Artificial Intelligence in Veterinary Diagnostic Imaging and Radiation Oncology
ACVR/ECVDI Expert Panel Report
April 2021

This report was prepared following panel discussions of a group of experts in artificial intelligence (AI) in veterinary radiology and included diplomates of the ACVR and ECVDI in diagnostic imaging and radiation oncology and medical physicists working to develop AI algorithms for veterinary radiology and radiation oncology applications. Panel members represented individuals from academia, teleradiology practice and commercial interests.

The objectives were to explore the current state of AI in veterinary radiology and radiation oncology, to understand the pitfalls and challenges associated with the responsible development of AI applications in veterinary radiology, to promote the expertise of radiologists and radiation oncologists in the development of AI applications in veterinary radiology and to encourage AI applications that will benefit veterinary radiologists/radiation oncologists and by extension the veterinary profession as a whole.

What is artificial intelligence?
It is beyond the scope of this report to define and describe artificial intelligence. For the purposes of this report, artificial intelligence is considered to be any technique where a computer algorithm has the ability to modify itself, without requiring the specific input of a human programmer. It includes and is not limited to the terms machine learning, deep learning, radiomics and neural networks.

Current state of AI in veterinary radiology and radiation oncology
Artificial intelligence techniques are being actively developed and are starting to be implemented in a number of facets of veterinary radiology, namely: image quality and enhancement, interpretation workflow and image interpretation.

The area that has gained the most attention is image interpretation and is where this report will focus primarily. There have been recent presentations and publication of a number of academic studies describing the development of image interpretation algorithms and automated measurement tools. At least two commercial products for small animal thoracic radiographic interpretation have been released in the past year or so. At least one veterinary teleradiology company is developing an interpretation algorithm to facilitate interpretation by its radiologists. Commercial software for automated measurements in equine orthopedics and small animal cardiology are available on the immediate horizon. Thus far, veterinary radiologists have had at least a degree of oversight with these endeavors. As far as the panel was aware, products have not yet been released by non-veterinary entities.

At the moment, the focus is in companion animals and primarily on the thorax. However, we can expect that other anatomic regions, species and modalities will follow.

In addition to image-based diagnosis, AI has great potential to facilitate the veterinary radiologist's workflow. The major teleradiology providers are actively implementing tools to increase efficiency, including automated personalized hanging protocols. We can expect AI technology to emerge to facilitate reporting, report measurements, track radiologist efficiency and oversee other administrative tasks. Another potential application is the triage of cases to allow prioritization of case interpretation.

Lastly, artificial intelligence is making incursions into the taking of medical images. AI techniques are being developed to give animal health technicians real time feedback and direction to improve radiographic image quality. Other developers are looking at how to accurately identify limb laterality in horses, potentially eliminating the need for radiographic markers. These applications can be more insidious at the image processing level to reduce image artifacts and it is unclear how these modifications may affect lesion detectability. Beyond radiography, other imaging modalities are integrating AI to help imaging technologists; for example smart scan applications in MRI or to minimize patient dose in CT.

In radiation oncology, applications are being developed and deployed in human medicine for normal tissue contouring, treatment planning, quality assurance, and lesion tracking during radiation delivery. A primary area of interest is the automation of normal tissue contouring, as this is currently a fairly labor-intensive process. However, at this time, there are no veterinary-specific applications and veterinary anatomy presents a particular challenge. Going forward, there will also be an interest in tools that can mine clinical databases for correlating patient outcomes with the radiation dose delivered to better inform treatment protocols. Also, AI algorithms that can integrate image interpretation, pathology and biomarker data with radiation oncology clinical data will be of value to understanding and improving patient prognosis and treatment outcomes.

A comparison with AI for image interpretation in human diagnostic imaging

With respect to image interpretation, the development of algorithms in veterinary medicine is not specifically following the human medicine blueprint.

In human medicine in North America, the US Food and Drug Administration (FDA) considers AI algorithms as "software in a medical device". There are stringent guidelines that clinical trials on emerging AI algorithms must follow to gain approval. There are no equivalent regulatory bodies to oversee the approval of AI algorithms for diagnostic use in veterinary medicine. This goes as much for image acquisition systems (i.e. X-ray machines) as for software for image interpretation. We are essentially self-regulated. This allows the development of new AI algorithms despite limited resources, but also leaves the veterinary profession vulnerable to misleading and potentially harmful technological or diagnostic claims and economic incentives.

Because of the regulatory requirements in human medicine, AI algorithms must undergo an extensive validation process before they can be used commercially. For this reason, algorithms are often much more specific, focusing on limited imaging findings and/or diagnosis (i.e. radiographic detection pulmonary nodules, detection of acute cerebral hemorrhage on CT for the diagnosis of hemorrhagic stroke). Currently, in veterinary medicine, the approach is much broader: i.e. the detection and classification of abnormalities on thoracic radiographs in dogs.

The other notable difference between human and veterinary radiology is that a radiologist sees most imaging studies in humans. Radiologists have therefore played a critical role in driving the development of algorithms to improve their own diagnostic accuracy. This is known as AI facilitated diagnosis or a clinical decision support system. Systems are divided into computer-aided detection algorithms (CADe) and computer-aided diagnosis systems (CADx). The distinction between detection and diagnosis is less clear in emerging AI algorithms in veterinary medicine. Also, in veterinary radiology, general practitioners (and perhaps eventually animal owners themselves) are a direct target audience for AI. Contrary to human medicine, direct AI image interpretation has the potential to obviate the expertise of veterinary radiologists. However, if developed responsibly, it could improve image interpretation for the large subset of patients whose images are not referred to a veterinary radiologist.

Challenges in developing AI for image interpretation in veterinary medicine

Artificial intelligence will impact all of veterinary medicine, including radiology. It is a tool with the potential to help both the general practitioner and the veterinary radiologist improve the level of care that we provide our patients. It nevertheless can cause widespread harm to our patients if it is not developed in a responsible, systematic manner. Indeed, the potential for harm by a single faulty AI algorithm could number in the tens of thousands of patients. The ethical and legal ramifications should therefore not be underestimated.

As veterinary radiologists, we hold the global expertise in image interpretation in animals, and that expertise is constantly evolving. We have a gatekeeper role to ensure our expertise is used to develop rigorous, evidence-based AI algorithms. In particular, a veterinary radiologist is more likely than a general practitioner to recognize when an AI algorithm has made an error. Conversely, we must be open to what AI might teach us about our diagnostic shortcomings and be willing to learn from what the technology may reveal.

In order to develop a diagnostically accurate AI algorithm, a training set of cases must be developed, that addresses the following requirements (See Annex 2):

- Adequate and appropriate dataset sizes
- High quality images:
  - Accurate radiographic positioning and technique
  - A representative cross-section of image acquisition systems
- Case variety representative of the eventual population to which the algorithm will be applied (species, breeds, ages, gender, disease conditions, etc…). This should include a representative subset of rare diseases as well as common ones, as well as unusual presentations of common diseases.
- Accurate findings and diagnosis on each training case:
  - Labeled by at least one or several experts (radiologists)
  - Verifed by a final "ground truth"

Once the AI algorithm is developed, it should undergo a validation step. Validation is performed on a smaller, but equally representative and high-quality case set (See Annex 2). This validation set should be completely independent of the original training set. With the validation step, diagnostic accuracy of the AI algorithm can be calculated. Finally, AI algorithms should periodically be re-assessed, in order to ensure that performance remains constant for the clinical caseload for which the algorithm is intended.

Now that we understand the steps to developing a diagnostically accurate AI algorithm, we can see the particular challenges to developing these algorithms in veterinary medicine. The first challenge is accessing a large data set. At the moment, this is primarily the purview of university teaching hospitals and large teleradiology companies. Going forward, there are issues for general practitioners and radiologists to consider around data ownership, access and utilization, as well as patient privacy. Radiographic images and report data are increasingly valuable commodities.

The second challenge is ensuring high radiographic quality. Indeed, AI algorithms can compensate for some errors in quality and positioning and in fact, their development should include less-than-perfect cases (i.e. underexposure/overexposure). However, the greater the variability of image quality used to develop the algorithm, the larger the requisite data set to develop a robust algorithm. Adding to this is the inherent variability and complexity of radiographic anatomy in some species: the classic Chihuahua-to Bulldog-to Great Dane challenge that we face every day.

The third and perhaps greatest challenge is obtaining an accurate description of the imaging findings and the closest approximation to a final diagnosis ("ground truth") on each case in the training case set. In human medicine, there are a multitude of labeling systems that have been devised (i.e. RADLEX). No standard has been established in veterinary medicine. The highest available level of expertise should be sought to ensure the most accurate coding possible. However, data labeling is an exceedingly resource-intensive exercise. In addition, accurate labeling becomes increasingly difficult in the face of non-binary imaging findings (i.e. air bronchograms no/yes vs unstructured interstitial pattern none/artifactual/mild/moderate/severe). Structured reporting and Natural Language Processing techniques can facilitate data labeling. Semi-automatic and automatic "black-box" AI learning techniques are not currently approved for use in image interpretation in humans. Until we have a more transparent understanding on how these techniques teach AI algorithms, we should remain wary about their use in veterinary image interpretation as well.

A final diagnosis (or "ground truth") is often elusive. Because clinical pathology, histopathology and surgical confirmation are not available on the majority of cases, ground truth is often presumed by clinical response to treatment and follow-up, requiring both availability and a thorough study of the patient's complete medical record. Because of its data mining power, AI itself may eventually offer a solution to this problem, assuming the AI algorithms for data extraction from the patient's medical record are themselves well developed. AI algorithms developed without the benefit of a "final" diagnosis cannot report diagnostic accuracy (sensitivity/specificity). They are instead a reflection of agreement with the presumed diagnosis. Regardless of the availability of a ground truth, AI algorithms should also show consistency, and confidence intervals should be available to users.

The final challenge in developing accurate AI algorithms is validation and ongoing assessment. A number of institutions in human medicine (i.e. the American College of Radiology-DSI Certify-AI and Assess-AI directories) have developed independently curated case sets for AI algorithms. There are not currently any widely available reference datasets for the validation and ongoing assessment of academically or commercially-developed veterinary AI-algorithms. Validation datasets will be increasingly valuable to compare the performance of different algorithms as these become available. To maintain validity, curators of validation data sets need to remain independent from algorithm developers.

For the moment, a form of algorithm validation in veterinary medicine occurs via publication of an academic research project that studies the agreement/accuracy of an AI algorithm. This is an increasingly active field of research. However, as with any traditional scientific research, the peer review process is lengthy. In other domains, open-source forums that rely on reader peer review are becoming a more rapid alternative. Also, specific author and research guidelines are needed to facilitate the review of AI research for publication going forward.

Challenges to developing AI in veterinary radiation oncology

Similar to diagnostic imaging, the primary challenge in radiation oncology is developing and accessing large data-rich patient and image databases to develop robust AI algorithms. Patient caseloads are smaller in radiation oncology; therefore, it takes much longer to accumulate a significant volume of cases for training and validation of an AI algorithm. Multi-institutional efforts are likely necessary to accrue large enough patient databases. Common diseases such as nasal tumors or meningiomas may initially lend themselves best to developing AI applications.

Also similar to diagnostic imaging, it can be challenging to extract accurate data from the patient's medical record and imaging reports Language, taxonomy, electronic health record systems, diagnostic imaging and radiation oncology databases used for patients can vary from one individual to the next and between institutions. Patient follow-up can also be difficult to obtain and then to classify in a meaningful enough manner to train an AI algorithm. Standardized and harmonized patient databases and Natural Language Processing techniques are tools that will be required to advance radiation oncology AI applications.

Obtaining ground truth is as much a challenge in radiation oncology as in diagnostic imaging and the definition of ground truth will vary. For example, necropsy results may be appropriate for determining lesion size, while patient outcome may be appropriate for determining the best treatment protocol. Indeed the development of AI applications in other specialty fields (i.e. pathology, medical oncology) will indirectly help radiation oncology.

Finally, any AI application for radiation oncology will need to undergo the same training, validation and ongoing assessment as described above for diagnostic imaging, requiring development and maintenance of high quality independent datasets.

Benefits to developing AI for veterinary radiology and radiation oncology

Although there are a number of potential pitfalls and challenges surrounding the development of AI for veterinary radiology, there are also a number of ways that this technology will help veterinary radiologists and general practitioners, raising the overall standard of care of our patients. Artificial intelligence should be seen as a potential aid to radiographic interpretation.

Despite initiatives by the ACVR and ECVDI to expand veterinary diagnostic imaging residency programs, demand for veterinary radiologists still far outstrips the profession's ability to train a sufficient number of specialists. Well-developed AI algorithms could help maximize use of radiologists' expertise at 2 different levels: as a triage tool and as an add-on (AI facilitated diagnosis: a radiologist "enhancer"). As a triage tool, AI could work at 2 other levels: to direct the general practitioner to consult a radiologist and to direct the radiologist toward the cases that most need their expertise (for example: eliminate normal studies from the radiologist's worklist).

AI could also improve the radiologist's diagnostic performance. For example, AI facilitated diagnosis applications could direct the radiologist's attention toward abnormalities or provide a preliminary interpretation. Additionally, both in diagnostic imaging and radiation oncology, radiomics have the potential to detect abnormalities within the imaging data that are not perceptible to the human visual system and could eventually be applied in conjunction with more traditional image interpretation. In addition, AI has the advantage of not being susceptible to fatigue and less susceptible to certain perception and interpretation biases and intra and interday variation. These strengths could improve radiologists' consistency and reduce interpretation error rates. In a similar manner, AI facilitated diagnosis could also be used as a training tool for residents, to improve resident performance. In radiation oncology, AI could provide similar error-reduction and resident-training benefits with respect to lesion detection and characterization, normal tissue contouring and treatment planning.

Finally, the development, validation and re-assessment of AI applications have the potential to open an entirely new field of expertise within our specialty. Veterinary radiologists will develop AI applications, and should also aspire to play essential gatekeeper roles in the validation and on-going assessment of AI applications.

How can the ACVR (DI and RO) and ECVDI promote the responsible development of AI applications in veterinary radiology and radiation oncology?

It should be noted that the ACVR (DI and RO) and ECVDI are certification entities and professional organizations, but are not regulatory bodies and have neither the resources nor the mandate to act as such. Our role with regards to the advent of AI in veterinary radiology is to maximize our involvement as experts and educators.

There are several levels where the ACVR and ECVDI can promote the responsible development of AI in veterinary radiology. ACVR and ECVDI diplomates are the current experts in veterinary diagnostic imaging and we need to understand the influence AI will have on our profession in order to guide the conversation on the role this emerging technology will play. The ACVR and ECVDI can:

1) Improve diplomate and trainee literacy with regards to AI applications and development:
    - Educate on basic AI techniques and terminology.
    - Educate on basic understanding of the development of datasets for machine learning.
    - Develop media to improve our professional literacy:
        - Annual ACVR and ECVDI scientific meetings: keynote speakers, Forum, dedicated abstract section
        - ACVR Continuing Education LMS
        - Supplemental issue in Veterinary Radiology and Ultrasound
        - ACVR and ECVDI Websites : Repository of regularly updated tools /resources
        - Training grants for diplomates wanting to pursue advanced education in AI (i.e. via ACVR-DSI)

2) Communicate and represent expertise in AI in veterinary radiology:
    - Stakeholder groups:
        - ACVR/ECVDI diplomates and other veterinary radiology expert groups (i.e. IVRA)
        - Other specialty organizations (i.e. ACVIM, ECVIM, ABVS, EBVS)
        - General practitioners (i.e. AVMA, AAHA, state practice associations and organizations)
        - Commercial interests
        - Regulatory bodies
    - Communication channels:
        - Taskforce/committee/panel representation
        - Written communications: scientific articles, news articles, guidelines/recommendations
        - Presentations, chats and discussions (live or virtual/recorded)

3) Facilitate networking and collaboration between veterinary radiologists/oncologists and other expert groups in AI:
    - Increase involvement and collaboration with the AAPM (American Association of Physicists in Medicine) and other equivalent expert groups (i.e. human radiologists and radiation oncologists).

- Provide a meeting time/space, virtual and/or live at Annual ACVR and ECVDI scientific meetings.
- Hold a Data Science Summit, including the various academic and commercial stakeholders
- Facilitate the creation of a Society.

4) Facilitate AI research and development by veterinary radiologists/radiation oncologists:
    - Fund AI-specific resident and diplomate research grants and collaborations with other expert groups.
    - Fund training grants for diplomates wanting to pursue advanced education in AI (i.e. via ACVR-DSI).
    - Develop a publication standard for materials and methods for AI research in Veterinary Radiology and Ultrasound.
    - Provide education on alternative forms of open-source publication.
    - Encourage the development of high-quality case sets for both the training and the independent validation of AI algorithms. It is likely beyond the resources of the ACVR/ECVDI to develop their own data sets, but could facilitate this process, for example through commercial/academic collaboration.
    - Encourage the development of a common lexicon for image coding and/or a model for structured reporting.

Beyond the ACVR (DI and RO) and ECVDI: to those developing AI

The ACVR and ECVDI are critical stakeholders in the development of AI in veterinary radiology and radiation oncology. Diplomates will be involved in developing AI and will be asked to judge the contribution of these technologies to improving patient care. We will also be end users of AI technology, hopefully for the improvement of our own practice.

The following are some of the directions that we would like to see this technology taking in the years to come:

- Development of AI to improve radiologist workflow and efficiency:
    - Computer assisted reporting, i.e. automatic captions, curated and updated evidence-based lists of differential diagnoses.
    - Improved DICOM reading platforms, i.e. automated personalized hanging protocols and other image interpretation preferences.
- Development of AI-facilitated triage tools:
    - The ability to differentiate normal vs abnormal with the highest possible accuracy
    - Automated triaging based on patient medical record data
- Development of AI-driven assessment and feedback tools:
    - Assessment of radiologist (and trainee) performance, i.e. efficiency, accuracy, etc.
- AI algorithm transparency and access
    - Statements by AI developers outlining training and validation methodology as well as the accuracy/agreement/confidence results (see Annex 1)
    - AI algorithms that can be widely integrated on the common platforms that most radiologists use.
    - A common lexicon for image coding, model for structured reporting
- AI algorithm validation:

    - Development of independent, cost-effective and widely-available training and validation case sets.
- Radiomics:
    - Learn what our limited visual system cannot glean from the image data

Panel Co-Chairs:
Kate Alexander, Dipl. ACVR, ACVR president
Andrew Parry, Dipl. ECVDI, ECVDI president
Chris Warren-Smith, Dipl. ECVDI, ECVDI president
Tracy Gieger, Dipl. ACVIM (Oncology) and Dipl. ACVR (RO), ACVR RO president

Panel Members Diagnostic Imaging:
Ryan Appleby, Dipl. ACVR, University of Guelph
Michael Bailey, Dipl. ACVR, IDEXX and AVMA
John Craig, PhD, EponaTech
Erin Epperly, Dipl. ACVR, Cornell University and Signal Pet
Adrien Hespiel, Dipl. ACVR, University of Tennessee
Steve Joslyn, Dipl. ECVDI, VetDB
Del Leary, PhD, Medical Physicist, Colorado State University
Hiro Murakami, Dipl. ACVR, Purdue University
Matt Paek, Dipl. ACVR, Synergy VIP
Tobias Schwarz, Dipl. ACVR and ECVDI, University of Edinburgh and EBVS representative
Seth Wallack, Dipl. ACVR, Vetology
Diane Wilson, Dipl. ACVR, Antech Imaging Services

Panel Members Radiation Oncology:
Parminder Basran, PhD, Medical Physicist, Cornell University
Johnson Darko and André Fleck, Medical Physicists, Grand River Regional Hospital and University of Waterloo
Ira Gordon, Dipl, ACVR (RO), Ethos Veterinary Health
Elias Gumpel, Dipl. ACVR (RO), Brisbane Veterinary Specialist Centre
Valerie Poirier, Dipl. ACVIM (Oncology) and ACVR (RO), University of Guelph
David Ruslander, Dipl. ACVIM (Oncology) and ACVR (RO), Veterinary Specialty Hospital of the Carolinas

Annex 1

What to look for in an AI system for image interpretation? Some preliminary guidelines for veterinary radiologists and general practitioners:

Artificial intelligence systems are beginning to appear on the market to facilitate radiographic interpretation. These have the potential for a larger-than-ever number of patients to benefit from radiographic imaging. However, it is important to understand that the technology does not offer a one-stop shop for a final radiographic diagnosis.

Here are some questions to consider when evaluating the performance of an AI system for use in your veterinary practice:

1) What is the expertise behind the development of the algorithm?
    - Veterinary radiologists should be involved at all stages of algorithm development
2) How was the algorithm developed:
    - Where was the case set data obtained?
        - Multiple institutions?
    - How was minimum image quality ensured?
    - How were images labeled and by who?
        - Image labeling should be performed by the highest level possible of expertise (i.e. multiple veterinary radiologists)
    - How was ground truth (final diagnosis) defined and determined?
        - Histopathology/clinical pathology/surgical diagnosis (highest level)
        - Presumed clinical diagnosis
            - By whom? Specialist vs referring veterinarian
        - Presumed imaging diagnosis by a single radiologist, multiple individual radiologists or a consensus of multiple radiologists
3) How was the algorithm validated?
    - Appropriate validation reference dataset – see Annex 2.
    - What are the confidence intervals?
    - Was accuracy measured vs a ground truth, or agreement vs expert interpretation?
4) How is algorithm performance periodically re-assessed?
5) What information does the AI readout provide?
    - What abnormalities are the system trained to detect? What abnormalities is it not able to detect?
    - Yes/no vs % likelihood of a list of possible imaging findings? Are confidence intervals reported?
    - A ranked list of differential diagnoses?
6) In the case where the AI algorithm leads to a misdiagnosis, who has the legal responsibility?
7) What are the legal ramifications around the imaging data that has been sent to the AI system?
    - Who legally owns the data?
    - What permissions does the company have concerning the use of the data?

Annex 2

Preliminary requirements for datasets used for an AI system for image interpretation

The challenges with AI algorithms in veterinary radiology result in training or validating dataset sizes that likely far exceed those found in human radiology. This relates to non-uniformity of diagnostic techniques, geographic distributions of disease and multiple breeds affected to highlight a few. Nonetheless, it is important to stress the requirements of both training and validation datasets to ensure the quality and accurate estimates of performance. What becomes obvious is the need for a large-scale institutional dataset collaboration to support research in veterinary radiology AI.

Training Reference Dataset:

1) Ground truth labelling performed by those who are routinely considered to be the reference standard for the specific problem at hand.
2) Appropriate labeling of the ground truth to reflect the task – e.g. classifying, measuring, segmenting, and must consider the discriminating power and its clinical relevance.
3) Ground truth labelling that is consistent with raters and populations.
4) Datasets should reflect the variances in imaging equipment, procedures and clinical settings.
5) Datasets should consider the disease's clinical spectrum and prevalence.
6) Datasets should reflect an even distribution of findings or measurements of all possible outcomes, normal and abnormal, to avoid model biases and underperformance.
7) At a minimum, the dataset size should reflect the reference test's statistical power for detection of the specified finding or measurement. The larger the dataset, the greater diagnostic accuracy.

Validation Reference Dataset:

1) The normal-to-abnormal ratio should reflect the prevalence of the target pathology in the relevant animal or breed specific population.
2) Reference dataset should be sourced from multiple institutions to introduce data heterogeneity.
3) Age, breed and sex characteristic, as well as health statuses, should reflect the same in the relevant target population.
4) The size of the reference dataset should reflect the statistical considerations and diagnostic accuracy.
5) Datasets used in training should not be used as validation reference datasets, and vice-versa. Data sets need to be independent, with independent developers/curators.