optimizing anesthesia protocols, adaptive radiation therapy planning, and individualized treatment strategies that respond to changes in a patient's clinical status over time.

## COMMON MACHINE LEARNING ALGORITHMS AND ARCHITECTURES

A variety of ML algorithms are used in veterinary research and clinical applications. While an exhaustive treatment is beyond the scope of this article, familiarity with the most commonly encountered algorithms will help veterinary professionals interpret the literature and understand the tools being developed for clinical use (**Fig. 3**).

### *Decision Trees and Ensemble Methods*

A decision tree is an intuitive algorithm that makes predictions by learning a series of if–then rules from the data. At each step, the algorithm splits the data based on the feature that most effectively separates the outcome classes. The resulting structure resembles a branching tree, making it one of the most interpretable ML models available.[26]

However, single decision trees are prone to overfitting—learning the noise in the training data rather than the underlying pattern—which limits their generalizability. To address this, ensemble methods combine multiple decision trees to improve performance. Random forests build many trees using random subsets of the data and features, then aggregate their predictions (typically by majority vote for classification tasks).[27] Gradient-boosted trees, such as XGBoost, build trees sequentially, with each new tree correcting the errors of the previous ones.[28] Both random forests and gradient-boosted methods are widely used in veterinary and human ML studies, including radiomics analyses, survival prediction models, and clinical decision support systems.[29,30]

### *Support Vector Machines*

Support vector machines (SVMs) are algorithms that find the optimal boundary (hyperplane) separating data points into different classes. The "support vectors" are the data points closest to the decision boundary, and the algorithm seeks to maximize the margin between classes. SVMs can handle linear and nonlinear classification problems through the use of kernel functions, which map the data into higher dimensional spaces where separation is easier.

SVMs have been applied in medicine for tasks such as disease classification from laboratory data, prognostic prediction, and image-based diagnosis.[6,26,31] While they can be highly effective for smaller datasets, they are generally less scalable to very large datasets and complex tasks compared to DL approaches.

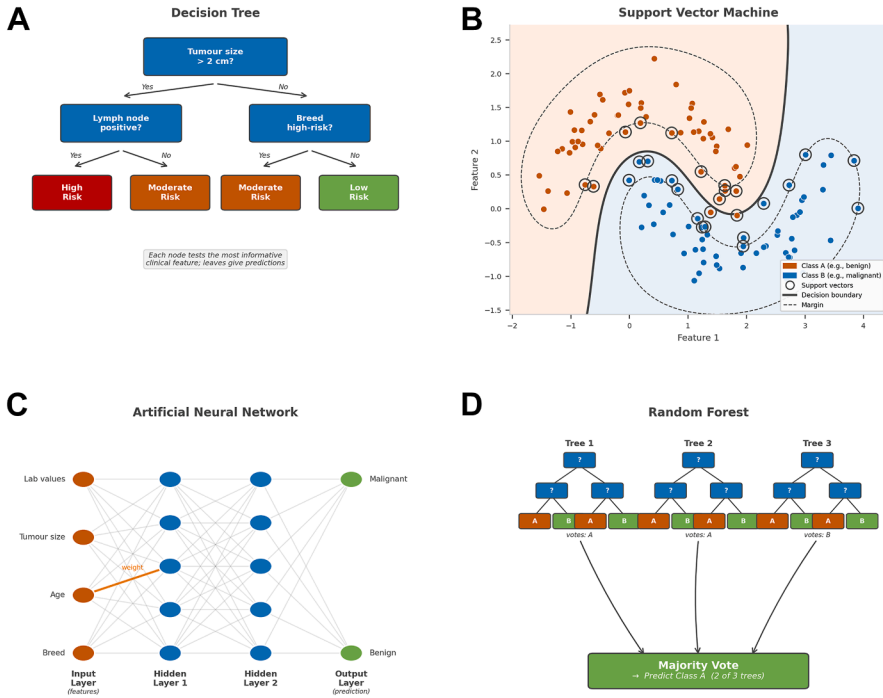



**Fig. 3.** Illustrations of 4 common machine learning algorithms. (*A*) Decision tree: a hierarchical series of binary splits on clinical features, leading to outcome predictions at the leaf nodes. (*B*) Support vector machine (SVM): a nonlinear decision boundary (*solid line*) separates 2 classes in feature space; support vectors (circled points) define the maximum-margin boundary, with the margin shown as dashed lines. (*C*) Artificial neural network (ANN): input clinical features are passed through 2 hidden layers of interconnected neurons (highlighted connection illustrates a single learnable weight) to produce a binary prediction at the output layer. (*D*) Random forest: an ensemble of independently trained decision trees, each producing a class vote; the final prediction is determined by majority vote across all trees.

## ARTIFICIAL NEURAL NETWORKS AND DEEP LEARNING ARCHITECTURES

Artificial neural networks (ANNs) are inspired by the structure of biological neural circuits. A basic ANN consists of an input layer (receiving data), 1 or more hidden layers (performing computations), and an output layer (producing predictions). Each connection between neurons carries a weight that is adjusted during training to minimize prediction error. When an ANN has multiple hidden layers, it is referred to as a deep neural network, and training such networks constitutes DL.[18]

Several specialized DL architectures have become particularly important in veterinary AI.

- *CNNs:* Designed for processing grid-like data such as images, CNNs use convolutional filters to automatically learn spatial features at multiple scales. They are the foundation of most image-based AI tools in veterinary radiology, histopathology, and cytology.[17,20,21]
- *Recurrent neural networks and long short-term memory networks:* Designed for sequential data, these architectures are used for time-series analysis (such

as electrocardiogram interpretation) and natural language processing (NLP) tasks.
- *Transformer models:* The architecture underlying LLMs such as GPT and similar systems, transformers process data in parallel rather than sequentially and have revolutionized NLP.

### TRANSFER LEARNING

Transfer learning is an approach in which a model pretrained on a large dataset is adapted for a new, related task using a much smaller dataset. Rather than training a model from scratch, transfer learning allows researchers to leverage feature representations already learned from large-scale data—for example, a CNN pretrained on millions of general images can be fine-tuned to classify veterinary histopathology slides using only a few hundred labeled examples. This is particularly valuable in veterinary medicine, where the data volumes required to train DL models from scratch are rarely available. Transferring from human medical models to veterinary data has only begun to be explored. As it relates to language model tasks, there is much to be sought after in terms of model performance and understanding to the veterinary context with direct human model adaptation.[14] This is a consideration of growing relevance as the breadth of human medical AI expands.

### LOGISTIC REGRESSION

Although technically a statistical method rather than an ML algorithm in the strictest sense, logistic regression is frequently included in ML studies as a baseline model. It predicts the probability of a binary outcome (such as metastasis present vs absent) based on a linear combination of input features. Logistic regression is highly interpretable and remains a valuable tool, particularly when the relationships between predictors and outcomes are approximately linear and the dataset is small. It is a common benchmark against which more complex ML models are compared.[26]

## KEY CONCEPTS FOR EVALUATING MACHINE LEARNING MODELS

Understanding how ML models are built, evaluated, and validated is critical for veterinary professionals who wish to assess the quality and clinical relevance of AI-based tools. Several key concepts recur throughout the literature; a glossary of key terms is provided in **Table 2**.

### TRAINING, VALIDATION, AND TESTING

The data used to develop an ML model are typically divided into 3 subsets. The training set is used to fit the model's parameters. The validation set is used during development to tune hyperparameters (model settings that are not learned from the data) and to monitor for overfitting. The test set is held out entirely during development and used only for a final, unbiased estimate of model performance. Rigorous separation of these datasets is essential; if the model has "seen" test data during training, performance estimates will be artificially inflated.[26]

When data are scarce—a common situation in veterinary research—cross-validation techniques are often used. In k-fold cross-validation, the data are divided into k subsets, and the model is trained and evaluated k times, each time using a different subset as the validation fold. Leave-one-out cross-validation, where each individual case serves as the validation fold in turn, is sometimes used for very small datasets.

**Table 2**
**Glossary of key artificial intelligence and machine learning terms for veterinary professionals**

| Term | Definition |
|---|---|
| Algorithm | A set of step-by-step mathematical instructions that a computer follows to solve a problem or make a prediction |
| Training data | The dataset used to teach a machine learning model to recognize patterns; the quality and diversity of training data directly influence model performance |
| Validation data | A subset of the dataset used during model development to tune hyperparameters and monitor for overfitting, without being used to fit the model's primary parameters |
| Testing data | A held-out subset of the dataset used only for the final, unbiased evaluation of a trained model's performance; the test set must not be used during model training or hyperparameter tuning |
| Feature | A measurable property or characteristic of the data used as input to the model (eg, patient age, tumor size, and pixel intensity) |
| Label | The known outcome or category assigned to each training example in supervised learning (eg, "metastasis present" or "metastasis absent") |
| Overfitting | When a model learns the noise or specific details of the training data rather than the general underlying pattern, resulting in poor performance on new data |
| Underfitting | When a model is too simple to capture the underlying patterns in the data, resulting in poor performance on both training and new data. Underfitting is typically addressed by increasing model complexity or providing more informative features |
| Hyperparameter | A model setting chosen by the developer before training begins; fine-tunes the model (eg, learning rate, number of trees in a random forest) |
| Neural network | A computational model inspired by biological neural circuits, consisting of interconnected layers of nodes (neurons) that process data |
| Convolutional Neural Network (CNN) | A deep learning architecture specialized for processing images, using convolutional filters to automatically detect visual features at multiple scales |
| Transfer learning | A technique in which a model pretrained on a large dataset (eg, millions of general images) is adapted for a new, smaller dataset (eg, veterinary radiographs), leveraging previously learned features |
| AUC-ROC | Area under the receiver operating characteristic curve; a metric summarizing a model's ability to discriminate between classes across all classification thresholds (0.5 = chance, 1.0 = perfect) |
| Data augmentation | A technique for artificially expanding a training dataset by applying transformations (eg, rotation, flipping, and cropping) to existing data, commonly used in image-based tasks |
| Radiomics | The extraction of large numbers of quantitative features from medical images, converting visual information into mineable data for predictive modeling |
| Pathomics | The extraction of large numbers of quantitative features from pathologic images, converting visual information into mineable data for predictive modeling |

For studies involving multiple samples from the same patient (such as multiple imaging slices), leave-one-patient-out cross-validation ensures that all data from a given patient is assigned to the same fold, preventing data leakage between testing and training datasets.[26]

## OVERFITTING AND UNDERFITTING

Overfitting occurs when a model learns the specific patterns and noise in the training data too well, resulting in excellent training performance but poor generalization to new data. Overfitting is a particular risk with complex models (such as deep neural networks) and small datasets—both common scenarios in veterinary ML research. Indicators of overfitting include a large gap between training and validation performance.

Underfitting, conversely, occurs when a model is too simple to capture the underlying relationships in the data, resulting in poor performance on both training and new data. Strategies to address overfitting include regularization techniques (which penalize model complexity), dropout (randomly disabling neurons during training), data augmentation (artificially expanding the training set by applying transformations such as rotations or flips to images), and early stopping (halting training when validation performance begins to deteriorate).[18,26] Selecting an appropriately complex model for the size and nature of the available data is a fundamental aspect of ML development.

## PERFORMANCE METRICS

Multiple metrics are used to evaluate ML model performance, and no single metric tells the complete story. The most commonly reported metrics in veterinary ML studies include

- *Accuracy:* The proportion of correct predictions among all predictions. While intuitive, accuracy can be misleading when classes are imbalanced (eg, if only 10% of patients in a dataset have the condition of interest, a model that simply predicts "negative" for every case achieves 90% accuracy without any diagnostic value).
- *Sensitivity (Recall):* The proportion of true positive cases correctly identified by the model. High sensitivity is important when missing a positive case carries significant consequences.
- *Specificity:* The proportion of true negative cases correctly identified. High specificity is important when false positives carry significant consequences.
- *Precision (positive predictive value):* The proportion of predicted positive cases that are truly positive. Unlike sensitivity and specificity, precision positive predictive value (PPV) is directly dependent on disease prevalence: even a highly sensitive and specific test will have a lower PPV when applied in a population where the condition is rare.
- *Area under the receiver operating characteristic curve (AUC-ROC):* A threshold-independent metric that summarizes the model's ability to discriminate between classes across all possible classification thresholds. An AUC of 0.5 indicates performance equivalent to chance, while 1.0 indicates perfect discrimination.[26]
- *F1 Score:* The harmonic mean of precision and sensitivity, providing a single metric that balances both concerns. It is particularly useful when class distributions are unequal.

When evaluating AI tools for clinical use, veterinary professionals should consider which metrics are most relevant to the intended clinical application. A screening

test may prioritize sensitivity, while a confirmatory test may prioritize specificity. Reporting only accuracy without sensitivity, specificity, and AUC is often insufficient for assessing clinical utility.

## CLASS IMBALANCE

Many veterinary datasets exhibit class imbalance, where one outcome is much more common than another. For instance, in a dataset of lymph node evaluations, the majority of lymph nodes may be nonmetastatic. Class imbalance can bias ML models toward the majority class, resulting in models that appear accurate overall but perform poorly on the minority (often more clinically important) class.

Common strategies for addressing class imbalance include oversampling the minority class (eg, using the synthetic minority oversampling technique, or SMOTE, which creates synthetic examples),[32] undersampling the majority class, adjusting class weights within the algorithm, and using metrics beyond accuracy (such as AUC, F1 score, or balanced accuracy) to evaluate performance.

## FEATURES, DATA, AND THE ROLE OF DOMAIN EXPERTISE

In ML, a "feature" is any measurable property or characteristic of the data used as input to the model. In a clinical prediction model, features might include patient age, breed, body weight, laboratory values, and clinical stage. In radiomics, features are quantitative descriptors extracted from medical images, such as texture metrics, shape descriptors, and intensity histograms.[30]

Feature engineering—the process of selecting, transforming, and creating features from raw data—is often the most impactful step in the ML pipeline for traditional (non–deep-learning) models. Domain expertise is critical at this stage: a veterinary professional's understanding of disease biology and clinical relevance can guide the selection of features that are biologically meaningful and likely to improve model performance. For example, knowing that certain imaging texture features correlate with tumor heterogeneity can inform feature extraction in a radiomics study.

DL models, by contrast, are often described as performing automatic feature extraction, learning directly from raw data (such as image pixels) without the need for manual feature engineering. While this is a significant advantage in terms of workflow, it does not eliminate the need for domain expertise. Veterinary professionals remain essential for defining clinically relevant tasks, curating and labeling training data, identifying potential sources of bias, and interpreting model outputs in clinical context.

## DATA QUALITY AND AVAILABILITY

The performance of any ML model is fundamentally limited by the quality and quantity of the data on which it is trained. The commonly cited principle of "garbage in, garbage out" applies with particular force in ML: biased, incomplete, or mislabeled data will produce unreliable models regardless of algorithmic sophistication.

Veterinary medicine faces several data-related challenges that distinguish it from human medical AI. These include the diversity of species and breeds encountered, smaller patient populations for many conditions, less standardized electronic health record systems, and limited availability of large, curated, annotated datasets.[7] Multi-institutional collaborations and data-sharing initiatives are increasingly recognized as essential for overcoming these limitations and developing AI tools that generalize across practice settings.

## CURRENT APPLICATIONS OF ARTIFICIAL INTELLIGENCE IN SMALL ANIMAL VETERINARY MEDICINE

While detailed discussions of specific applications are provided in subsequent articles in this issue, a brief overview of the current landscape provides useful context for the foundational concepts discussed earlier.

### DIAGNOSTIC IMAGING

AI-assisted interpretation of radiographs is one of the most visible and commercially advanced applications of AI in veterinary medicine. Several commercial products use DL (primarily CNNs) to detect abnormalities on thoracic and musculoskeletal radiographs, including cardiomegaly, pulmonary patterns, and orthopedic conditions. These tools are typically designed to support, rather than replace, veterinary interpretation by highlighting regions of interest and providing probabilistic assessments.[4,8,10,11,23,33] Radiomics—the extraction and analysis of quantitative imaging features—is also being applied to CT and MRI data for tasks such as tumor characterization, prognostication, and treatment response prediction.[30]

### CYTOLOGY AND HISTOPATHOLOGY

Computer vision algorithms are being developed and validated for automated cytology and histopathology interpretation. Applications include mitotic figure detection, automated classification, and differentiation of neoplastic from inflammatory conditions. These tools have the potential to improve consistency, reduce workload for pathologists, and expand access to expert-level interpretation in settings where specialist availability is limited.[12,34,35]

### CLINICAL DECISION SUPPORT AND PREDICTIVE ANALYTICS

ML models are being developed to support clinical decision-making in areas such as disease prognosis, treatment response prediction, and risk stratification. Examples include models predicting survival outcomes in oncology patients, identifying patients at risk for specific complications, and estimating disease progression. Electronic health record data, combined with ML techniques, can also support population-level surveillance and identification of emerging disease trends.[29,36]

### NATURAL LANGUAGE PROCESSING AND LARGE LANGUAGE MODELS

NLP techniques, and particularly LLMs such as GPT-4,[37–39] are being explored for a variety of veterinary applications. These include automated generation of medical records and discharge summaries, client communication, literature review and synthesis, and educational applications. While the capabilities of LLMs are impressive, their tendency to produce plausible but inaccurate information (often termed "hallucination") necessitates careful human oversight in clinical settings.[13–15]

## CHALLENGES AND CONSIDERATIONS FOR CLINICAL ADOPTION

### *Interpretability and Explainability*

A persistent concern with many ML models, particularly DL systems, is their "black box" nature: they can make accurate predictions without providing a clear explanation of how they arrived at their conclusions. This lack of transparency can be a barrier to clinical trust and adoption. Efforts to address this include the development of explainability techniques such as gradient-weighted class activation mapping, which

highlights the regions of an image most influential in a model's classification,[40] and SHAP (SHapley Additive exPlanations) values, which quantify the contribution of each feature to a prediction.[41]

### *Generalizability and External Validation*

A model that performs well on data from the institution where it was developed may not perform as well when applied to data from a different institution, geographic region, or patient population. This is because the model may have learned features specific to the training environment (such as imaging equipment characteristics or local disease prevalence) rather than the underlying biology. External validation—testing a model on independent data from a different source—is an important step in establishing that an AI tool is robust and clinically useful. Multi-institutional validation studies, though logistically challenging, are particularly informative.[7,10]

### *Regulatory and Ethical Considerations*

The regulatory landscape for AI in veterinary medicine is still evolving. Unlike human medical devices, which are subject to rigorous regulatory oversight by bodies such as the US Food and Drug Administration, veterinary AI tools currently face less formal scrutiny. Questions of liability (when an AI-assisted diagnosis is incorrect), data privacy, algorithmic bias, and the appropriate role of AI in clinical decision-making are all areas of active discussion within the veterinary profession.[16,42] These topics are addressed in greater detail in subsequent articles in this issue.

### *Artificial Intelligence Literacy and Education*

As AI tools become increasingly common in clinical practice, the need for AI literacy among veterinary professionals grows correspondingly. Understanding the basics of how these tools work—their capabilities, limitations, and appropriate use cases—is essential for effective and responsible adoption. Integrating AI and data science concepts into veterinary curricula and continuing education programs is an important step toward ensuring that the profession is prepared for this technological transformation.[16,43]

## SUMMARY

AI and ML represent a significant and accelerating area of innovation in veterinary medicine. The foundational concepts discussed in this article—the distinctions among AI, ML, and DL; the major learning paradigms; common algorithms and architectures; and key principles of model evaluation—provide the vocabulary and conceptual framework needed for veterinary professionals to engage with this rapidly evolving field. As AI tools become more prevalent in clinical practice, an informed and critically engaged veterinary profession will be essential for ensuring that these tools are developed, validated, and applied in ways that improve the quality of care for our patients.

## CLINICS CARE POINTS

- When evaluating an AI diagnostic tool for clinical use, veterinary professionals should inquire about the composition and diversity of the training dataset, including species, breeds, and imaging equipment represented, as models trained on narrow datasets may not generalize well to the broader patient population.
- AI-assisted diagnostic tools are designed to support, not replace, clinical judgment; findings flagged by AI systems should be interpreted within the context of the full clinical picture, and discordant results should prompt additional investigation rather than blind reliance on the algorithmic output.

- Accuracy alone is insufficient for evaluating an AI tool's clinical utility; sensitivity, specificity, AUC, and the clinical context (such as disease prevalence in the target population) should all be considered.
- DL models can achieve impressive performance metrics during development but may fail when applied to data from different equipment, institutions, or patient populations; external validation data should be considered strong evidence of a tool's robustness.
- Veterinary professionals should be aware that AI outputs may reflect biases present in training data and should maintain a low threshold for critical appraisal, especially for underrepresented patient populations or atypical presentations.

## DECLARATION OF ARTIFICIAL INTELLIGENCE AND ARTIFICIAL INTELLIGENCE-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this article, the author used Claude (Anthropic, San Francisco, CA) in order to assist with formatting and highlighting of section headings, spell-checking, and reformatting sentence structure. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

## DISCLOSURE

C. Pinard is the CEO and cofounder of ANI.ML Health Inc. and its research subsidiary (ANI.ML Research).

## REFERENCES

1. Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. Nat Med 2022;28(1):31–8.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44–56.
3. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. Nature 2020;577(7788):89–94.
4. Appleby RB, Basran PS. Artificial intelligence in veterinary medicine. J Am Vet Med Assoc 2022;260(8):819–24.
5. Akinsulie OC, Idris I, Aliyu VA, et al. The potential application of artificial intelligence in veterinary clinical practice and biomedical research. Front Vet Sci 2024;11:1347550.
6. Cihan P, Gokce E, Kalipsiz O. A review of machine learning applications in veterinary field. Kafkas Univ Vet Fak Derg 2017;23(4):673–80.
7. Basran PS, Appleby RB. The unmet potential of artificial intelligence in veterinary medicine. Am J Vet Res 2022;83(5):385–92.
8. Hennessey E, DiFazio M, Hennessey R, et al. Artificial intelligence in veterinary diagnostic imaging: a literature review. Vet Radiol Ultrasound 2022;63(S1):851–70.
9. Pinard CJ. The future is here: an introduction to the veterinary oncology collection on artificial intelligence and informatics. Vet Oncol 2025;2:10.
10. Joslyn S, Alexander K. Evaluating artificial intelligence algorithms for use in veterinary radiology. Vet Radiol Ultrasound 2022;63(6):871–9.
11. Nyquist ML, Fink LA, Mauldin GE, et al. Evaluation of a novel veterinary dental radiography artificial intelligence software program. J Vet Dent 2024;41(1):3–12.

12. Rai T, Morisi A, Bacci B, et al. Keeping pathologists in the loop and an adaptive F1-score threshold method for mitosis detection in canine perivascular wall tumours. Cancers 2024;16(3):644.
13. Chu CP. ChatGPT in veterinary medicine: a practical guidance of generative artificial intelligence in clinics, education, and research. Front Vet Sci 2024;11:1395934.
14. Pinard CJ, Poon AC, Lagree A, et al. Precision in parsing: evaluation of an open-source Named Entity Recognizer (NER) in veterinary oncology. Vet Comp Oncol 2025;23(1):102–8.
15. Poore TJ, Pinard CJ, Shabbir A, et al. Context matters: comparison of commercial large language tools in veterinary medicine. Preprint 2025. arXiv:2510.01224.
16. Bertin FR, Lawrence J, Niessen SJM, et al. The influence, promise, and potential perils of artificial intelligence in veterinary medicine: a call for improved awareness and literacy. J Vet Intern Med 2026;40(1):aalaf045.
17. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.
18. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, MA: MIT Press; 2016.
19. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
20. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90.
21. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. June 27-30 2016, Las Vegas (NV), 2016:770-778.
22. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115–8.
23. Banzato T, Wodzinski M, Tauceri F, et al. An AI-based algorithm for the automated detection of cardiogenic pulmonary edema on thoracic radiographs in dogs. Vet Radiol Ultrasound 2022;63(3):292–7.
24. Garcia R, Aguilar J, Toro M, et al. A systematic literature review on the use of machine learning in precision livestock farming. Comput Electron Agric 2020;179:105826.
25. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016;529(7587):484–9.
26. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer; 2009.
27. Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
28. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Aug 13-17 2016, San Francisco (CA), 2016:785-794.
29. Bohannan Z, Pudupakam RS, Koo J, et al. Predicting likelihood of in vivo chemotherapy response in canine lymphoma using ex vivo drug sensitivity and immunophenotyping data in a machine learning model. Vet Comp Oncol 2021;19(1):160–71.
30. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48(4):441–6.
31. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.
32. Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321–57.

33. Hespel A, Zhang Y, Basran PS. Artificial intelligence 101 for veterinary diagnostic imaging. Vet Radiol Ultrasound 2022;63(S1):817–27.
34. Zuraw A, Aeffner F. Whole-slide imaging, tissue image analysis, and artificial intelligence in veterinary pathology: an updated introduction and review. Vet Pathol 2022;59(1):6–25.
35. Fragoso-Garcia M, Wilm F, Bertram CA, et al. Automated diagnosis of 7 canine skin tumors using machine learning on H&E-stained whole slide images. Vet Pathol 2023;60(6):865–75.
36. Obermeyer Z, Emanuel EJ. Predicting the future — big data, machine learning, and clinical medicine. N Engl J Med 2016;375(13):1216–9.
37. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Adv Neural Inf Process Syst 2017;30:5998–6008.
38. Devlin J, Chang MW, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT. June 2-7 2019, Minneapolis (MN), 2019:4171-4186.
39. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. Adv Neural Inf Process Syst 2020;33:1877–901.
40. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 2020;128(2):336–59.
41. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30:4765–74.
42. Duggirala HJ, Johnson JL, Tadesse DA, et al. Artificial intelligence and machine learning in veterinary medicine: a regulatory perspective on current initiatives and future prospects. Am J Vet Res 2025;86:S16–21.
43. Bertin FR, Lawrence J, Niessen SJM, et al. Use of, and attitudes toward, artificial intelligence in members of the American College of Veterinary Internal Medicine and the European College of Veterinary Internal Medicine – companion animals. J Vet Intern Med 2026;40(1):aalaf036.