

Artificial intelligence in animal behaviour, veterinary behaviour and neurology

Edited by

Holger Andreas Volk, Jasmin Nicole Nessler, Anna Zamansky, Sebastian Meller and Annika Bremhorst

Published in

Frontiers in Veterinary Science



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-6585-8
DOI 10.3389/978-2-8325-6585-8

Generative AI statement

Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Artificial intelligence in animal behaviour, veterinary behaviour and neurology

Topic editors

Holger Andreas Volk — University of Veterinary Medicine Hannover, Germany

Jasmin Nicole Nessler — University of Veterinary Medicine Hannover, Germany

Anna Zamansky — University of Haifa, Israel

Sebastian Meller — University of Veterinary Medicine Hanover Foundation, Germany

Annika Bremhorst — University of Bern, Switzerland

Citation

Volk, H. A., Nessler, J. N., Zamansky, A., Meller, S., Bremhorst, A., eds. (2025).

Artificial intelligence in animal behaviour, veterinary behaviour and neurology.

Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6585-8

Table of contents

- 04 **Editorial: Artificial intelligence in animal behaviour, veterinary behaviour and neurology**
Jasmin Nicole Nessler
- 07 **ChatGPT and scientific papers in veterinary neurology; is the genie out of the bottle?**
Samira Abani, Holger Andreas Volk, Steven De Decker, Joe Fenn, Clare Rusbridge, Marios Charalambous, Rita Goncalves, Rodrigo Gutierrez-Quintana, Shenja Loderstedt, Thomas Flegel, Carlos Ros, Thilo von Klopman, Henning Christian Schenk, Marion Kornberg, Nina Meyerhoff, Andrea Tipold and Jasmin Nicole Nessler
- 14 **Can ChatGPT diagnose my collapsing dog?**
Samira Abani, Steven De Decker, Andrea Tipold, Jasmin Nicole Nessler and Holger Andreas Volk
- 25 **Machine learning approaches to predict and detect early-onset of digital dermatitis in dairy cows using sensor data**
Jennifer Magana, Dinu Gavojdian, Yakir Menahem, Teddy Lazebnik, Anna Zamansky and Amber Adams-Progar
- 34 **BovineTalk: machine learning for vocalization analysis of dairy cattle under the negative affective state of isolation**
Dinu Gavojdian, Madalina Mincu, Teddy Lazebnik, Ariel Oren, Ioana Nicolae and Anna Zamansky
- 44 **Temporal–spatial variability of grazing behaviors of yaks and the drivers of their intake on the eastern Qinghai-Tibetan Plateau**
Xiaoqian Yang, Umar Daraz, Jianguo Ma, Xingxin Lu, Qingshan Feng, Huaide Zhu and Xiao-Bo Wang
- 56 **Automation in canine science: enhancing human capabilities and overcoming adoption barriers**
Nareed Farhat, Dirk van der Linden, Anna Zamansky and Tal Assif
- 69 **Machine learning algorithms predict canine structural epilepsy with high accuracy**
Thomas Flegel, Anja Neumann, Anna-Lena Holst, Olivia Kretzschmann, Shenja Loderstedt, Carina Tästensen, Sarah Gutmann, Josephine Dietzel, Lisa Franziska Becker, Theresa Kalliwoda, Vivian Weiß, Madlene Kowarik, Irene Christine Böttcher and Christian Martin
- 78 **Research on motion characterization of goose neck in narrow space**
Fu Zhang, Haoxuan Sun, Jiajia Wang, Xinyue Wang, Yubo Qiu, Xiahua Cui and Shaukat Ali
- 90 **Research on herd sheep facial recognition based on multi-dimensional feature information fusion technology in complex environment**
Fu Zhang, Xiaopeng Zhao, Shunqing Wang, Yubo Qiu, Sanling Fu and Yakun Zhang



OPEN ACCESS

EDITED AND REVIEWED BY
Alasdair James Charles Cook,
University of Surrey, United Kingdom

*CORRESPONDENCE
Jasmin Nicole Nessler
✉ jasmin.nessler@tiho-hannover.de

RECEIVED 07 April 2025
ACCEPTED 12 May 2025
PUBLISHED 24 June 2025

CITATION
Nessler JN (2025) Editorial: Artificial
intelligence in animal behaviour, veterinary
behaviour and neurology.
Front. Vet. Sci. 12:1607623.
doi: 10.3389/fvets.2025.1607623

COPYRIGHT
© 2025 Nessler. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Editorial: Artificial intelligence in animal behaviour, veterinary behaviour and neurology

Jasmin Nicole Nessler*

Foundation Clinic for Small Animals, Neurology and Neurosurgery Service, University of Veterinary Medicine Hannover, Hanover, Germany

KEYWORDS

artificial intelligence, machine learning (ML), deep learning, behavior, veterinary neurology

Editorial on the Research Topic

Artificial intelligence in animal behaviour, veterinary behaviour and neurology

Recent technological breakthroughs—ranging from computer vision and sensor-based analytics to robotics and large language models—are transforming veterinary science. This Research Topic brings together nine studies illustrating how these innovations can address both longstanding and emerging challenges in animal health, behavior, and welfare.

A central theme in several of the contributions is precision livestock management. One study proposed a three-phase monitoring system that integrates a multi-part detection network for flock inventory, a facial classification model for identity recognition, and a facial expression analysis network for health assessment (Zhang, Zhao et al.). By refining deep learning architectures, the authors achieved marked gains in detection accuracy and efficiency, underlining how targeted enhancements such as multi-link convolution fusion blocks and re-parameterizable convolution can elevate both performance and speed. Their framework exemplifies how machine learning (ML) solutions can help farmers more efficiently monitor livestock in large-scale or complex settings, potentially reducing labor requirements and supporting better animal welfare outcomes.

Several articles in this Research Topic highlight the synergy between bioinspired engineering and animal-based research. One study, motivated by goose neck biomechanics, designed a bionic robotic arm that prioritizes both flexibility and bearing capacity in confined spaces (Zhang, Sun et al.). By systematically analyzing the goose neck's movements in narrow environments, the authors derived parameters that could inform more advanced and adaptive robotic systems. Their approach bridges animal physiology with practical robotics, opening new possibilities for applications in veterinary procedures, agritech, and beyond.

While precision livestock approaches often focus on optimizing herd management, another cluster of papers underscores how machine learning augments veterinary diagnostics and clinical practice. A large-scale study demonstrated that Bayesian Networks and Random Forests can accurately predict structural epilepsy in dogs (Flegel et al.). This approach, validated through tens of thousands of ML models, showcases how data-driven methods can strengthen clinical decision-making. The authors' success in isolating features highly predictive of structural epilepsy paves the way for improved diagnostic protocols, potentially reducing uncertainties for both clinicians and pet owners.

Expanding on these applications of ML, other researchers explored automated approaches in canine science, reflecting the growing need to analyze complex, high-volume behavioral data (Farhat et al.). By systematically reviewing existing literature and conducting empirical work with animal behavior researchers, they pinpointed both the promise and challenges of automation. Their recommendations could lead to greater acceptance of automated methods, ultimately promoting more robust and reproducible behavioral studies in companion animals.

Meanwhile, in an investigation of grazing behaviors on the Qinghai-Tibetan Plateau, researchers combined field observations with meta-analysis to uncover how temporal and spatial factors shape yak foraging patterns (Yang et al.). Their findings highlight how local climate conditions, forage quality, and livestock traits intertwine to influence grazing behaviors. Such insights can help refine herd management practices, ensuring that rangelands remain sustainably productive under changing environmental pressures.

A complementary line of research explores non-invasive indicators of cattle welfare, specifically vocalization patterns under negative affective states (Gavojdian et al.). By categorizing low- and high-frequency calls, the authors developed deep learning and explainable machine learning frameworks to identify individual cows and gauge stress responses. This work reinforces the importance of acoustic monitoring as part of a precision livestock farming toolkit, offering a tangible route toward earlier detection of distress or discomfort in herd animals.

Another study tackled the early detection and prediction of digital dermatitis—a prevalent cause of lameness in dairy herds—using sensor-based ML models (Magana et al.). The application of automated pipelines, including the Tree-Based Pipeline Optimization Tool (TPOT), shows promise for alerting farmers to emerging disease outbreaks before clinical signs fully manifest. By leveraging subtle shifts in cow behavior, these models could lead to timely interventions that improve overall herd health and productivity.

As these data-driven methods proliferate, the role of generative AI has become a topic of both excitement and caution. One team posed the question: “Can ChatGPT diagnose my collapsing dog?” and delved into how large language models might intersect with veterinary triage or diagnostics (Abani, De Decker et al.). Their findings underscore the need for critical evaluation of AI-suggested “diagnoses,” given that statistical pattern-matching differs fundamentally from true clinical reasoning. In a related article, the authors examined ChatGPT’s potential influences on research in veterinary neurology (Abani, Volk et al.). While acknowledging the efficiencies in literature exploration, they also highlight ethical concerns, such as overreliance on AI-generated texts or the spread of misinformation.

Collectively, these nine contributions exemplify how innovation in artificial intelligence, robotics, and data analytics can reshape the practice of veterinary science, from farm management and clinical diagnostics to animal welfare and fundamental research. While each study addresses a distinct question, they converge on a shared principle: harnessing technology to generate timely, actionable insights that benefit both animals and the humans who care for them. Yet, these advancements are not without challenges: ensuring data quality, building transparent

models, and maintaining ethical oversight remain vital to the responsible adoption of such tools.

We hope this Research Topic serves as a stepping stone for cross-disciplinary collaboration. Whether you are a veterinarian, researcher, engineer, or policy-maker, the findings outlined here illustrate the rapid evolution of veterinary science in our data-driven era. By embracing these methods while engaging in thoughtful discourse on their limitations, we can collectively advance a more sustainable, innovative, and compassionate future for animal care.

Author’s personal thoughts on this editorial

This editorial was created by a large language model (LLM) (1) with the prompt: “Write an editorial... (see acknowledgments below)”. No additional revisions were made on the created text. While good human writing is often characterized by an engaging narrative style and storytelling – even in scientific publications this editorial feels rather flat: While the information extracted from the abstracts provided to the LLM is accurately conveyed and presented in highly readable language, a thematic summary of related abstracts, with a more integrative discussion and a take-home-message would be preferable for a well-crafted editorial. It appears that outstanding writing still relies on the human ability to bring depth and new ideas to the narrative.

Author contributions

JN: Formal analysis, Writing – original draft, Methodology, Writing – review & editing.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. This editorial was created by large language model (LLM) (1) with the prompt: “Write an editorial with referencing the papers below. Find the description on how to write the editorial: Editorials frame the aims and objectives of the research within your Topic, as well as placing its findings in a broader context. Your Editorial should present the contributing articles of the Research Topic but should not be just a table of contents. Editorials should not include unpublished or original data. Editorials have a maximum word limit of 1,000 for Topics with 5–10 articles, and may include 1 figure. The word limit can be increased by 100 words for each additional article in the Topic, up to a maximum of 5,000 words for 50 articles or more. Articles

published within a Research Topic should not be listed in the reference list but rather the intext citation should be hyperlinked directly to the article. Editorials should have the title format: “Editorial: [Title of Research Topic].” This was then followed by the reference APA style and the abstracts—or first introductory paragraphs—of the papers in this Research Topic. No additional revisions were made, allowing readers to observe what a LLM can accomplish.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Reference

1. OpenAI. (2025). *ChatGPT o1 (Large language model) [Software]*. Available online at: <https://openai.com> (accessed April 7, 2025).



OPEN ACCESS

EDITED BY

Daisuke Hasegawa,
Nippon Veterinary and Life Science
University, Japan

REVIEWED BY

Hussein M. El-Husseiny,
Tokyo University of Agriculture and
Technology, Japan

*CORRESPONDENCE

Jasmin Nicole Nessler
✉ jasmin.nessler@tiho-hannover.de

RECEIVED 04 August 2023

ACCEPTED 18 September 2023

PUBLISHED 05 October 2023

CITATION

Abani S, Volk HA, De Decker S, Fenn J,
Rusbridge C, Charalambous M, Goncalves R,
Gutierrez-Quintana R, Loderstedt S, Flegel T,
Ros C, Klopman Tv, Schenk HC, Kornberg M,
Meyerhoff N, Tipold A and Nessler JN (2023)
ChatGPT and scientific papers in veterinary
neurology; is the genie out of the bottle?
Front. Vet. Sci. 10:1272755.
doi: 10.3389/fvets.2023.1272755

COPYRIGHT

© 2023 Abani, Volk, De Decker, Fenn,
Rusbridge, Charalambous, Goncalves,
Gutierrez-Quintana, Loderstedt, Flegel, Ros,
Klopman, Schenk, Kornberg, Meyerhoff,
Tipold and Nessler. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

ChatGPT and scientific papers in veterinary neurology; is the genie out of the bottle?

Samira Abani^{1,2}, Holger Andreas Volk^{1,2}, Steven De Decker³,
Joe Fenn³, Clare Rusbridge⁴, Marios Charalambous¹,
Rita Goncalves⁵, Rodrigo Gutierrez-Quintana⁶,
Shenja Loderstedt⁷, Thomas Flegel⁷, Carlos Ros⁸,
Thilo von Klopman⁹, Henning Christian Schenk¹⁰,
Marion Kornberg¹¹, Nina Meyerhoff¹, Andrea Tipold^{1,2} and
Jasmin Nicole Nessler^{1*}

¹Department of Small Animal Medicine and Surgery, University of Veterinary Medicine Hannover, Hannover, Germany, ²Centre for Systems Neuroscience, University of Veterinary Medicine Hannover, Hannover, Germany, ³Department of Veterinary Clinical Science and Services, Royal Veterinary College, University of London, London, United Kingdom, ⁴Faculty of Health and Medical Sciences, School of Veterinary Medicine, University of Surrey, Guildford, United Kingdom, ⁵Department of Veterinary Science, Small Animal Teaching Hospital, University of Liverpool, Neston, United Kingdom, ⁶Small Animal Hospital, School of Biodiversity, One Health and Veterinary Medicine, University of Glasgow, Glasgow, United Kingdom, ⁷Department of Small Animal Medicine, Leipzig University, Leipzig, Germany, ⁸Memvet Referral Veterinary Center, Palma de Mallorca, Spain, ⁹Department of Neurology, Small Animal Clinic Hofheim, Hofheim, Germany, ¹⁰Department of Neurology, Lüneburg Small Animal Clinic, Lüneburg, Germany, ¹¹AniCura Small Animal Clinic, Trier, Germany

KEYWORDS

ChatGPT, artificial intelligence (AI), machine learning, generative AI, scientific writing, ethics, integrity, plagiarism

Introduction

The new wave of technology known as generative artificial intelligence (AI), while both awe-inspiring and intimidating due to its transformative potential, has become a worldwide phenomenon this year (1). Generative AI refers to a class of AI models that can create new content, such as text, images, sounds, and videos, based on patterns and structures learned from existing data (2, 3).

Since its launch in November 2022, ChatGPT® (OpenAI®, L.L.C., San Francisco, CA, USA), a state-of-the-art language model, has taken the world by storm and captured the attention of the scientific community (GPT stands for “generative pretrained transformer”) (4, 5). The model is trained to predict the next word in a sentence using hundreds of gigabytes of online textual data and has been further fine-tuned with both supervised and reinforcement learning techniques (6, 7). Although models like ChatGPT® can generate highly plausible and sometimes remarkably coherent text in a wide range of contexts, they rely solely on statistical representations of language and lack any understanding of the meanings behind their generated output (8).

Many challenges and concerns arise regarding the dual-use nature of Large Language Models (LLMs), such as ChatGPT®, particularly in scholarly publishing (9). While some believe that journals must take action against such tools to battle the flood of AI-generated manuscripts that could potentially ruin the scholarly publishing industry, others argue that these technologies will break down language barriers and increase global participation in scholarly conversations (10, 11).

As AI chatbots become more integrated into our daily lives, it is predictable that the way internet users find information online will change (12). However, while using AI chatbots makes finding information faster and simpler compared to traditional search methods, there

is still a tendency for such models to “hallucinate” and create inaccurate fictional answers, which can mislead individuals (12, 13). Moreover, unlike search engines, ChatGPT® engages users in conversation or provides a concise answer, rather than directing them to numerous websites that offer a plethora of information from various sources and letting users decide what they trust. Additionally, ChatGPT® (free version, June 2023) does not browse the web to gather real-time information; instead, its knowledge is restricted to the limited dataset it learned until 2021 (5). Consequently, its responses may not reflect the most recent information and are constrained by the content it was trained on OpenAI (5). LLMs like ChatGPT® have applications far beyond search engines. Among others, ChatGPT® is even capable of generating a credible scientific manuscript when provided with the appropriate prompt. Surprisingly, the AI-generated papers were “convincing enough” to initiate debates in academic community concerning the use of such tools in scholarly publishing (14).

Almost everywhere in the research community, the disruptive potential of AI tools is being eagerly debated. There are some observers who worry that in the worst-case scenario AI-generated technologies will increase the output of “pseudo-scientific papers” even faster and cheaper (15, 16). Another major concern related to the utilization of LLMs like ChatGPT® in scientific writing is so called “AI-based plagiarism” or “copyright laundering” (17, 18). Chan (17), define it as *AI-giarism*, referring to “the unethical use of AI to create content that is plagiarized from either original human-authored work or directly from AI-generated content, without appropriately acknowledging the original sources or AI’s contribution.” Concerning this, shortly after the launch of ChatGPT®, some publishers established new policies regarding the disclosure of such tools, while others have taken a step further and announced a complete ban on text generated by ChatGPT® or any other AI tools (19). However, it remains uncertain whether there is currently an enforcement tool capable of consistently detecting AI-generated text (15).

While there has been a rapidly growing academic literature investigating the potential and limitations of ChatGPT® across diverse domains, there is no published work (as of June 2023) on the potential uses and misuses of ChatGPT® and its ethical boundaries in the context of scientific writing within the field of veterinary medicine. First and foremost, this opinion article reflects the authors’ viewpoints and experiences rather than providing a comprehensive, critical review. In the current study, we present a user-based experience to determine if ChatGPT® can generate convincing scientific papers in the field of veterinary neurology. We evaluated the ChatGPT® generated abstracts as well as introduction sections with references for three research papers focusing on different subjects in veterinary neurology. We used an AI output detector and a plagiarism detector to analyse the generated content. Furthermore, we asked thirteen Board-Certified neurologists by the European College of Veterinary Neurology or American College of Veterinary Internal Medicine (subspecialty of Neurology), who have experience in writing and reading

scientific papers, to try to evaluate whether the sets were original or AI-generated. The purpose of this study is to focus on the potential limitations and advantages associated with the application of ChatGPT® in scientific writing within the field of veterinary neurology.

Materials and methods

We selected three research papers authored by the team of the Department of Small Animal Medicine and Surgery at the University of Veterinary Medicine Hannover, Hannover, Germany focusing on different subjects in the field of veterinary neurology, including SARS-CoV-2 scent detection in dogs (20), potential biomarkers for steroid-responsive meningitis-arteritis (21), and staining of cannabinoid receptor type one (22). These articles were published between November 2022 and April 2023 in three different journals: BMJ (23), PLOS ONE (24), and Scientific Reports (25). As stated on the OpenAI® homepage, ChatGPT® does not have access to data beyond 2021 (26). Therefore, we assume that the model has not accessed any of these publications. Given the diverse research interests and familiarity levels of the specialist reviewers with the subjects, the three subjects were classified into three familiarity classes:

In Test 1, the subject of SARS-CoV-2 scent detection in dogs was categorized as less familiar, assuming limited familiarity by the reviewers with this topic.

In Test 2, the subject of potential biomarkers for steroid-responsive meningitis-arteritis in dogs was classified as highly familiar.

Lastly, in Test 3, the staining of cannabinoid receptor type one was classified as a moderately familiar subject.

We hypothesized that reviewers might show different levels of ability in distinguishing between AI-written manuscripts and human-written ones, depending on their familiarity with the subject or research interest. We utilized two different plagiarism detection tools to analyse both the original and AI-generated manuscripts. The first tool used is a paid Internet-based similarity detection service called Turnitin® (27), offering a similarity index ranging from 0 to 100 percent. A higher value on Turnitin indicates a greater level of text redundancy with existing sources. The second tool used is a free online plagiarism checker platform called Plagiarism Detector (28), which provides a plagiarized score ranging from 0 to 100 percent. A higher plagiarized score indicates more detected plagiarism. In addition, we evaluated all the original and generated manuscripts using two different AI-generated content detectors. The first AI-detector is called the AI Detector (29), which provides scores indicating the percentage of human-written content. A score of 100% indicates the absence of any detected AI-written content. The second AI-detector is known as the AI Text Classifier (OpenAI®) (30), which categorizes contents as either very unlikely, unlikely, unclear, possible, or likely to be AI-generated.

Generating the abstract and introduction with references

We utilized ChatGPT® to generate abstracts, introductions, and references for three mentioned scientific papers based on the title, keywords, journal style and characteristics of each author.

Prompts

To generate scientific abstracts and introduction sections with references, we used different prompts. We prompted ChatGPT® with following request “Write an academic abstract with a focus on (subject) in the style of (author characteristics, i.e., position, gender and age) at (university name) for publication in (journal name), (keywords).” Additionally, during a separate chat session, we asked “Write an introduction on (subject), including the following keywords” Subsequently, we requested “Generate 15 references to support the content.”

During our interactions with ChatGPT® for generating scientific papers (May 2023), we noticed that the model sometimes disobeyed the requests. For example, when we prompted it to generate references to support the content, the model answered: “As an AI language model, I don’t have direct access to a specific database or the ability to provide real-time references. I can provide general information and knowledge based on what has been trained on. For accurate and up-to-date references, it’s recommended to consult academic databases, research papers, and scholarly sources.” In contrast, when the prompt was changed to “Generate 15 fictitious references to support the content” the model generated 15 fictitious references and responded with the following sentence: “Please note that these references are fictional and not based on real publications.” Although it feels rather odd to be lectured by an AI model, we acknowledge that the model demonstrates some knowledge of producing fake or inaccurate data. In addition, ChatGPT®’s generated responses are sensitive to the way it is prompted, and it can generate different responses even for the same prompt multiple times (31). To enhance the believability and persuasiveness of the responses generated by ChatGPT®, we prompted it multiple times with additional refinements, such as font and reference styles, within the generated text.

Evaluation

We requested thirteen expert reviewers, all Diplomates of the European College of Veterinary Neurology or American College of Veterinary Internal Medicine (SDD, JF, CIR, MC, RG, RGQ, SL, TF, CaR, TK, HS, MK, NM), to evaluate three tests; all three tests are available in the [Supplementary material](#). Each test contained two sets of abstracts and introductions with references, with one set being written by ChatGPT® and the other by a veterinary neurologist. The reviewers were asked to indicate whether the manuscripts were written by human, generated by an AI algorithm, or if they were unable to determine the source. All specialist reviewers were blinded to the information who generated the text they evaluated. All manuscript headings, as well as any references

containing the names of authors or reviewers involved in this current study, were deleted, to avoid bias in the evaluation. To ensure the accuracy of our study, we requested that reviewers avoid opening all three tests simultaneously for comparison. Additionally, we asked them not to use the internet, AI output detector software, or any external sources such as books or scientific papers during the survey. The process for each of the three tests consisted of the following stages:

Initial assessment stage: abstract identification. The reviewer was requested to read both abstracts and determine which one was written by AI, without referring to the introduction or references.

Final assessment stage: introduction and reference review. The reviewer was asked to read the introduction and references for the corresponding abstract. Considering all the information provided (abstract, introduction, and references), the reviewers were asked to select the manuscript they thought was written by an AI, and to provide supporting evidence and an explanation for their choice.

Results

In Test 1 and Test 3, which were referred to as less familiar and moderately familiar subjects, respectively, only four out of 13 specialist reviewers (31%) correctly identified the AI-generated abstracts (initial assessment stage). In the final assessment stage of both Test 1 and Test 3, which included the introduction and references, nine out of 13 reviewers (69%) successfully identified the generated abstracts and introductions. In Test 1, four reviewers, and in Test 3, five reviewers, who initially could not identify the AI-generated abstracts, identified the manuscripts as being generated by the AI algorithm after reading the introduction and references. In the initial assessment stage of Test 2, which was classified as a highly familiar subject in the field of veterinary neurology, seven out of 13 reviewers (54%) were able to correctly identify the AI-generated abstracts. In the final assessment stage, which included the introduction and references, 10 out of 13 participants (77%) correctly identified the AI-generated manuscript. Furthermore, five reviewers, who initially misidentified the human-written abstract as AI-generated in Test 2, re-evaluated their decision after reading the introduction and references. They subsequently correctly identified the manuscript as being written by a human ([Figure 1](#)).

Specialist reviewers provided various explanations to determine whether the manuscript was generated by AI or written by a person. They commented that AI-generated manuscripts sometimes contain “incredibly human-sounding” texts and might have better English proficiency and structure. As a result, it was occasionally impossible to differentiate the AI-generated manuscripts from the original ones without relying on scientific background information and references. However, some reviewers also pointed out that AI-generated manuscripts were verbose, superficial, and less creative compared to human-written ones.

In addition, both the original and AI-generated manuscripts, were analyzed using the paid Internet-based similarity detection service called Turnitin® (27) and the free online plagiarism checker website called Plagiarism Detector (28). When using Turnitin®, it indicated a similarity index of 100 (higher value indicates more similar text) for all original published manuscripts in Test 1, Test 2, and Test 3. In comparison, the AI-generated manuscripts

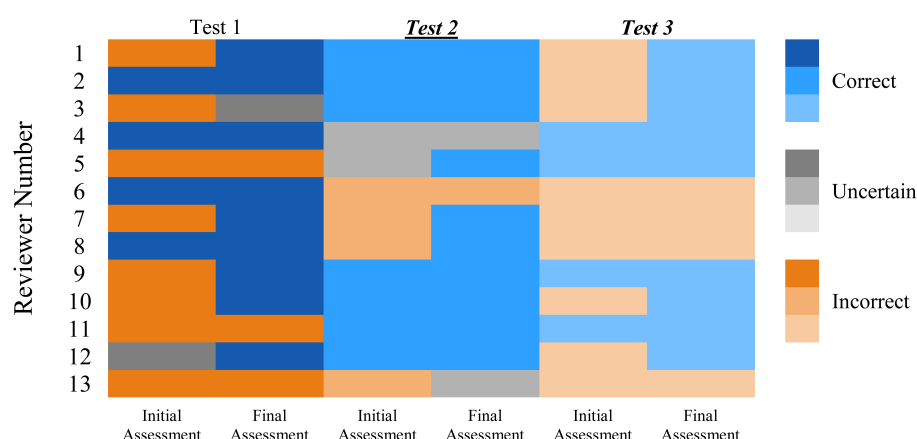


FIGURE 1

Presents the performance of specialist reviewers in identifying ChatGPT®-generated scientific manuscripts vs. original published manuscripts in three tests, each containing two stages; the initial assessment, referred to as abstract detection, and the final assessment, which includes the introduction with references. The reviewers' assessments are represented using colors, with red indicating "Incorrect" (misidentification), green indicating "Correct" (accurate identification), and yellow indicating "Uncertain" (indicating uncertainty). Reviewers demonstrated a better performance in identifying ChatGPT®-generated manuscripts in **Test 2** (the highly familiar subject of steroid-responsive meningitis-arteritis in dogs) vs. **Test 1** (SARS-CoV-2 detection dogs) and **Test 3** (Cannabinoid receptors in canine epilepsy), which were referred to as less familiar and moderately familiar subjects. In the final assessment, reviewers demonstrated improved performance in three tests, indicating that the reviewers were able to refine their assessments and make more accurate conclusions by incorporating specific information and references.

in Test 1, Test 2, and Test 3 had similarity indices of 2%, 8%, and 18%, respectively, on Turnitin®. When utilizing the free online plagiarism checker website, it indicated plagiarism scores of 62%, 74%, and 58% for the original manuscripts in Test 1, Test 2, and Test 3, respectively. In contrast, the website detected 0% plagiarized content in the AI-generated manuscripts of Test 1, Test 2, and Test 3, indicating the potential absence of any detected plagiarism.

When the original manuscripts were run through the online platform AI Detector (29), it indicated that all the original manuscripts contained 0% AI-written content. Among the AI-generated manuscripts, the AI Detector identified AI-generated manuscript in Test 2, indicating that 100% of the content was written by an AI algorithm. However, it identified two other AI-generated manuscripts in Test 1 and Test 3 as having 0% content written by an AI algorithm. The AI Text Classifier (30) indicated that the original manuscript in Test 1 unlikely, in Test 2 and Test 3 was very unlikely to being generated by an AI algorithm. Nonetheless, the AI Text Classifier correctly identified all the abstracts and introductions generated by ChatGPT® as possibly AI-generated.

Discussion

The present study showed that experts in the field increasingly struggled to distinguish between ChatGPT®- and human-written abstracts with a decrease in subject knowledge. In the subject matter which was least familiar, only four out of 13 reviewers correctly identified the text written by AI, whereas when the topic was more familiar, this rate increased to around half, with seven out of 13 reviewers correctly identifying the AI-written

abstract. The expert reviewers did, however, show improvement in accurately detecting the text when they had access to additional sections, including the introduction and reference list. This led to an increase in their performance, with approximately two out of three texts being correctly identified. A software tool, AI Text Classifier (OpenAI®), on the other hand, correctly identified the generated texts. The reviewers commented on different factors to judge whether a manuscript was written by AI or a human, sometimes relying on the same aspect of the text to reach contrasting conclusions. Hence, human reviewers can be tricked by sophisticated AI-generated articles, as LLMs like ChatGPT® are trained to mimic human writing styles and produce coherent, plausible sounding texts. Furthermore, scientific manuscripts may sometimes present poor grammar or readability, when written by humans. Ultimately, relying solely on naturalness, fluency and writing patterns may not always assist reviewers in distinguishing between content written by humans and content generated by AI (32). Specialists comments which highlights this were:

"The English in manuscript B [AI-generated] is better..."
[Reviewer 11, Test 3]

"I found this [Test 1] difficult. I think abstract B [human written] is worded in a slightly more robotic manner, although this could just be concise wording." [Reviewer 2, Test 1]

On the other hand, the inclusion of the introduction, along with references, improved the reviewers' ability to differentiate between human-written and AI-generated articles. Many reviewers' comments focused on the textual content and emphasized that the detailed information helped them judge whether the texts were original or contained fabricated information.

“My only reason to select abstract A [AI-generated] as written by AI is the ‘generic references’; Davis, Johnson, and Smith can appear a bit random, the reference list comes across as potentially legit.” [Reviewer 10, Test 3]

Language models, like ChatGPT[®], could generate convincing scientific manuscripts with fabricated data that might also include fictitious or inaccurate information (see the generated manuscripts in [Supplementary material](#)).

Despite recent improvements, LLMs like ChatGPT[®] are still prone to “hallucination.” Thus, this raises concerns regarding the integrity of utilizing such models in academic writing. Moreover, plagiarism detection tools may not flag AI-generated manuscripts. Since ChatGPT[®] is trained on vast datasets from the internet, some content may include the intellectual property of authors without explicit permission or proper citation (17). Consequently, AI-generative models might create new texts that resemble the original works without acknowledgment, leading to “AI-giarism.” Furthermore, reviewers demonstrated varying levels of confidence and success in distinguishing between AI-written manuscripts and the original published work, depending on their level of subject knowledge. Specialist reviewers found Test 1, which we referred to as the less familiar subject, and Test 3, which we referred to as a moderately familiar subject, more challenging due to their unfamiliarity with these subjects.

“I struggled more with the COVID abstracts [SARS-CoV-2 scent detection in dogs] as it is not a topic I know much about...” [Reviewer 9, Test 1]

“The references were difficult to assess: the specific topic [staining of cannabinoid receptor type one] falls outside my personal field of research interest and seem to originate all from more ‘basic science’ journals.” [Reviewer 10, Test 3]

In contrast, for Test 2, which focused on the highly familiar subject area of steroid-responsive meningitis-arteritis in dogs, specialist reviewers showed better performance in identifying the ChatGPT[®]-generated manuscript. In conclusion, specialist reviewers mentioned that when the manuscript was related to their research field, it was easier to distinguish between human-written and AI-written texts.

“In Abstract B [AI-generated]—SRMA can cause severe neurological deficits—but that is rare. SRMA is only life-threatening if not treated. I don’t think a neurologist would write this sentence as the first introductory sentence.” [Reviewer 11, Test 2]

It seems that AI-based generative tools have already touched the scientific publishing flow and as they continue to improve, they will create opportunities, as well as threats in every aspect of science in the future (33). We believe that AI-models, if trained and tested properly on diverse, reliable, and representative datasets, with a focus on transparency and accuracy, could serve as valuable tools for a wide range of applications in research and publishing. These include AI-tools to assist with semantic search, writing, editing, summarizing papers, statistical analysis, submission screening, citation validation, and peer reviewing (33,

34). In the age of globalization, diversity, equity, and inclusivity bring new ideas, dynamics, and creativity to every domain, and science is not an exception. However, English is a dominant or “standard” language in science, thus can create language barriers for non-native English-speaking researchers who are enthusiastic about contributing to scholarly publishing. Undeniably, publishing in the world’s influential scientific journals requires the privilege of being a so-called “native English speaker.” AI tools like ChatGPT[®] have the potential to address the often neglected challenge of the language barrier in science, offering valuable assistance to non-native English speaker researchers and facilitating the transfer of scientific knowledge in both directions. Some researchers already utilize AI tools such as Grammarly[®] (35), Paperpal[®] (36), Hemingway Editor (37), and Writefull (38) to assist in editing and improving their manuscripts, aiming to increase readability, clarity, and correctness (34, 39).

Language models might become a “game-changer” in scientific writing, enabling researchers to use these models as tools to translate knowledge into their language or improve their English scientific writing skills. This could ultimately be advantageous not only for the researchers but also for the other side of scholarly literature—journals, by reducing the rejection of scientific papers due to poor English.

In the context of the implications of AI technology on scientific writing, journals defined different policies. Some publishers, such as Frontiers (40) and Springer Nature[®] (41), state that AI writing tools like ChatGPT[®] cannot be credited as authors. However, researchers are allowed to use such tools if their usage is appropriately documented in relevant sections.

Limitations

The current study’s findings should be interpreted with caution due to limitations in study design. The reviewers might have read the original papers before the survey was conducted and might have known which abstracts were written by humans. In addition, reviewers knew that each test contained two manuscripts, with one set written by an AI-algorithm and the other by a human. As a result, in real-world scenarios, journal reviewers might show different performance in identifying AI-generated manuscripts. These might be less. Moreover, there are numerous language models available for streamlining scientific writing. However, for this study, we chose to use ChatGPT[®] due to its rapidly growing popularity, as well as its free and user-friendly platform that can be utilized by individuals without requiring any technical expertise (42).

Conclusion

Although this study initiates and opens a conversation regarding the utilization of language models like ChatGPT[®] in the scientific publishing flow in veterinary neurology, the use of AI technology in scientific practice remains debatable in academia. While there is an accurate concern regarding the misuse of AI-based tools in scientific publishing, only simply setting strict boundaries may not always be the most effective way to

prevent malpractice. LLMs can either pose a threat to science integrity and transparency or assist researchers, depending on how they are used. Therefore, we suggest integrating education on both proper use and potential misuse of AI-based tools in academia, as part of good scientific practice for both pre- and post-graduate students in university programs. These training programs could effectively raise awareness and address the ethical considerations associated with AI models in research integrity. The growing popularity of AI requires interdisciplinary scientific collaboration to establish clear and comprehensive guidelines and recommendations for the utilization of AI tools by publishers and journals, thus ensuring the integrity and transparency of the published literature.

Author contributions

SA: Conceptualization, Data curation, Methodology, Writing—original draft, Formal analysis, Software, Visualization. HV: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing—review and editing. SD: Writing—review and editing. JF: Writing—review and editing. CRu: Writing—review and editing. MC: Writing—review and editing. RG: Writing—review and editing. RG-Q: Writing—review and editing. SL: Writing—review and editing. TF: Writing—review and editing. CRO: Writing—review and editing. TK: Writing—review and editing. HS: Writing—review and editing. MK: Writing—review and editing. NM: Writing—review and editing. AT: Conceptualization, Methodology, Supervision, Writing—review and editing. JN: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This publication was funded by the “Central Innovation Program for small- and medium-sized enterprises” grant no. KK5066602LB1—the German Federal Ministry for Economic Affairs and Climate Action, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—491094227 “Open Access

Publication Funding” and the University of Veterinary Medicine Hannover, Foundation.

Acknowledgments

We acknowledge the authors of three selected research articles which we used for this study. Test 1: Canine real-time detection of SARS-CoV-2 infections in the context of a mass screening event; <http://dx.doi.org/10.1136/bmjgh-2022-010276>. Test 2: Immunoglobulin profiling with large high-density peptide microarrays as screening method to detect candidate proteins for future biomarker detection in dogs with steroid-responsive meningitis-arthritis; <https://doi.org/10.1371/journal.pone.0284010>. Test 3: Hippocampal expression of the cannabinoid receptor type 1 in canine epilepsy; <https://doi.org/10.1038/s41598-023-29868-3>.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2023.1272755/full#supplementary-material>

References

1. Wodecki B. UBS: ChatGPT May Be the Fastest Growing App of All Time. (2023). Available online at: <https://aibusiness.com/nlp/ubs-chatgpt-is-the-fastest-growing-app-of-all-time> (accessed August 1, 2023).
2. Callaway E. How generative AI is building better antibodies. *Nature*. (2023) 617:235. doi: 10.1038/d41586-023-01516-w
3. Generative Models. Available online at: <https://openai.com/research/generative-models> (accessed July 21, 2023).
4. Huang K. Alarmed by A.I. Chatbots, Universities Start Revamping How They Teach. *The New York Times*. (2023). Available online at: [nytimes.com](https://www.nytimes.com) (accessed July 7, 2023).
5. OpenAI. Available online at: <https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html> (accessed July 7, 2023).
6. Mearian L. Q&A: ChatGPT isn't Sentient, it's a Next-word Prediction Engine. Available online at: <https://www.computerworld.com/article/3688934/chatgpt-is-not-sentient-it-s-a-next-word-prediction-engine.html> (accessed July 7, 2023).
7. Rudolf J, Tan S, Tan S. ChatGPT: bullshit spewer or the end of traditional assessments in higher education? *J Appl Learn Teach*. (2023) 6. doi: 10.37074/jalt.2023.6.1.9
8. Spitale G, Biller-Andorno N, Germani F. AI model GPT-3 (dis)informs us better than humans. *Sci Adv*. (2023) 9:eadh1850. doi: 10.1126/sciadv.adh1850

9. Van Dis EAM, Bollen J, Van Rooij R, Zuidema W, Bockting CL. ChatGPT: five priorities for research. *Nature*. (2023) 614:224–6. doi: 10.1038/d41586-023-00288-7
10. Thorp HH. ChatGPT is fun, but not an author. *Science*. (2023) 379:313. doi: 10.1126/science.adg7879
11. Berdejo-Espinola V, Amano T. AI tools can improve equity in science. *Science*. (2023) 379:991. doi: 10.1126/science.adg9714
12. Stokel-Walker C. AI chatbots are coming to search engines — can you trust the results? *Nature*. (2023). doi: 10.1038/d41586-023-00423-4
13. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare*. (2023) 11:887. doi: 10.3390/healthcare11060887
14. Quach K. Scientists Tricked into Believing Fake Abstracts written by ChatGPT were Real. (2023). Available online at: https://www.theregister.com/2023/01/11/scientists_chatgpt_papers/?ref=allganize-korea (accessed July 7, 2023).
15. Berinard J. Fake scientific papers are alarmingly common. *Science*. (2023) 380:568–69. doi: 10.1126/science.ad6523
16. Tay A. AI Writing Tools Promise Faster Manuscripts for Researchers. (2021). Available online at: <https://www.nature.com/nature-index/news/artificial-intelligence-writing-tools-promise-faster-manuscripts-for-researchers> (accessed July 7, 2023).
17. Chan CKY. Is AI Changing the Rules of Academic Misconduct? An In-Depth Look at Students' Perceptions of "AI-Giarism". (2023). doi: 10.48550/arXiv.2306.03358
18. Hern A. AI Bot ChatGPT Stuns Academics with Essay-writing Skills and Usability. (2022). Available online at: <https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability> (accessed July 21, 2023).
19. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*. (2023) 613. doi: 10.1038/d41586-023-00191-1
20. Ten Hagen NA, Twele F, Meller S, Wijnen L, Schulz C, Schoneberg C, et al. Canine real-time detection of SARS-CoV-2 infections in the context of a mass screening event. *BMJ Glob Health*. (2022) 7:e010276. doi: 10.1136/bmjgh-2022-010276
21. Nessler JN, Tipold A. Immunoglobulin profiling with large high-density peptide microarrays as screening method to detect candidate proteins for future biomarker detection in dogs with steroid-responsive meningitis-arthritis. *PLoS ONE*. (2023) 18:e0284010. doi: 10.1371/journal.pone.0284010
22. Kostic D, Nowakowska M, Freundt Revilla J, Attig F, Rohn K, Gualtieri F, et al. Hippocampal expression of the cannabinoid receptor type 1 in canine epilepsy. *Sci Rep*. (2023) 13:3138. doi: 10.1038/s41598-023-29868-3
23. *BMJ Journal*. Available online at: <https://journals.bmj.com/home> (accessed July 21, 2023).
24. *PLOS ONE*. Available online at: <https://journals.plos.org/plosone/> (accessed July 21, 2023).
25. *Scientific Reports*. Available online at: <https://www.nature.com/srep/> (accessed July 21, 2023).
26. Raf. Do the OpenAI API Models have Knowledge of Current Events? Available online at: <https://help.openai.com/en/articles/6639781-do-the-openai-api-models-have-knowledge-of-current-events> (accessed July 7, 2023).
27. Turnitin. Available online at: <https://www.turnitin.com/> (accessed July 21, 2023).
28. Plagiarism Detector. Available online at: <https://plagiarismdetector.net/> (accessed July 21, 2023).
29. AI Detector. Available online at: <https://plagiarismdetector.net/ai-content-detector> (accessed July 21, 2023).
30. AI Classifier. New AI Classifier for Indicating AI-written text. Available online at: <https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text> (accessed July 21, 2023).
31. Gao CA, Howard FM, Markov NS, Dyer EC, Ramesh S, Luo Y, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digit Med*. (2023) 6:75. doi: 10.1038/s41746-023-00819-6
32. Clark E, August T, Serrano S, Haduong N, Gururangan S, Smith NA. Human Evaluation of Generated Text. Available online at: www.nltk.org/ (accessed July 21, 2023).
33. Flanagan A, Bibbins-Domingo K, Berkswits M, Christiansen SL. Nonhuman "authors" and implications for the integrity of scientific publication and medical knowledge. *JAMA*. (2023) 329:637–9. doi: 10.1001/jama.2023.1344
34. De Waard A. Guest post – AI and scholarly publishing: a view from three experts. (2023). Available online at: <https://scholarlykitchen.sspnet.org/2023/01/18/guest-post-ai-and-scholarly-publishing-a-view-from-three-experts/> (accessed July 7, 2023).
35. Grammarly. Available online at: <https://www.grammarly.com/> (accessed July 21, 2023).
36. Paperpal. Available online at: <https://paperpal.com/> (accessed July 21, 2023).
37. Hemingway Editor. Available online at: <https://hemingwayapp.com/> (accessed July 21, 2023).
38. Writefull. Available online at: <https://www.writefull.com/> (accessed July 21, 2023).
39. Overcoming the language barrier in science communication. *Nat Rev Bioeng*. (2023) 1:305. doi: 10.1038/s44222-023-00073-1
40. Author Guidelines - AI Use by Authors. Available online at: <https://www.frontiersin.org/guidelines/author-guidelines> (accessed July 21, 2023).
41. AI Authorship. Available online at: <https://www.springer.com/jp/editorial-policies/artificial-intelligence--ai-/25428500#:~:text=While%20legal%20issues%20relating%20to%20AI-generated%20images%20and,have%20created%20images%20in%20a%20legally%20acceptable%20manner> (accessed July 21, 2023).
42. Chow AR. How ChatGPT Managed to Grow Faster Than TikTok or Instagram 4:40 PM EST. (2023). Available online at: <https://time.com/6253615/chatgpt-fastest-growing/> (accessed June 1, 2023).



OPEN ACCESS

EDITED BY

Lisa Alves,
University of Cambridge, United Kingdom

REVIEWED BY

Sam Long,
Veterinary Referral Hospital, Australia

*CORRESPONDENCE

Samira Abani
✉ samira.abani@tiho-hannover.de

RECEIVED 23 June 2023

ACCEPTED 19 September 2023

PUBLISHED 10 October 2023

CITATION

Abani S, De Decker S, Tipold A, Nessler JN and Volk HA (2023) Can ChatGPT diagnose my collapsing dog? *Front. Vet. Sci.* 10:1245168. doi: 10.3389/fvets.2023.1245168

COPYRIGHT

© 2023 Abani, De Decker, Tipold, Nessler and Volk. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Can ChatGPT diagnose my collapsing dog?

Samira Abani^{1,2*}, Steven De Decker³, Andrea Tipold^{1,2},
Jasmin Nicole Nessler¹ and Holger Andreas Volk^{1,2}

¹Department of Small Animal Medicine and Surgery, University of Veterinary Medicine Hannover, Hannover, Germany, ²Centre for Systems Neuroscience, University of Veterinary Medicine Hannover, Hannover, Germany, ³Department of Veterinary Clinical Science and Services, Royal Veterinary College, University of London, London, United Kingdom

KEYWORDS

ChatGPT, artificial intelligence, diagnosis, machine learning, natural language processing, language model, GPT-3.5, generative AI

Introduction

Not long ago, the latest advances in artificial intelligence (AI) were mostly evident to insiders who closely followed the most up-to-date research articles and conference presentations. However, in 2022, generative AI broke into the public consciousness (1). Generative AI refers to a class of AI models that create new data based on statistical probable patterns and structures learned from existing data. The release of text-to-image models like DALL-E 2 and Stable Diffusion, text-to-video systems like Make-A-Video, and specially chatbots like ChatGPT (Chat Generative Pre-trained Transformer) allowed individuals without technical expertise to explore and harness the power of generative AI technology (1–4).

ChatGPT is a Natural Language Processing (NLP) model developed by OpenAI in San Francisco, California that generates text in response to user inquiries (5). ChatGPT is based on the GPT-3.5 architecture, which is a substantial upgrade of the GPT-3 model released by OpenAI in 2020. GPT-3.5 is essentially a smaller version of GPT-3, with 6.7 billion parameters compared to GPT-3's 175 billion parameters (the values that a neural network tries to optimize during training). Despite having fewer parameters, GPT-3.5 produces impressive results in many areas of natural language processing tasks, such as language understanding, text generation, and language translation (2, 4, 6, 7). In comparison to earlier models, ChatGPT is unique as it has been trained on a large dataset from a vast web corpus and has been further fine-tuned for the specific task of generating conversational responses. As a result, it generates human-like responses to user queries or prompts (8–10).

Shortly after its launch, ChatGPT reached 100 million monthly active users, making it the fastest-growing consumer application in history (11). This phenomenal surge not only highlights its effectiveness with diverse tasks but also reflects a widespread and profound curiosity of people wanting to interact with human-like computer interfaces.

Like any other lead innovation, ChatGPT's introduction triggered a range of optimistic and skeptical responses (10, 12–14). The Insider reported that "The newest version of ChatGPT passed the US medical licensing exam with flying colors—and diagnosed a 1 in 100,000 [medical] condition in seconds," (14). Shen et al. (15) stated that "ChatGPT and other Large Language Models (LLMs) may have unintended consequences and become double-edged swords."

Regardless of whether you are an optimist or a skeptic in this debate, it is expected that these language AI models will persist and have an enormous impact on every aspect of society. Consequently, a crucial question arises: Are we prepared for the benefits and challenges presented by emerging AI technologies? (12).

Over the past few years, medical experts have been cautioning the public about the potential consequences of relying on inaccurate health information obtained from internet search engines, such as Google (16). Veterinary surgeons frequently encounter situations where pet owners rely on inaccurate diagnoses obtained from sources like online medical searches, commonly referred to as “Dr. Google,” or misleading information shared by the so called “Facebook Experts” (17, 18). Given the remarkable ability of language models to engage individuals through human-like conversation, it is conceivable that in the future, people may increasingly rely on them to quickly find information and seek medical advice (14).

While “foresight is not about predicting the future,” it is crucial to prepare for probable scenarios. In one possible scenario, following the veterinary consultation, pet owners may use ChatGPT to verify whether the proposed diagnosis is correct and if the treatment is appropriate. They can take it a step further and utilize ChatGPT to obtain possible diagnoses from a list of clinical signs or generate treatment plans, even though it is not specifically designed for these purposes (19). If the diagnosis, treatment, or prognosis contradicts what the veterinarians provided, what impact will it have on the relationship and trust between the pet owner and the veterinarian? Are we prepared to handle such situations with pet owners?

While it cannot be advisable at the moment to rely on ChatGPT for diagnostic applications, in another likely scenario, veterinary surgeons may utilize ChatGPT as a “second eye” in complex situation or as a decision support tool to assist in their clinical reasoning for finding accurate diagnoses, treatment recommendations, or even use it as a search engine to access background information.

Despite recent investigations into the potential benefits and obstacles of LLMs like ChatGPT in healthcare education, research, and practice, there is a lack of data regarding their implementation within the field of veterinary medicine (20–25). This opinion article discusses the potential advantages and challenges associated with integrating of AI-powered chat systems like ChatGPT in veterinary medicine.

Furthermore, here we report, as an example, a user experience of utilizing ChatGPT to diagnose episodic conditions. We used case histories and materials, for which a board-certified neurologist should have no problems to reach a diagnosis. The aim of this publication is to start a discussion about the potentials and threats of using ChatGPT in veterinary medicine. It will not be an all-comprehensive study about all of ChatGPT’s potentials or pitfalls, but our initial thoughts and opinions. We believe that the findings of this opinion article can provide valuable information for veterinary professionals, enabling them to educate pet owners and help them set realistic expectations regarding the use of AI-powered chat systems such as ChatGPT.

Materials and methods

The electronic database of the Small Animal Referral Hospital, Royal Veterinary College, University of London, was retrospectively searched for data of dogs being diagnosed with four disorders characterized by an episodic nature: idiopathic epilepsy,

structural epilepsy, paroxysmal dyskinesia, and syncope. The following inclusion criteria were used for patient database research and were mandatory for case selection: A complete case history, clinical/neurological examination, and final clinical diagnoses.

The diagnosis of idiopathic epilepsy (tier II), and structural epilepsy, was made based on the consensus statement by the International Veterinary Epilepsy Task Force (26). For cases being diagnosed with paroxysmal dyskinesia, the diagnostic approach followed the guidelines outlined in the European College of Veterinary Neurology’s (ECVN’s) consensus statement (27). Furthermore, cases with syncope were included in the study, if they underwent cardiovascular examination and echocardiography as part of the diagnostic process. Twenty cases, five cases for each disorder, were randomly selected, and data, including signalment, case history, and findings from physical/neurological examinations, were extracted from the medical records for stepwise evaluations.

In the first evaluation, the original signalment, as well as the case history of each case were entered into ChatGPT, and the generated responses were recorded. In the second step, separate chat sessions were conducted to input additional physical/neurological examination findings, and the respective responses were also recorded (see the example below).

In the second evaluation, the grammar and choice of words in the medical records were slightly modified without changing the context, in order to assess the reproducibility and overall quality of the AI-generated output. The same approach was conducted again. As ChatGPT utilizes previous interactions with users to generate personalized and tailored responses, a different user, with a distinct IP address, added modified records to ChatGPT (28) (see the example below).

Prompts

ChatGPT’s generated responses are significantly influenced by the way it is prompted, and different questions can lead to different answers. Therefore, throughout both evaluations, the conversation was initiated and concluded in a consistent manner to maintain consistency (29). In the first chat session, the conversation started with “Act as a veterinarian” and then the signalment as well as case history were added. In the second step, the conversation started with “Act as a Board-Certified Veterinary Neurologist in the Queen Mother Hospital Neurology department” and additional clinical examinations were entered. For the second evaluation, the same approach was conducted again.

ChatGPT’s diagnostic performance for identifying cases of idiopathic epilepsy

In the first evaluation using original materials, ChatGPT identified idiopathic epilepsy based solely on case histories in two out of the five cases. In the second step, when clinical examinations were included, it successfully recognized four out of five cases. In one case, it correctly identified idiopathic epilepsy solely based on signalment and case history, and it maintained its accuracy after including additional clinical/neurological

examination. However, in another case, the inclusion of clinical examination information altered its generated diagnosis, resulting in a change of “diagnosis” from idiopathic epilepsy to epilepsy. In three other cases, providing additional clinical/neurological examinations improved the ChatGPT’s diagnostic capabilities, leading to a change in “diagnosis” from epilepsy to idiopathic epilepsy.

In the second evaluation, where the grammar and choice of words from the medical records were slightly modified, in the first step, when relying solely on the case history, ChatGPT did not correctly identify any of the five cases of idiopathic epilepsy. However, in the second step, when clinical/neurological examinations were included, ChatGPT’s performance improved, correctly diagnosing idiopathic epilepsy in three out of the five cases.

The results suggest that the inclusion of clinical/neurological examination improved ChatGPT’s diagnostic capabilities for idiopathic epilepsy. In addition, in one chat, the model emphasized that idiopathic epilepsy is typically observed in young dogs aged between 6 months and 5 years. However, according to the International Veterinary Epilepsy Task Force, idiopathic epilepsy is typically seen in young to middle-aged dogs between 6 months and 6 years old (26). From our perspective, it is concerning that Chat GPT’s generated content lacks scientific literature to support its numerous factual statements, which, despite appearing remarkably plausible, are not always correct. Furthermore, even subtle modifications to the grammar and wording in the medical records had a noticeable impact on its diagnostic capabilities.

An example of a query for a case with idiopathic epilepsy. ChatGPT’s generated answers is in italics

First evaluation utilizing the original medical records

The prompt for the case history

Act as a veterinarian on 3rd October 2006. My dog Milo, is an intact male Border Collie. He is 2 years and 9 months old. Milo has experienced multiple seizure like events since June 24th, 2006. In total, he has had eight events- on August 26th and September 3rd, 2006, he had two and three episodes within a 24-h period. These episodes typically happen when Milo is at rest and last ~10-15 seconds. During these episodes, he shows limb rigidity, paddling, and a vacant expression and distinctive facial expression with his ears back. Following each episode, Milo appears distressed and he is unsteady on his feet for about one to two minutes. I have noticed some mild changes in his behavior as Milo being more passive around other dogs. However, this might coincide with his return from a two-month stay in kennels. Otherwise, Milo has been in good health, and there is no known exposure to toxins. What is the most likely diagnosis?

ChatGPT’s “Diagnosis”

Based on the clinical signs described, the most likely diagnosis for Milo would be epilepsy. Epilepsy is a neurological

disorder characterized by recurrent seizures. The seizure episodes that Milo has experienced, including the limb rigidity, paddling, vacant expression, and distinctive facial expression with ears back, are all consistent with seizures. The fact that the seizures occur mostly at rest and that Milo is unsteady on his feet after the episodes also support this diagnosis.

Although the cause of epilepsy is often unknown, it can be inherited or caused by brain injury or disease. In some cases, epilepsy may be triggered by environmental factors such as toxins, but as you mentioned that there was no known exposure to toxins, this is less likely to be the cause. It is important to note that there are other conditions that can cause seizures, such as metabolic disorders, infections, and brain tumors, so further diagnostic tests may be required to rule out these possibilities. Milo’s veterinarian may recommend a complete physical and neurological examination, blood tests, and possibly imaging tests such as a CT or MRI scan to help confirm the diagnosis and rule out any underlying causes. Treatment for epilepsy typically involves medications such as anticonvulsants to help control the seizures. Your veterinarian may also recommend making changes to Milo’s diet or lifestyle to help manage his condition. It is important to work closely with your veterinarian to develop a comprehensive treatment plan for Milo to help minimize the frequency and severity of his seizures and improve his overall quality of life.

The prompt for the case history and neurological examination

Act as a Board-Certified Veterinary Neurologist in the Queen Mother Hospital Neurology department on 3rd October 2006, your patient is a lovely dog Milo, he is a 2 years and 9 months old male - entire Border Collie. He presented to the QMHA after experiencing several seizure-like episodes that started on the 24th of June 2006. Since then, he has had a total of eight known seizures, including two episodes of cluster seizures that occurred on the 26th of August and the 3rd of September 2006. The owner describes limb rigidity, paddling, and a vacant expression and distinctive facial expression with his ears back, lasting around 10–15 s. These episodes occur mostly at rest, and are followed by a post-ictal period of one to two minutes when Milo is distressed and unsteady on his feet. The owner has also noticed some mild behavioral changes, such as Milo being more passive around other dogs, although this may have coincided with his return from a two-month stay in kennels. He has been otherwise well in himself and there was no known exposure to toxins. During the physical examination, Milo appeared bright and alert and was in good body condition. All parameters were within normal limits, except for mild proprioceptive deficits on the left hind leg. Postural reactions such as hemistanding were also very slightly reduced in the left hind leg. The neurolocalisation was right forebrain. What is the most likely diagnosis?

ChatGPT’s “Diagnosis”

Based on the clinical signs and physical examination findings, the most likely diagnosis for Milo is idiopathic epilepsy. The classic presentation of seizures characterized by limb rigidity, paddling,

and a vacant expression lasting around 10–15 s, followed by a post-ictal period of one to two minutes, is consistent with seizures of an epileptic nature. The cluster seizures also suggest a diagnosis of idiopathic epilepsy. The mild proprioceptive deficits and postural reaction abnormalities on the left hind suggest a right forebrain lesion, which can be seen with idiopathic epilepsy.

However, other differential diagnoses should be considered, such as structural brain lesions, infectious diseases, metabolic disorders, and toxicities. An MRI or CT scan would be necessary to rule out structural brain lesions. Additionally, blood work and cerebrospinal fluid analysis should be performed to assess for infectious diseases and metabolic disorders. Nevertheless, the most likely diagnosis based on the clinical presentation is idiopathic epilepsy.

Second evaluation utilizing medical records with minor grammar and word choice adjustments

The prompt for the case history

Act as a veterinarian on 3rd October 2006. My dog, Milo, is a 2¾-year-old male Border Collie, and he hasn't been neutered. Since June 24th, 2006, Milo has had several episodes that seem like seizures, probably totaling eight in number. I believe he had multiple seizures on August 26th and September 3rd, 2006. During these, he experienced two seizures within a 24-hour period on August 26th and three seizures on September 3rd. Typically, these episodes occur when Milo is at rest and last around 10 to 15 seconds each. During these seizures, he becomes rigid in his limbs, starts paddling, and his ears are pulled back. After each seizure, Milo appears distressed and struggles to stand steadily for about one to two minutes. I've also observed some slight changes in his behavior, particularly when he's around other dogs, where he seems more passive than before. However, it's worth noting that this change might coincide with his return from a two-month stay in kennels. Apart from these episodes, Milo has been generally healthy, and there's no known exposure to any harmful substances. What is the most likely diagnosis?

ChatGPT's "Diagnosis"

Based on the information provided, the most likely diagnosis for Milo's symptoms would be epileptic seizures. Seizures in dogs can present in various ways, and the symptoms you've described, such as becoming rigid, paddling, and distressed behavior afterward, are consistent with seizure activity. Cluster seizures, where multiple seizures occur in a short period, can also be a sign of epilepsy. It's important to consider other potential causes as well, such as metabolic issues, toxins, or other neurological conditions. A thorough physical examination, along with neurological and possibly blood tests, could help the veterinarian rule out other possible causes.

Since there's a mention of a change in behavior coinciding with Milo's return from kennels, stress and environmental changes might also be contributing factors. Changes in routine

or environment can sometimes trigger stress-related symptoms in dogs.

Given the complexity of the situation, your best course of action would be to schedule an appointment with a veterinarian as soon as possible. They can perform a comprehensive evaluation, including a detailed medical history and potentially additional diagnostic tests, to determine the underlying cause of Milo's seizures and recommend an appropriate treatment plan.

The prompt for the case history and neurological examination

Act as a Board-Certified Veterinary Neurologist in the Queen Mother Hospital Neurology department on 3rd October 2006. Your patient is Milo, a 2¾-year-old intact male Border Collie, was brought to the QMH due to multiple episodes of seizures that began on June 24th, 2006. He has experienced a total of eight known seizures so far, including two instances of cluster seizures on August 26th and September 3rd, 2006, where he had two and three seizures, respectively, within a 24-h period. The owner reports that during these episodes, Milo displays limb stiffness, paddling movements, a vacant expression, and a distinctive facial expression with his ears pulled back. Each episode lasts ~10–15 s and primarily occurs when Milo is at rest. Following the seizures, there is a post-ictal period lasting one to two minutes during which Milo appears distressed and has difficulty maintaining balance. The owner also noticed some mild changes in Milo's behavior, such as him being more passive around other dogs. However, it's worth noting that this change coincided with Milo returning home after a two-month stay in kennels. Milo has been generally healthy, with no known exposure to toxins. During the physical examination, Milo was alert and had a good overall body condition score. All measured parameters fell within normal ranges. The examination revealed mild deficits in proprioception in his left hind leg and slight reductions in postural reactions, including hemistanding, in the same leg. The neurolocalization indicated involvement of the right forebrain. What is the most likely diagnosis?

ChatGPT's "Diagnosis"

Based on the information provided, the most likely diagnosis for Milo's condition is idiopathic epilepsy. Idiopathic epilepsy is a common neurological disorder in dogs characterized by recurrent seizures with no underlying structural or metabolic cause identified. The history of Milo's seizures, their frequency and characteristics, as well as the clinical signs he displays during and after the episodes, all align with the typical presentation of epilepsy.

Key points supporting the diagnosis of idiopathic epilepsy in Milo's case include:

- Age of Onset:** Milo is a 2¾-year-old dog, which falls within the typical age range for the onset of idiopathic epilepsy in Border Collies.
- Recurrent Seizures:** Milo has experienced multiple episodes of seizures, including cluster seizures, which is a common pattern seen in dogs with epilepsy.

3. **Seizure Characteristics:** The description of Milo's seizures, including limb stiffness, paddling movements, vacant expression, and distinctive facial expression with ears pulled back, is consistent with generalized tonic-clonic seizures often seen in epilepsy.
4. **Duration of Seizures:** The duration of Milo's seizures, lasting approximately 10-15 seconds, is in line with the typical duration of epileptic seizures.
5. **Post-Ictal Period:** The post-ictal period of distress and balance difficulties Milo experiences after the seizures is a characteristic feature of epilepsy.
6. **Behavioral Changes:** The mild behavioral changes, such as increased passivity around other dogs, could be associated with the effects of seizures or the post-ictal period.
7. **Normal Physical Examination:** The normal physical examination and blood parameters suggest that there might not be an underlying metabolic or structural cause for the seizures.
8. **Neurolocalization:** The involvement of the right forebrain, as indicated by the clinical signs and neurological examination findings, is consistent with a diagnosis of epilepsy.

Given Milo's history, presentation, and the absence of other significant findings, idiopathic epilepsy is the most likely diagnosis. It is important to note that a diagnosis of idiopathic epilepsy is one of exclusion, meaning that other potential causes for seizures should be ruled out before confirming the diagnosis. Additional diagnostic tests, such as blood work, imaging (such as an MRI), and possibly cerebrospinal fluid analysis, might be considered to rule out other potential causes and further support the diagnosis of idiopathic epilepsy. Treatment options and management strategies can then be discussed with the owner to help control Milo's seizures and improve his quality of life.

ChatGPT's diagnostic performance for identifying cases of structural epilepsy (intracranial neoplasia)

In the first evaluation using original materials, ChatGPT identified brain tumors as a potential differential diagnosis in four out of five cases, and structural abnormalities (without specifying brain tumors) in one out of five cases. ChatGPT also listed metabolic disorders, intoxication, infection, encephalitis, and other conditions as potential differential causes. In the second step, when clinical/neurological examinations were included, the model identified brain abnormalities in two out of five cases, brain tumors in two out of five cases and forebrain lesions in one out of five cases as a potential differential diagnosis.

In the second evaluation, where the grammar and choice of words from the medical records were slightly modified, in the first step, when relying solely on the case history, ChatGPT identified brain tumor as a potential diagnosis in one case out of five cases, and epilepsy in three out of five cases and idiopathic epilepsy one out of five cases. It also listed

epilepsy, infectious diseases affecting the nervous system, or metabolic disorders as potential differential causes. However, in the second step, when clinical examinations were included, in one case out of five the inclusion of clinical examination information altered its generated diagnosis, resulting in a change of differential diagnosis from brain tumors to neurological disorder. In one case out of five, providing additional clinical/neurological examinations improved the ChatGPT's diagnostic capabilities, leading to a change in suggesting the "diagnosis" from epilepsy to brain tumor.

These results reveal that while ChatGPT may be able to generate potential differential diagnoses, and the output appears plausible on the surface, a closer investigation reminds us that ChatGPT is neither a thinking machine nor an AI model with medical-specific training. When considering the inclusion of a slightly modified materials, it failed to accurately diagnose conditions and generated erroneous answers (Table 1).

ChatGPT's diagnostic performance for identifying cases of paroxysmal dyskinesia

When relying solely on the case history, ChatGPT did not correctly diagnose any of the five cases with paroxysmal dyskinesia. Similarly, in the second step, which included clinical/neurological examinations, the model did not diagnose any of the five cases.

In the second evaluation, where the grammar and choice of words from the medical records were slightly modified, in the first step, when relying solely on the case history, ChatGPT identified paroxysmal dyskinesia as a potential differential diagnosis in one out of five cases. In the second step, when clinical/neurological examinations were included, it changed differential diagnosis from diagnosing paroxysmal dyskinesia to idiopathic epilepsy. Nevertheless, it correctly diagnosed paroxysmal dyskinesia in one out of the five cases.

The limited diagnostic ability of ChatGPT in identifying cases with paroxysmal dyskinesia can be linked to two reasons. Firstly, the identification and characterization of canine paroxysmal dyskinesias pose significant challenges due to their infrequent and unpredictable nature (27). Since these episodes may not occur while the dog is at the veterinary clinic, veterinarians face difficulties in directly observing and diagnosing them (27). Secondly, there is lack of comprehensive and well-defined information about this disorder in the veterinary literature on internet. It is important to note that ChatGPT's access to data only extended until 2021, which further restricts its recognizing capabilities for paroxysmal dyskinesia. The model may not be up to date with the latest advancements and insights, such as breed-specific paroxysmal dyskinesias in veterinary medicine (30).

ChatGPT's diagnostic performance for identifying cases of syncope

In the first evaluation using original materials, ChatGPT identified cardiovascular syncope as a potential differential

TABLE 1 Representative examples of ChatGPT-generated diagnoses compared with the final diagnosis.

Case number	Final diagnosis at RVC	ChatGPT's generated diagnoses in the first evaluation		ChatGPT's generated diagnoses in the second evaluation	
		With case history	With case history and clinical/neurological examination	With case history	With case history and clinical/neurological examination
01	IE (Tier II)	Epilepsy	IE	Epilepsy, metabolic disorders, infectious diseases, or structural abnormalities in the brain	Seizure
02	IE (Tier II)	IE	IE	Epilepsy, metabolic disorders, infectious diseases, or structural abnormalities in the brain	IE
03	IE (Tier II)	IE	Epilepsy	Epilepsy	IE
04	IE (Tier II)	Epilepsy	IE	Epilepsy	Generalized seizure
05	IE (Tier II)	Epilepsy	IE	Epileptic seizures	IE
06	PD	Seizure	Focal seizure	Seizure-DDx: epilepsy, metabolic disorders, toxicities, infections, or other neurological conditions	Movement disorder
07	PD	Neurological disorder	IVDD	Neurological issues, musculoskeletal problems, injury, infectious or inflammatory conditions	DDx: IVDD, spinal cord injury, metabolic, toxicology, Infectious or inflammatory process
08	PD	Seizure	Seizure	IE	Diagnosis is not possible without further diagnostic tests (bloodwork, imaging or CSF)
09	PD	IE	IE	Canine epilepsy or another form of seizure disorder	IE
10	PD	IE	IE	Neurological disorder, such as seizures or a movement disorder	IE
11	Syncope-primary DCM	Syncope-DDx: collapsing trachea, cardiac arrhythmia or a congenital heart defect, brachycephalic airway syndrome	Exercise-induced collapse syndrome	Syncope or respiratory issues	Syncope
12	Syncope-pulmonary hypertension	Seizure	Syncope-heart disease	Syncope	Difficult to determine without further diagnostic tests, probably a potential neurological issue
13	Syncope-hepatic mass-IMHA-pulmonary hypertension (suspect pulmonary thromboembolism)	Syncope-DDx: seizure, cardiac issues, neurological problems such as brain tumors, vestibular disease	PSS with episodes of hepatic encephalopathy	Syncope	Diagnosis not clear, collapsing episodes with potential cardiovascular or hepatic involvement
14	Syncope-sick sinus syndrome	Syncope-DDx: most likely cardiovascular issues, such as heart disease or arrhythmias, neurologic conditions, or metabolic disorders	Syncope-DDx: cardiac arrhythmias, heart disease, neurological conditions, or other systemic disorders	Syncope	Syncope

(Continued)

TABLE 1 (Continued)

Case number	Final diagnosis at RVC	ChatGPT's generated diagnoses in the first evaluation		ChatGPT's generated diagnoses in the second evaluation	
		With case history	With case history and clinical/neurological examination	With case history	With case history and clinical/neurological examination
15	Syncope-angiostrongylus vasorum-pulmonary hypertension	Syncope-lungworm infection residual damage or complications	Syncope-arrhythmia	Syncope-relapse or complications related to lungworm disease	Syncope-cardiovascular issue
16	Stre (neoplasia)	Seizures-DDx: metabolic disorders, brain tumors, or other neurological conditions	Seizure-DDx: structural brain abnormalities or metabolic disorders	Epilepsy	Epilepsy
17	StrE (neoplasia)	Seizures-DDx: metabolic disorders, toxicity or structural abnormalities	Epilepsy-forebrain lesion	IE	IE
18	StrE (neoplasia)	Epilepsy-DDx: brain tumors, infections, metabolic disorders, or toxicities	Idiopathic epilepsy, brain tumors or inflammatory brain diseases	Epilepsy	IE
19	StrE (neoplasia)	Epilepsy-DDx: brain tumors, toxins, metabolic disorders, or infections	Focal seizure-DDx: structural abnormalities, metabolic disturbances, infectious diseases, toxins, or idiopathic (unknown) reasons.	Epilepsy	Epilepsy-DDx: brain tumor or inflammatory brain disease
20	StrE (neoplasia)	Seizure-DDx: brain tumors, encephalitis, metabolic disturbances, toxicity, or another neurological conditions	Tumor, specifically in the right forebrain	Epilepsy, brain tumors, infectious diseases affecting the nervous system, or metabolic disorders	Neurological disorder

IE, idiopathic epilepsy; PD, paroxysmal dyskinesia; StrE, structural epilepsy; DDx, differential diagnosis; IVDD, intervertebral disc disease; CSF, cerebrospinal fluid; DCM, dilated cardiomyopathy; IMHA, immune-mediated hemolytic anemia.

diagnosis based solely on case histories in four out of five cases. ChatGPT also listed collapsing trachea, brachycephalic airway syndrome, seizure, neurological problems such as brain tumors, vestibular disease, and or metabolic disorders as potential differential causes. In the second step, when clinical examinations were included, the model identified cardiovascular syncope as a potential differential diagnosis in three out of five cases. In two out of five cases, when additional clinical examination results were included, the model changed the potential differential diagnosis from cardiovascular syncope to exercise-induced collapse syndrome and portosystemic shunt with episodes of hepatic encephalopathy.

In the second evaluation, when grammar and choice of words from the medical records were slightly modified, ChatGPT correctly diagnosed all of the five syncope cases when only examining the case history. However, in the second step,

when clinical examinations were included, ChatGPT's generated diagnoses changed; it correctly identified cardiovascular syncope as a potential diagnosis in four out of five cases. In one case model ChatGPT responded "difficult to determine without further diagnostic tests, probably a potential neurological issue."

Although it can be quite challenging to differentiate cardiovascular syncope from seizure activity and other causes of collapse, ChatGPT has demonstrated sufficient performance in identifying cardiovascular syncope based solely on the case history. However, the inclusion of clinical examination information altered ChatGPT's generated diagnoses for cardiovascular syncope. This finding suggests that due to the lack of both medical-specific training in this model and a clinical reasoning algorithm, the excessive inclusion of physical exam findings for other organs could introduce confounding factors that may confuse and hinder its diagnostic abilities.

Pitfalls and challenges

Bias

Despite considerable potentials of LLMs technology for research and clinical applications, there are redoubtable challenges and risks particularly in terms of validating these models for integration into animal healthcare. Veterinary professionals must carefully consider the potential biases that can arise from the limited datasets used to train ChatGPT. These biases limit its capabilities and have the potential to result in factual inaccuracies. What is particularly concerning is that these biases may appear scientifically plausible, which is a phenomenon known as “hallucination” (25). Indeed, when veterinary surgeons excessively rely on ChatGPT’s responses, potential erroneous outcomes can lead to serious consequences for patient care.

Veterinary surgeons, being aware of the biases, should be able to identify if the information, “diagnosis” or treatment provided does not only sound plausible, but also is plausible for the individual patient. The owners on the other hand might be misled by the information provided by ChatGPT. The current study highlighted that despite ChatGPT providing a reasonably logic sounding reply, it was incorrect in quite a few cases with its diagnostic “judgement.” In comparison to “Dr Google,” which provides links to information and brief summaries, ChatGPT provides a more personalized and plausible sounding reply, which can be misleading and will make it even more difficult for non-medical trained individuals such as owners to differentiate between correct and incorrect information. Moreover, as stated on the OpenAI homepage, “ChatGPT is sensitive to tweaks in the input phrasing or attempts with the same prompt multiple times. For example, when presented with one phrasing of a question, the model might claim not to know the answer, but with a slight rephrase, it can answer correctly” (31). Hence, as observed in the current study, the effectiveness of this generative AI model for clinical applications might be hindered by its inability to reproduce consistent results (Table 1 illustrates examples of AI-generated diagnoses from two evaluations performed by different users).

The phrase “garbage in, garbage out” concisely describes the concept that the quality of the output generated by an AI model is directly correlated with the quality of the data it is trained on (32). To address the propensity of language models for hallucinations and routine biases, some studies emphasize the potential of training domain-specific language models, while others propose augmenting LLMs with domain-specific external tools for specific medical tasks (33, 34). Nevertheless, a version of ChatGPT as a Veterinary Support System would need to be trained and validated based on current, reliable scientific data, such as textbooks, academic literature, as well as comprehensive collection of medical records from multiple institutions. The model should also provide citations, ensuring that the information provided is accurate and up to date before its integration into clinical practice.

During our interactions with ChatGPT, we have consistently noticed that the diagnoses provided by the model always conclude with a recommendation to visit a veterinarian for a comprehensive diagnostic evaluation. Additionally, we consider OpenAI’s acknowledgment of potential limitations, such as the use of outdated data and the potential for bias in generated

responses, to be an essential step in demonstrating a commitment to advancing AI technology responsibly (31). This was also highlighted in the current study, where ChatGPT performed markedly worse to “diagnose” paroxysmal dyskinesia compared to the other conditions. The poorer performance could be explained in parts as the data used to train ChatGPT did not include the rapidly increasing number of publications about paroxysmal dyskinesia. The data for the current version of ChatGPT included information until 2021.

Privacy

Additionally, privacy concerns are an important consideration when using an AI tool. According to the OpenAI website, the content provided by users is actively and continuously collected to improve the service or conduct research (35). ChatGPT utilizes both supervised learning and reinforcement learning, incorporating human feedback in its fine-tuning process. This approach has the potential to enhance its capabilities and progressively offer users more relevant and accurate assistance over time (2, 28). It is advisable to be cautious when entering sensitive data like patients’ information on Chat GPT to ensure data security and privacy.

Potential advantages

Clinical practice and research

The integration of AI technologies in clinical research and decision making could be highly advantageous due to their capability to collect and analyse large amounts of data (36).

Language models like ChatGPT can assist e.g. clinicians to summaries case history, and clinical data to improve efficiencies in clinical decision making. ChatGPT can also help in finding relevant clinical scientific background information and help researchers across various stages of the research process, from study design to scientific literature writing.

These models can or will be able in the future to efficiently analyze and summarize concisely vast amounts of scientific literature, identify relevant studies, highlight research gaps, and extract information that goes beyond the expertise of an individual researcher (20). They can accelerate data collection, automate processes such as summarizing patient data and extracting information from diverse sources. Additionally, these models have the capability to assist in manuscript writing by generating well-structured drafts that conform to journal guidelines (37). However, as documented in the current study, clinicians and researchers need to be trained to use the tool, e.g., how to prompt and interpret the often plausible written text appropriately, and to be aware of its limitations (see Limitation Section).

Clinical documentation and communication are a vital component of good clinical practice and patient care (38). The challenges concerning the time and accuracy of clinical documentation is not a new dilemma. Interns and residents often spend long hours outside of office hours on documentation, a phenomenon known as “pajama time,” which is associated

significantly with burnout (39–41). Furthermore, these documentations may often contain numerous pages of extraneous information, which can lead to overlooking key aspects (42). Large generative LLMs present a unique opportunity to assist veterinary clinicians with these hidden, time-consuming administrative tasks in their day-to-day workflow by generating high quality clinical documentation and discharge summaries in real-time.

LLMs as clinical decision support systems and remote diagnostic solutions

Clinical Decision Support Systems based on LLMs have the potential to utilize patient histories, physical findings, laboratory and imaging results to suggest or revise differential diagnoses or recommend complementary tests for further confirmation or ruling out of diseases and constructing therapy plans (43).

Based on our interaction with ChatGPT, we believe that despite its limitations and the fact that it is not designed to answer veterinary practice questions, the performance of language models like ChatGPT represents a significant improvement over using Google search, even without any content-specific training. Therefore, we are cautiously optimistic about the future potential of utilizing LLMs in our field to improve clinical decision-making and optimize the overall clinical workflow. Additionally, integrating AI-enabled chatbot-based symptom checker applications could improve accessibility and support users with self-triaging especially during periods of limited veterinary services, such as night hours or weekends (44).

Training and clinical care

Even though continuous training for veterinary clinicians beyond their increasingly specialized fields is desirable and crucial to ensure the best practice, staying up to date in all areas of expertise is challenging and cannot be guaranteed. LLMs could serve as valuable resources for veterinary clinicians, facilitating the evaluation and integration of up-to-date scientific articles and guideline recommendations, and ultimately improving clinical routines.

Owner education

Language Models could enable the instant generation of personalized patient education materials that cover a wide range of topics, including diet, medication usage, and potential side effects. These materials could provide comprehensive information to clients in a concise and easily understandable manner. By utilizing Language Model, veterinarians can enhance efficiency and increase client satisfaction by minimizing post-visit inquiries (45).

Limitations

The current study has several limitations in terms of its study design, and caution must be taken when attempting to generalize its findings. The study relied on information obtained

from an electronic database of Small Animal Referral Hospital, Royal Veterinary College, University of London that combines notes from owner reports, and summaries from veterinary students and clinicians. As a result, this data source may not accurately reflect how pet owners describe their pets' clinical signs in real-world scenarios. Pet owners are not one homogenous group, but a heterogeneous group varying in educational and socioeconomic background. It will therefore be difficult to conduct a truly representative study of how it is used by the general public for various conditions. As soon these AI tools will become part of general internet search engines, then there will be another leap in their usage. Then enhanced tools similar to for example google analytics might provide us with further and deeper insights.

Since the results generated by ChatGPT are highly sensitive to factors such as the presentation of information and the specific wording of questions, it is predictable that the results in everyday situations may vary significantly from those observed in this study. This review not only presents the challenges and opportunities associated with LLMs like ChatGPT in veterinary medicine but also highlights the importance of conducting further research to investigate the best practices for integrating such models in veterinary medicine. Furthermore, training of veterinary students, professionals and owners will be required to overcome the above highlighted limitations. In this regard, the findings presented in this review should only serve as a starting point for further exploration and discussions.

Author contributions

SA, HV, and JN designed the study, with input from AT and SD. HV and SA identified the cases to be used for analysis. SA performed the analysis and wrote the first draft of the manuscript. All authors reviewed, modified sections as appropriate, and have approved the final version of the manuscript.

Funding

This work was funded by the Central Innovation Program for small- and medium-sized enterprises grant no. KK5066602LB1 - the German Federal Ministry for Economic Affairs and Climate Action, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—491094227 Open Access Publication Funding and the University of Veterinary Medicine Hannover, Foundation.

Acknowledgments

We would like to express our gratitude to the veterinarians at Small Animal Referral Hospital, Royal Veterinary College, University of London, whose invaluable contributions made this study possible.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

- Maslej N, Fattorini L, Brynjolfsson E, Etchemendy J, Ligett K, Lyons T, et al. *The AI Index 2023 Annual Report*. Stanford, CA: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University (2023).
- Ray PP. A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Int Things Cyb Phys Sys.* (2023) 3:121–54. doi: 10.1016/j.iotcps.2023.04.003
- Zhou C, Li Q, Li C, Yu J, Liu Y, Wang G, et al. A comprehensive survey on pretrained foundation models: a history from BERT to ChatGPT. *arXiv:2302.09419*. doi: 10.48550/arXiv.2302.09419
- Rudolph J, Tan S, Tan S. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *J Appl Learn Teach.* (2023) 6:1. doi: 10.37074/jalt.2023.6.1.9
- Peres R, Shreier M, Schweidel D, Sorescu A. On ChatGPT and beyond: how generative artificial intelligence may affect research, teaching, and practice. *Int J Res Mark.* (2023) 3:1. doi: 10.1016/j.ijresmar.2023.03.001
- Gill SS, Kaur R. ChatGPT: vision and challenges. *Int Things Cyb Phys Sys.* (2023) 3:262–71. doi: 10.1016/j.iotcps.2023.05.004
- Brown BT. *Language Models are Few-Shot Learners.* (2020). doi: 10.48550/arXiv.2005.14165
- Cotton DRE, Cotton PA, Shipway JR. Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innovat Edu Teach Int.* (2023) 4:1–12. doi: 10.1080/14703297.2023.2190148
- Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? *Lancet Infect Dis.* (2023) 23:405–6. doi: 10.1016/S1473-3099(23)00113-5
- Zhuo TY, Huang Y, Chen C, Xing Z. Red teaming ChatGPT via Jailbreaking: bias, robustness, reliability and toxicity. Technical Report. *arXiv preprint arXiv:2301.12867*. (2023). doi: 10.48550/arXiv.2301.12867
- Chow AR. *How ChatGPT Managed to Grow Faster Than TikTok or Instagram.* Time. (2023). Available online at: <https://time.com/6253615/chatgpt-fastest-growing/> (accessed June 1, 2023).
- Kasneji E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn Indiv Diff.* (2023) 103:102274. doi: 10.1016/j.lindif.2023.102274
- Deng J, Lin Y. The benefits and challenges of ChatGPT: an overview. *Front Comp Intell Sys.* (2023) 2:81–3. doi: 10.54079/fcis.v2i2.4465
- Brueck H. *The Newest Version of ChatGPT Passed the US Medical Licensing Exam with Flying Colors-and Diagnosed a 1 in 100,000 Condition in Seconds.* (2023). Available online at: <https://www.businessinsider.nl/the-newest-version-of-chatgpt-passed-the-us-medical-licensing-exam-with-flying-colors-and-diagnosed-a-1-in-100000-condition-in-seconds/> (accessed June 1, 2023).
- Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. *Radiology.* (2023) 307:e230163. doi: 10.1148/radiol.230163
- “Don't google it” Belgian campaign of the year. Available online at: <https://dadbbrussels.prezly.com/dont-google-it-belgian-campaign-of-the-year> (accessed June 1, 2023).
- Pergande AE, Belshaw Z, Volk HA, Packer RMA. “We have a ticking time bomb”: a qualitative exploration of the impact of canine epilepsy on dog owners living in England. *BMC Vet Res.* (2020) 16:443. doi: 10.1186/s12917-020-02669-w
- Nettifee JA, Munana KR, Griffith EH. Evaluation of the impacts of epilepsy in dogs on their caregivers. *J Am Anim Hosp Assoc.* (2017) 53:143–9. doi: 10.5326/JAAHA-MS-6537
- Introducing ChatGPT. Available online at: <https://openai.com/blog/chatgpt> (accessed June 14, 2023).
- Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst.* (2023) 47:33. doi: 10.1007/s10916-023-01925-4
- Baumgartner C. The opportunities and pitfalls of ChatGPT in clinical and translational medicine. *Clin Translat Med.* (2023) 4:1206. doi: 10.1002/ctm2.1206
- Chrupala G. Putting natural in natural language processing. *arXiv preprint arXiv:2305.04572*. (2023). doi: 10.18653/v1/2023.findings-acl.495
- Liévin V, Hother CE, Winther O. Can large language models reason about medical questions? *arXiv preprint arXiv:2207.08143*. (2023). doi: 10.48550/arXiv.2207.08143
- DiGiorgio AM, Ehrenfeld JM. Artificial intelligence in medicine and ChatGPT: de-tether the physician. *J Med Syst.* (2023) 47:3. doi: 10.1007/s10916-023-01926-3
- Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare.* (2023) 11:887. doi: 10.3390/healthcare11060887
- De Risio L, Bhatti S, Muñana K, Penderis J, Stein V, Tipold A, et al. International veterinary epilepsy task force consensus proposal: diagnostic approach to epilepsy in dogs. *BMC Vet Res.* (2015) 11:148. doi: 10.1186/s12917-015-0462-1
- Cerda-Gonzalez S, Packer RA, Garosi L, Lowrie M, Mandigers PJJ, O'Brien DP, et al. International veterinary canine dyskinesia task force ECVN consensus statement: terminology and classification. *J Vet Intern Med.* (2021) 35:1218–30. doi: 10.1111/jvim.16108
- Markovski Y. *How Your Data is Used to Improve Model Performance.* OpenAI Help Center. (2023). Available online at: <https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance> (accessed June 14, 2023).
- Sieh J. *Best Practices for Prompt Engineering with OpenAI API.* OpenAI Help Center. (2023). Available online at: <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api> (accessed June 14, 2023).
- Whittaker DE, Volk HA, De Decker S, Fenn J. Clinical characterisation of a novel paroxysmal dyskinesia in Welsh terrier dogs. *Vet J.* (2022) 281:105801. doi: 10.1016/j.tvjl.2022.105801
- Limitations-OpenAI. (2023). Available online at: <https://openai.com/blog/chatgpt> (accessed June 15, 2023).
- Moisset X, Ciampi de Andrade D. Neuro-ChatGPT? Potential threats and certain opportunities. *Rev Neurol.* (2023) 179:517–9. doi: 10.1016/j.neurol.2023.02.066
- Venigalla A. *BioMedLM: A Domain-Specific Large Language Model for Biomedical Text.* (2022). Available online at: <https://www.mosaicml.com/blog/introducing-pubmed-gpt> (accessed June 14, 2023).
- Jin Q, Yang Y, Chen Q, Lu Z. GeneGPT. Augmenting large language models with domain tools for improved access to biomedical information. *arXiv preprint arXiv:2304.09667*. (2023). doi: 10.48550/arXiv.2304.09667
- Privacy policy-OpenAI. (2023). Available online at: <https://openai.com/policies/privacy-policy> (accessed June 15, 2023).
- Sanmarchi F, Bucci A, Golinelli D. A step-by-step Researcher's Guide to the use of an AI-based transformer in epidemiology: an exploratory analysis of ChatGPT using the STROBE checklist for observational studies. *medRxiv.* (2023). doi: 10.1101/2023.02.06.23285514
- Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell.* (2023) 6:1169595. doi: 10.3389/frai.2023.1169595
- Douglas-Moore JL, Lewis R, Patrick JRJ. The importance of clinical documentation. *Bull R Coll Surg Engl.* (2014) 96:18–20. doi: 10.1308/rcsbull.2014.96.1.18
- Thornell C. *Research Shows Gender, Specialty, Geography Among Top Factors Contributing to “Pajama Time” Work for Clinicians.* Athenahealth Knowledge Hub. (2022). Available online at: <https://www.athenahealth.com/knowledge-hub/clinical-trends/research-shows-gender-specialty-geography-among-top-factors-contributing-to-pajama-time-work-for-clinicians> (accessed June 14, 2023)
- Kennedy S. UC Davis Health Launches AI Tool to Curb Documentation Burden. *HealthITAnalytics.* (2023). Available online at: <https://health.ucdavis.edu/news/headlines/uc-davis-health-launches-new-invasive-e-coli-disease-vaccine-trial/> (accessed June 14, 2023)
- Kennedy S. KS Health System Unveils Generative AI Partnership. *HealthITAnalytics.* (2023). Available online at: <https://healthitanalytics.com/news/ks-health-system-unveils-generative-ai-partnership> (accessed June 14, 2023)

42. Kahn D, Stewart E, Duncan M, Lee E, Simon W, Lee C, et al. Prescription for note bloat: an effective progress note template. *J Hosp Med.* (2018) 13:378–82. doi: 10.12788/jhm.2898
43. Nori H, King N, McKinney S, et al. Capabilities of GPT-4 on Medical Challenge Problems. arXiv preprint arXiv:2303.13375. (2023). Available online at: <https://arxiv.org/abs/2303.13375> (accessed March 3, 2023).
44. You Y, Gui X. Self-Diagnosis through AI-enabled Chatbot-based symptom checkers: user experiences and design considerations. *AMIA Annu Symp Proc.* (2021) 2020:1354–63. doi: 10.48550/arXiv.2305.04572
45. Deik A. Potential benefits and perils of incorporating ChatGPT to the movement disorders clinic. *J Movement Disord.* (2023) 16:158–62. doi: 10.14802/jmd.23072



OPEN ACCESS

EDITED BY

T. Bas Rodenburg,
Utrecht University, Netherlands

REVIEWED BY

Mariska van der Voort,
Wageningen University and Research,
Netherlands
Oleksiy Guzha,
Swedish University of Agricultural Sciences,
Sweden

*CORRESPONDENCE

Dinu Gavojdian

✉ gavojdian_dinu@animalsci-tm.ro

RECEIVED 18 September 2023

ACCEPTED 09 November 2023

PUBLISHED 30 November 2023

CITATION

Magana J, Gavojdian D, Menahem Y,
Lazebnik T, Zamansky A and
Adams-Progar A (2023) Machine learning
approaches to predict and detect early-onset
of digital dermatitis in dairy cows using sensor
data.

Front. Vet. Sci. 10:1295430.

doi: 10.3389/fvets.2023.1295430

COPYRIGHT

© 2023 Magana, Gavojdian, Menahem,
Lazebnik, Zamansky and Adams-Progar. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Machine learning approaches to predict and detect early-onset of digital dermatitis in dairy cows using sensor data

Jennifer Magana¹, Dinu Gavojdian^{2*}, Yakir Menahem³,
Teddy Lazebnik^{4,5}, Anna Zamansky⁶ and Amber Adams-Progar¹

¹Department of Animal Sciences, Washington State University, Pullman, WA, United States, ²Cattle Production Systems Laboratory, Research and Development Institute for Bovine, Balotesti, Romania,

³Department of Computer Science, Holon Institute of Technology, Holon, Israel, ⁴Department of Mathematics, Ariel University, Ariel, Israel, ⁵Department of Cancer Biology, University College London, London, United Kingdom, ⁶Tech4Animals Laboratory, Information Systems Department, University of Haifa, Haifa, Israel

The present study aimed to employ machine learning algorithms based on sensor behavior data for (1) early-onset detection of digital dermatitis (DD) and (2) DD prediction in dairy cows. Our machine learning model, which was based on the Tree-Based Pipeline Optimization Tool (TPOT) automatic machine learning method, for DD detection on day 0 of the appearance of the clinical signs has reached an accuracy of 79% on the test set, while the model for the prediction of DD 2 days prior to the appearance of the first clinical signs, which was a combination of K-means and TPOT, has reached an accuracy of 64%. The proposed machine learning models have the potential to help achieve a real-time automated tool for monitoring and diagnosing DD in lactating dairy cows based on sensor data in conventional dairy barn environments. Our results suggest that alterations in behavioral patterns can be used as inputs in an early warning system for herd management in order to detect variances in the health and wellbeing of individual cows.

KEYWORDS

animal behavior, dairy cattle, digital dermatitis, sensor data, machine learning

1 Introduction

Digital dermatitis (DD) is one of the most prevalent infectious diseases in dairy cows worldwide, being responsible for substantial economic losses due to impaired production and reproduction, higher risks of culling, and treatment costs while having detrimental effects on animal welfare (1–3).

DD in cattle is regarded as a complex disease influenced by multiple microbes. While its exact pathogenesis is still not fully understood (4, 5), foot lesions are often associated with various phylotypes of *Treponemes*. Consequently, the *Treponema* genus is considered the primary causal agent (6). Other bacteria species such as *Porphyromonas*, *Fusobacterium*, and *Dichelobacter* are believed to act synergistically (7–9). DD in cattle manifests in the form of ulcerative or growth-like skin lesions, primarily located digitally and on the coronary band of the hoof. The hind legs are affected in over 90% of cases (10), and this condition is typically associated with lameness. It can also coexist with other issues such as foot rot, sole ulcers, sole hemorrhages, and white line disease.

The highly contagious nature and reduced treatment responses of DD (11–13) were shown to result in increased prevalence of up to 91% at the herd level while affecting up to 41% of the animals (14–16). Diagnosis for DD is based on the visual inspection of the feet using the Mortellaro-stage (M-stage) scoring system described and modified by Berry et al. (17), where lesion type and size are differentiated.

Although the etiopathogenesis of bovine DD is not well understood (18), in recent years, several studies have been conducted to identify risk factors associated with the occurrence of DD in dairy cattle. At the individual animal level, the main risk factors for developing DD were found to be breed, milk yield, parity, lactation stage, presence of metabolic diseases, interindividual differences in the immune response, and animal behavior (3, 19, 20), while the main risk factors at the farm level are represented by the housing system, flooring type, plan of nutrition, general farm biosecurity, and preventive practices used to mitigate bovine digital dermatitis (21, 22).

With the recent advent of precision livestock farming (PLF), an increasing body of research addresses machine learning approaches for the early detection of cattle diseases (23–27). Furthermore, several studies have already successfully applied computer vision for detecting and classifying DD in cattle (28–30). As a result, sensor-based behavior monitoring technologies are promising, more affordable, and operationally simpler alternatives for disease monitoring and diagnostics. Currently, a wide range of commercially validated systems are available (31), which monitor behaviors such as feeding, ruminating, activity, and lying. Some of these behavioral patterns have been directly linked to DD in cattle, with ill animals spending more time lying down than their healthy counterparts and devoting less time to feeding and rumination (32, 33). However, to the best of our knowledge, the use of machine learning for the detection of DD from behavioral sensor data has not yet been explored in cattle.

The present study aimed to employ machine learning algorithms based on sensor behavior data for (1) *early-onset detection of DD* and (2) *DD prediction*, with the ultimate goal of setting up early warning tools for DD prediction. These warning tools would then enable farmers and veterinarians to better monitor and manage DD in commercial settings, resulting in a decrease in DD prevalence and severity while improving animal welfare.

2 Materials and methods

2.1 Animal management and data collection

All procedures involving animals used in the current study were approved by the Washington State University Institutional Animal Care and Use Committee (IACUC), with the approval code ASAF#6770. The data collection process occurred over 60 consecutive days at the Washington State University Knott Dairy Center (KDC) in Pullman, Washington, USA (GPS: 46.6937°N, 117.2423°W).

The experimental cattle facility at the KDC experimental farm houses 180 Holstein pedigreed purebred cows, with lactating animals being housed in a free-stall barn with individual cubicles, using composted manure as bedding. Cows are milked twice per day using a 6×6 ‘herring-bone’ milking parlor, have *ad libitum* access to two water troughs, and are fed a total mixed ration twice per day. The KDC farm practices zero-grazing for lactating cows (indoor housing year-round),

with movement alleys and the outside paddock having concrete flooring. During the dry period, the cows are housed in deep-bedded packs with access to grazing areas. Each cow at the KDC experimental farm was fitted with a CowManager® (CowManager B.V., Harmelen, Netherlands) ear tag that continuously records animal behavior, rumination, and ear temperature 24h per day. The measurements of interest in this study were activity (non-active, active, and highly active), eating time, rumination time, and ear temperature. All behavioral data were calculated as the proportion of time each cow spent exhibiting each behavioral pattern and computed in hours devoted to that behavior per 24h.

The CowManager® sensor is a molded microchip that has been adapted into a cattle ear identification tag (Supertag; Dalton ID Ltd., Oxfordshire, UK). A three-dimensional accelerometer within the sensor continuously registers the activities of the cow, with the raw data being sent through a wireless connection via routers to a central computer. The raw sensor data are continuously collected, and each minute is classified into one of the four measurement categories: “ruminating,” “eating,” “resting,” and/or “active,” with a proprietary model of the sensor. Data obtained are subsequently expressed as minutes of behavior per hour as well as hours per day and were retrieved through a web-based application. The ear temperature was presented as the average/day and expressed in °C. The sensor used has been previously validated to effectively monitor the behavior of free-stall housed dairy cattle (34).

Cattle were enrolled in the study if they met two criteria: (1) no lesions for at least 7 days prior to the first observation of an active lesion and (2) had at least 2 consecutive days of DD lesion observed. During the study, 21 cows that were between the 1st and 5th lactation periods developed DD. Each cow that developed a DD episode was then matched with a healthy counterpart that had the same parity, reproduction status (open/pregnant), and lactation period (early/mid/late). Lactation periods were classified as early (< 100 days in milk [DIM]), mid (101–199 DIM) or late (> 199 DIM). Therefore, the final dataset included 21 cows with DD and 21 healthy cows. As a prevention method for DD, an acidified copper-, sulfate-, and zinc footbath solution was placed at the exit of the milking parlor. The footbath solution was replaced twice a week, following the recommendations of a hoof specialist. The observer for this study was trained by a hoof specialist to evaluate digital dermatitis (DD) lesions. All hoofs were visually assessed during the first milking in the morning inside the milking parlor, looking exclusively at the hind feet. To date, there has been no golden standard for making observations to determine if a cow has DD. In most on-farm cases, cows are observed for lameness, and those with signs of lameness are further evaluated for hoof disorders. The DD lesion scoring system implemented in this study follows the widely used M-stage scoring system (35). When observed, lesions were categorized as active (red and painful with hair on lesions) or digressing (no hair or little hair, no pain, and scabbing on lesions). Lesion size was categorized as either small (<0.635 cm), medium (0.635–3.81 cm), or large (>3.81 cm) based on the lesion diameter. The same observer recorded the DD status and lesion size daily during the trial to avoid interobserver biases.

2.2 Machine learning models

The data used in this study are of time series type, which is a sequence of data points measured at successive points in time spaced

at uniform time intervals. All measurements were continuous and aggregated to time frames of a single day. We analyzed 2,520 entries in total (i.e., the daily data about a cow) as part of our dataset. Namely, we analyzed the data, which are the product of 7 days over 6 features, collected for 60 cows. Due to the challenging nature of the problem, we took a two-step approach to tasks of increasing difficulty. The first task was detection, namely whether the cow has DD or not on day 0, looking at data from all days prior to day 0 (-7 days). The second, and a more challenging task, was the prediction/forecasting of DD episodes, especially classifying whether the cow will have DD or not on day 0 based on data x days before day 0 (where the optimal value of x needs to be determined).

2.3 Detection machine learning model

The first task we address is providing a machine learning classifier of whether a specific cow has DD or not on day 0. In this section, we describe the training process of the machine learning model. We first divided the dataset into training and testing cohorts such that the training cohort contained 80% of the dataset, while the remaining 20% belonged to the test cohort. Importantly, we ensured the distribution of the target feature in both the training and test cohorts using the Monte Carlo method, taking the best random split out of $n = 100$ attempts. The training cohort was then used to train the model, and the testing cohort was used to evaluate its performance. Importantly, samples from the same individual were either included in the training or testing cohort in order to avoid potential data leakage between the two. Moreover, to make sure the results were robust, we further divided the training cohort using the k -fold cross-validation method (36) with $k = 5$. Using the training cohort, we then used the Tree-Based Pipeline Optimization Tool (TPOT) automatic machine learning library method (37). Formally, given a dataset $D \in \mathbb{R}^{c \times n}$ with $c \in \mathbb{N}$ features and $n \in \mathbb{N}$ samples, we utilized TPOT, which uses a GA-based approach, to generate and test ML pipelines based on the popular scikit-learn library (38). Formally, we run the TPOT classifier search method to obtain an ML pipeline that aims to optimize the classifier's mean accuracy over the k -folds (39). Once the pipeline was obtained, we further aimed to improve the model's performance over the training cohort using the grid-search hyperparameters method (40) such that the hyperparameter value ranges were chosen manually (41). Finally, the obtained model was evaluated using the testing cohort. This model development process was similar in nature to other recent studies in sensory data of dairy cattle (42); however, rather than manually testing multiple ML models, we used the automatic machine learning approach, which performed this task more efficiently in terms of time.

2.4 Prediction machine learning model

Two important concepts in the context of time series forecasting are 'lag' and 'window'. A 'lag' in time series prediction is a way of referencing past data points, e.g., a lag of 1 would mean the previous data point, a lag of 2 would mean the data point two periods back, and so forth. A (rolling) 'window' refers to a fixed-size subset of a time series dataset. The aim was to take a portion of the data of a particular length (window size) and move those data across the time series.

Having a window allowed us to create aggregated features such as moving averages, sums, and standard deviations. The question, then, becomes, what lag and what window size would yield better performance of the model for prediction? Obviously, with lag 0, we are back to the prediction problem. Going to lags 1, 2, and 3 will decrease our accuracy, but it will also mean that we are able to make the prediction sooner. We thus had a time series task with some lag $l \in \mathbb{N}$ and a window size $w \in \mathbb{N}$. In this representation, the disease occurrence prediction takes a binary classification form. However, naturally, the number of negative samples is much larger than the number of possible samples, as these occur once for each cow, by definition. Hence, to balance the data, we undersampled the negatively labeled samples using the K-means method (43) such that the number of clusters equals the number of positive samples. Building on these grounds, we repeat the same computational process as the one used to obtain the disease detection classifier. In addition, to investigate the influence of the lag and window size parameters, the disease occurrence predictor was obtained for all possible combinations of these parameters. To control the balancing method, we also used the class weight fixing method, where the number of samples is kept the same but the weight of each label is different to count for the differences in the labels' group sizes. Both models were implemented using the Python programming language (version 3.8.1) (44) and set at a value of p of ≤ 0.05 to be statistically significant.

3 Results

3.1 Digital dermatitis detection on day 0

For a preliminary exploration, we computed the Pearson correlation matrix (45) between the sensor's data and the target variable (presence/absence of DD in the cow). Figure 1 presents the matrix. As can be noticed, most of the values were less than 0.3, which strongly indicates that the inputted space was mostly linearly independent (46). Hence, a non-linear-based model should be investigated for the proposed challenge.

To this end, we investigated the pairwise relationship of the inputted features and their relationship with the target feature, as presented in Figure 2, such that the red (square) markers indicate DD sick cows while the green (circle) markers indicate healthy cows. The lines indicate the kernel density estimate of each pairwise distribution. In more pairwise plots as well as the features' histograms, it can be visibly observed that there is no clear separation between the target feature sets.

Based on the above, for disease detection, we obtained an ensemble model that combines a random forest model (47, 48) and a k -nearest neighbors algorithm (49), which received a second-order polynomial extension of the inputted features after min-max normalization (50). For this model, we obtained an accuracy of 81.2% for the training set with 5-fold cross-validation. More importantly, for the testing set, we obtained an accuracy of 79.2%. These results indicate that the proposed model was well-fitting, due to the relatively small difference between the mean performance over the training and testing sets. Nonetheless, the standard deviation of 4.6% indicated that the model might be somewhat non-data stable (51). Considering the standard deviation with a probability of 95% CIs, we estimated that the proposed model would have an accuracy of at least 72%.

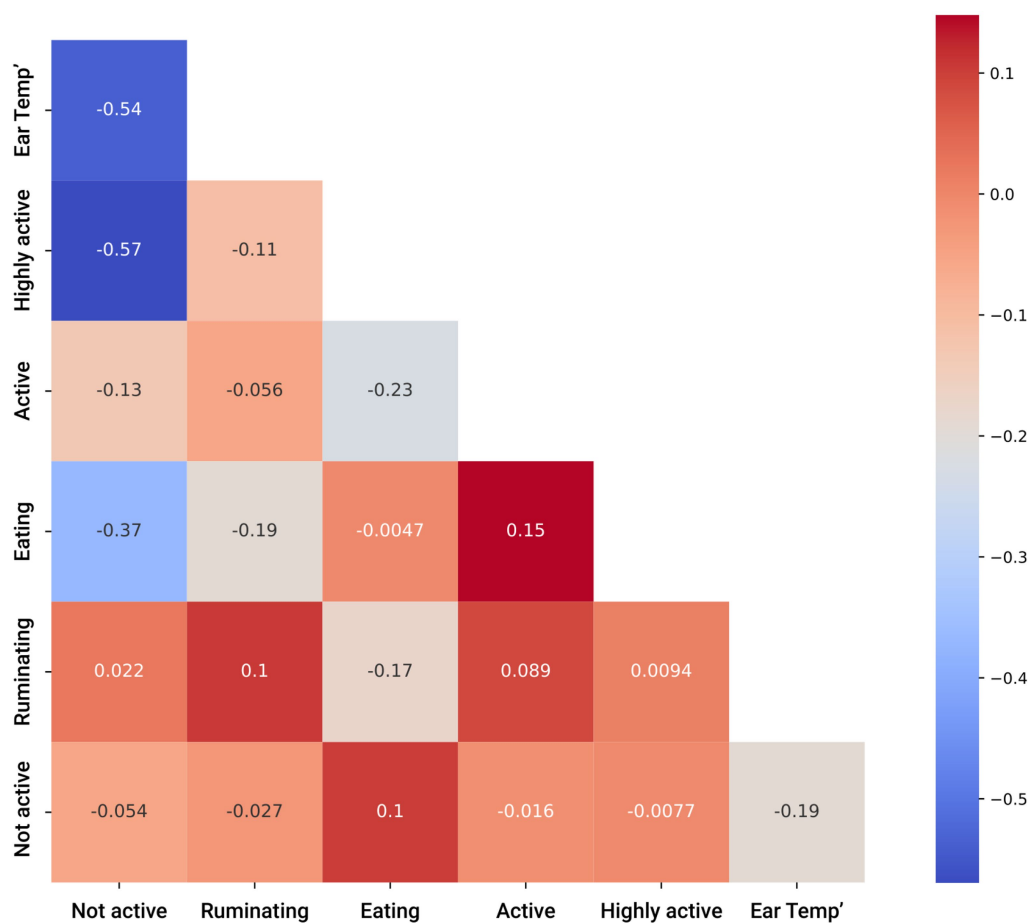


FIGURE 1
Pearson's coloration matrix between the input features for the disease detection model.

To learn which features contribute the most to the model's classification capabilities, we computed the model's features by removing one feature at a time, evaluating how this influences the model's accuracy, and normalizing these results once all values were obtained. We repeated this process on the entire dataset with a 5-fold cross-validation. Figure 3 shows the results of this analysis, where it can be observed that 'activity' is the most important feature, followed by the 'not-active' feature.

In a complementary manner, Figure 4 presents the feature that is important using the SHapley Additive exPlanations (SHAP) value, which connects optimal credit allocation with local explanations using the classic Shapley values from game theory. Specifically, it shows how the value of each feature corresponds to the model's prediction. One can notice that higher values of ear temperature and non-activity are associated with a higher probability of DD occurrence.

3.2 Digital dermatitis prediction prior to day 0

To find the optimal parameters for window and lag, Figure 5 shows a sensitivity analysis of the model's accuracy, which was computed for the test set as a function of the lag and window size of the prediction. We can see that, for example, 2 days prior to the appearance of the first clinical signs, we have an accuracy of 64% by

looking at a window from 3 days ago. One day prior to day 0, the accuracy increases to 71% (with a window of 3 days). It is important to point out that a 50% accuracy of a binary prediction, such as the one presented in this case, indicates a random choice, thus taking into account that a larger window or looking more days ahead yields low performance, indicating that the model failed to learn any significant pattern and, as a result, more or less guessing the result with some minor (false) bias obtained from slightly overfitting of the training set. In addition, the results are comparable as we downsampled the train and test set sizes to be identical for all cases such that the train and test sets include 98 and 28 samples, respectively.

4 Discussion

In this study, we present a machine learning model for DD detection on the first day for the appearance of clinical signs with an accuracy of 79% and a model for the prediction of DD 2 days prior to the appearance of the first clinical signs with an accuracy of 64%. The accuracy attained for the detection of DD was higher in our study when compared to reports by Cernek et al. (28), which applied computer vision approaches for detecting DD in cattle.

In the current study, activity was found to be the most important sensor feature for DD detection. Similarly, Tsai et al. (52) reported that

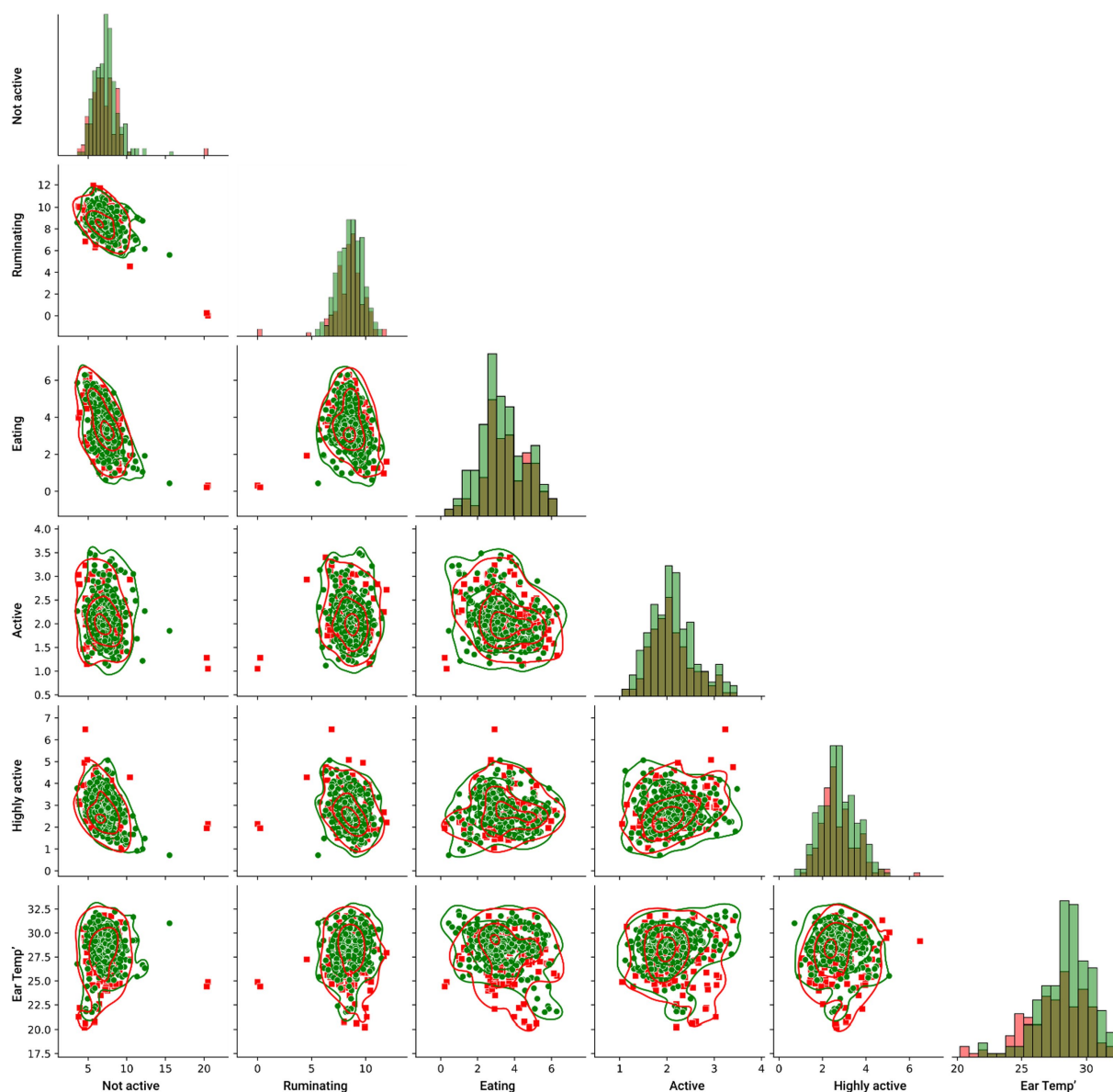


FIGURE 2

Pair plot between the features of the model, divided by the target features such that the red (square) markers indicate DD sick cows while the green (circle) markers indicate healthy cows. The lines indicate the kernel density estimate of each pairwise distribution.

activity and, most importantly, changes in time devoted to walking represent valid indicators for disease detection in cattle. The same authors outline that the current use of PLF needs an improvement in detection accuracy at the farm level. Our results are also in line with the findings by Soriani et al. (53), who reported changes in lying time for cows affected by lameness. Contrary to our findings, these authors found a significant decrease in the time devoted to ruminating during the first days of subclinical diseases or health disorders.

Barker et al. (54) validated the combined use of an animal neck-mounted sensor with a location device to classify cattle behavior, with feeding behavior patterns being used for lameness detection. Interestingly, feeding behavior has not played a significant role in DD detection or prediction in our case. This can be explained by the differences in sensor devices, as well as differences in the machine learning models used, and deserves further exploration.

Regarding temperature variations, Harris-Bridge et al. (55) found a significant temperature rise at the foot level in dairy cows with DD using infrared thermography, with similar results reported by Pirkkalainen et al. (56) for rectal temperature in DD vs. healthy cows, with authors attributing the temperature rises to the effects of inflammation at the foot level. However, such temperature fluctuations have not been observed in our study, most likely due to the sensor being placed in the animal's ear, and thus, the assumed temperature rise in cows with DD might have occurred only in the plantar region.

The results of our study highlight the potential applications of behavioral sensor data extracted from commercially available sensors for the prediction of digital dermatitis. Current findings are in accordance with those reported by Benaissa et al. (57), which found that cattle sensor behavior data are strongly linked to the animals' health and welfare. Furthermore, Hosseinioorbin et al. (58) found that both

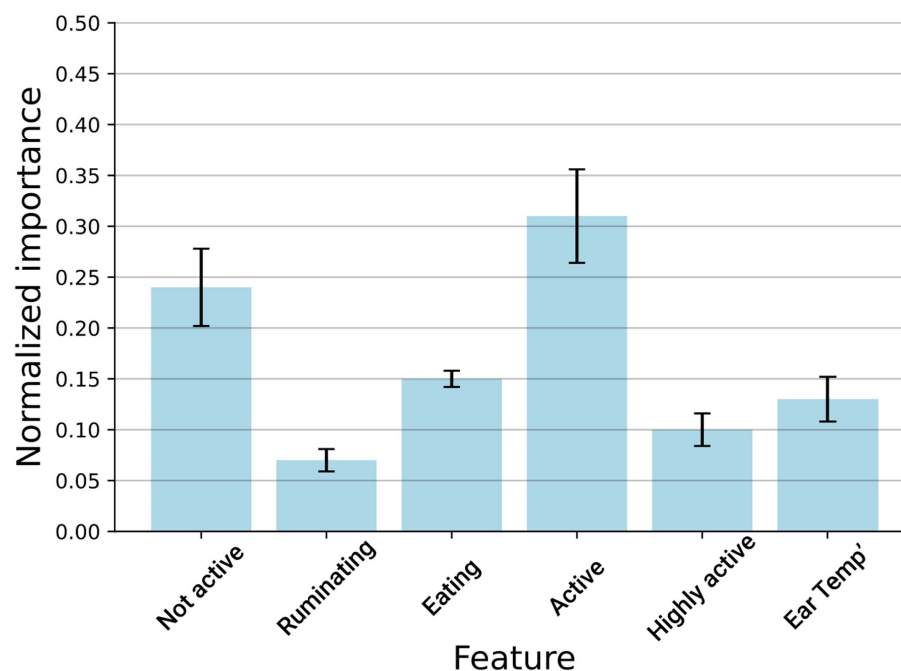


FIGURE 3

Disease detection model's feature importance measures the relative information gained from each feature. The results are shown as the mean \pm standard deviation of 5-fold cross-validation performed on the entire dataset.

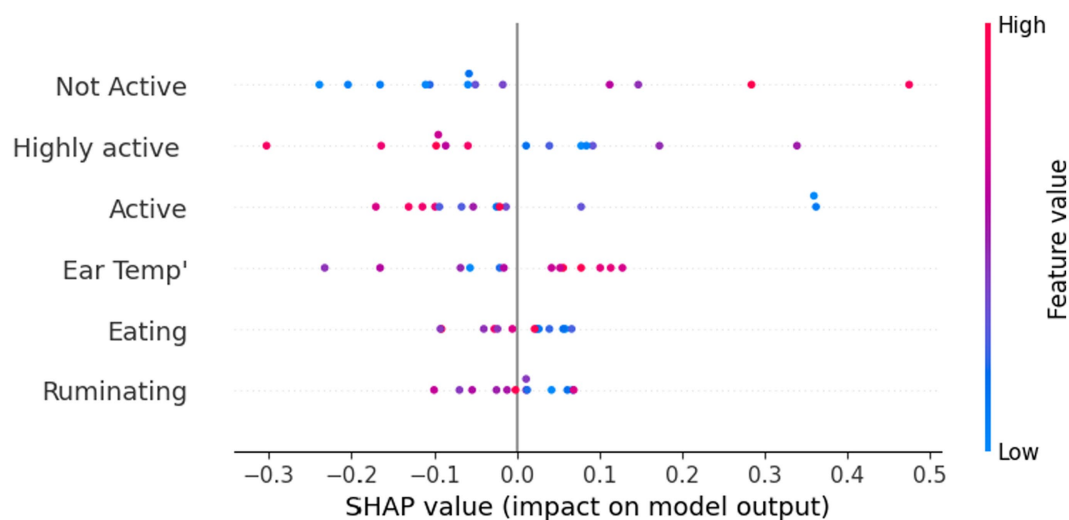


FIGURE 4

Disease detection model's feature importance is measured by the SHapley Additive exPlanations (SHAP) value of each feature.

lameness and infectious diseases can be detected via the use of cattle behavior. However, further studies are needed to expand this exploration, focusing on studying the forecasting parameters of lag and window revealed in this study. The use of other, more complex sensor systems that provide more fine-grained behavioral data can potentially increase the performance of the machine learning models presented here.

The main drawback for the lack of adoption of sensor data combined with other PLF systems in commercial settings is the additional cost to be incurred for the farms. It is important to note that farmers are more likely to adopt a PLF system if the system provides

useful information they can use to make informed decisions. The activity monitoring device used in the current study is currently used by farmers to identify potentially unhealthy animals; however, the system is not able to differentiate between different types of sickness. One of our intentions for this study is to demonstrate how activity monitoring device data can be used to identify cows with DD. If the system could incorporate this type of alert for cows with DD, the system would have added value and may be more widely adopted. Overcoming bottlenecks such as user adoption could result in better monitoring of the herd by improving estrous and early disease

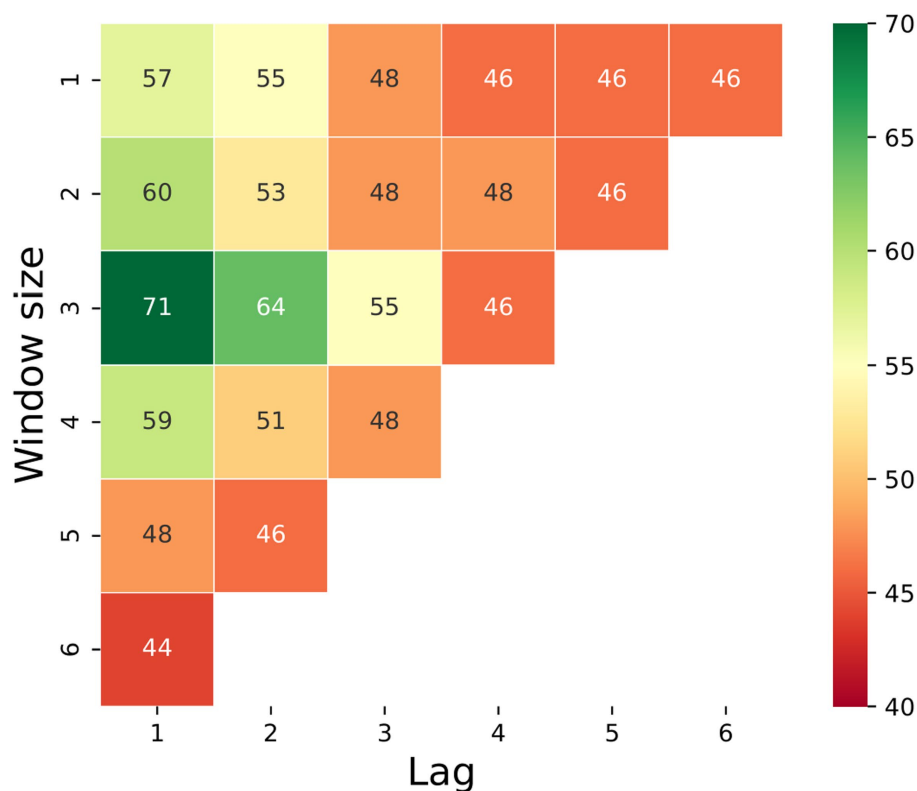


FIGURE 5

Heatmap of the models' accuracy on the test set (presented in percentage) as a function of their lag and window size. Notably, a 50% accuracy of a binary prediction indicates a random choice; thus, all results below shows that the model failed to learn any significant pattern.

detection, which would then translate into improved overall farm efficiency. For instance, in a study that coupled accelerometer and GPS location data, Cabezas et al. (59) found a high accuracy of 93% for classifying four dairy cattle behavioral patterns. The grouping of these two sensors was also used to track the social interactions between cows and social behavior, which was significantly linked to both health and animal welfare (60). These aspects are of high importance, given that Proudfoot et al. (61) reported their observation of sick cows isolating themselves and avoiding both allogrooming and agonistic interactions with herd-mates, mainly throughout the use of less frequented cubicles located at the far ends of the barn and away from resources such as feeding alleys and water troughs.

This study is not without limitations. First, the number of developed DD cases during the trial-period and that qualified for enrolment in the study-herd could be increased, hopefully leading to higher performance for both detection and prediction models. Therefore, for our future studies, we plan to include more farms with different barn designs and test the machine learning models in more diverse farming settings. Furthermore, the currently commercially available behavior sensors are focused on monitoring a rather limited number of behavioral patterns, providing data mainly on feeding, ruminating, and activity time budgets. The detection of behaviors that are less frequent or are being expressed during shorter periods of time, such as social interactions, the resting position of the animal, or drinking bouts and rates, and even changes in the behavioral circadian rhythm could be altered during a disease episode; however, to date, the validation of sensors to monitor such behaviors remains a challenge. To overcome these shortcomings, several authors

recommend the integrated use of additional PLF tools, such as image analysis-based systems, pressure sensors, radio-frequency identification, and ultra-wideband technology (62–64); thus, progress on this front is expected.

5 Conclusion

In conclusion, a machine learning model that is capable of predicting and detecting bovine digital dermatitis in cows housed under free-stall conditions based on behavior sensor data has been proposed and tested in this exploratory study. The model for DD detection on day 0 of the appearance of the clinical signs has reached an accuracy of 79%, while the model for the prediction of DD 2 days prior to the appearance of the first clinical signs has reached an accuracy of 64%. The proposed machine learning models might help to achieve a real-time automated tool for monitoring and diagnosing DD in lactating dairy cows based on behavior sensor data in conventional dairy barn environments. Our results suggest that alterations in behavioral patterns at individual levels can be used as inputs in an early warning system for herd management in order to detect variances in the health and wellbeing of individual cows.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

All procedures used in the current study were approved by the Washington State University Institutional Animal Care and Use Committee (IACUC), approval code ASAF#6770. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

JM: Formal analysis, Investigation, Writing – original draft. DG: Data curation, Formal analysis, Funding acquisition, Investigation, Resources, Writing – review & editing. YM: Data curation, Formal analysis, Methodology, Validation, Investigation, Writing – original draft. TL: Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft. AZ: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – review & editing. AA-P: Conceptualization, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was

supported by a grant from the Ministry of Research, Innovation, and Digitization, CNCS – UEFISCDI, project number PN-III-P1-1.1-TE-2021-0027, within PNCDI III.

Acknowledgments

During the trial's implementation, DG received support to join Washington State University through a Romanian – U.S. Fulbright Visiting Scholar grant. We thank the Washington State University Knott Dairy Center staff for providing assistance during data collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Palmer MA, O'Connell NE. Digital dermatitis in dairy cows: a review of risk factors and potential sources of between-animal variation in susceptibility. *Animals*. (2015) 5:512–35. doi: 10.3390/ani5030369
- Plummer PJ, Krull A. Clinical perspectives of digital dermatitis in dairy and beef cattle. *Vet Clin North Am Food Anim Pract*. (2017) 33:165–81. doi: 10.1016/j.cvfa.2017.02.002
- Weber J, Becker J, Syring C, Welham Ruiters M, Locher I, Bayer M, et al. Farm-level risk factors for digital dermatitis in dairy cows in mountainous regions. *J Dairy Sci*. (2023) 106:1341–50. doi: 10.3168/jds.2022-22243
- Wilson-Welder JH, Alt DP, Nally JE. The etiology of digital dermatitis in ruminants: recent perspectives. *Vet Med*. (2015) 6:155–64. doi: 10.2147/VMRR.S62072
- Evans NJ, Murray RD, Carter SD. Bovine digital dermatitis: current concepts from laboratory to farm. *Vet J*. (2016) 211:3–13. doi: 10.1016/j.tvjl.2015.10.028
- Alsaod M, Locher I, Jores J, Grimm P, Brodard I, Steiner A, et al. Detection of specific *Treponema* species and *Dichelobacter nodosus* from digital dermatitis (Mortellaro's disease) lesions in Swiss cattle. *Schweiz Arch Tierheilkd*. (2019) 161:207–15. doi: 10.17236/sat00201
- Rasmussen M, Capion N, Klitgaard K. Bovine digital dermatitis: possible pathogenic consortium consisting of *Dichelobacter nodosus* and multiple *Treponema* species. *Vet Microbiol*. (2012) 160:151–61. doi: 10.1016/j.vetmic.2012.05.018
- Knappe-Poindecker M, Gilhuus M, Jensen TK, Klitgaard K, Larssen RB, Fjeldaas T. Interdigital dermatitis, heel horn erosion, and digital dermatitis in 14 Norwegian dairy herds. *J Dairy Sci*. (2013) 96:7617–29. doi: 10.3168/jds.2013-6717
- Wilson-Welder JH, Alt DP, Nally JE. Digital dermatitis in cattle: current bacterial and immunological findings. *Animals*. (2015) 5:1114–35. doi: 10.3390/ani5040400
- Solano L, Barkema HW, Mason S, Pajor EA, LeBlanc SJ, Orsel K. Prevalence and distribution of foot lesions in dairy cattle in Alberta. *Canada J Dairy Sci*. (2016) 99:6828–41. doi: 10.3168/jds.2016-10941
- Orsel K, Plummer P, Shearer J, De Buck J, Carter SD, Guatteo R, et al. Missing pieces of the puzzle to effectively control digital dermatitis. *Transbound Emerg Dis*. (2018) 65:186–98. doi: 10.1111/tbed.12729
- Corlevic AT, Beggs DS. Host factors impacting the development and transmission of bovine digital dermatitis. *Ruminants*. (2022) 2:90–100. doi: 10.3390/ruminants2010005
- Capion N, Larsson EK, Nielsen OL. A clinical and histopathological comparison of the effectiveness of salicylic acid to a compound of inorganic acids for the treatment of digital dermatitis in cattle. *J Dairy Sci*. (2018) 101:1325–33. doi: 10.3168/jds.2017-13622
- Relun A, Lehebel A, Chesnin A, Guatteo R, Bareille N. Association between digital dermatitis lesions and test-day milk yield of Holstein cows from 41 French dairy farms. *J Dairy Sci*. (2013) 96:2190–200. doi: 10.3168/jds.2012-5934
- Jury A, Syring C, Becker J, Locher I, Strauss G, Ruiters M, et al. Prevalence of claw disorders in Swiss cattle farms. *Schweiz Arch Tierheilkd*. (2021) 164:779–90. doi: 10.17236/sat00327
- Kofler J, Suntinger M, Mayerhofer M, Linke K, Maurer L, Hund A, et al. Benchmarking based on regularly recorded claw health data of Austrian dairy cattle for implementation in the cattle data network. *Animals*. (2022) 12:808. doi: 10.3390/ani12070808
- Berry SL, Read DH, Famula TR, Mongini A, Dopfer D. Long-term observations on the dynamics of bovine digital dermatitis lesions on a California dairy after topical treatment with lincomycin HCl. *Vet J*. (2012) 193:654–8. doi: 10.1016/j.tvjl.2012.06.048
- Vanhoudt A, Orsel K, Nielen M, van Werven T. An observational study on the management of digital dermatitis through a repeated risk assessment on 19 Dutch dairy herds. *J Dairy Sci*. (2021) 104:947–56. doi: 10.3168/jds.2020-18730
- Holzhauser M, Hardenberg C, Bartels CJ, Frankena K. Herd- and cow-level prevalence of digital dermatitis in the Netherlands and associated risk factors. *J Dairy Sci*. (2006) 89:580–8. doi: 10.3168/jds.S0022-0302(06)72121-X
- Schopke K, Gomez A, Dunbar KA, Swale HH, Dopfer D. Investigating the genetic background of bovine digital dermatitis using improved definitions of clinical status. *J Dairy Sci*. (2015) 98:8164–74. doi: 10.3168/jds.2015-9485
- Oliveira VHS, Sorensen JT, Thomsen PT. Associations between biosecurity practices and bovine digital dermatitis in Danish dairy herds. *J Dairy Sci*. (2017) 100:8398–408. doi: 10.3168/jds.2017-12815
- Yang DA, Gates MC, Muller KR, Laven R. Bayesian analysis of herd-level risk factors for bovine digital dermatitis in New Zealand dairy herds. *BMC Vet Res*. (2019) 15:125. doi: 10.1186/s12917-019-1871-3
- Van Hertem T, Maltz E, Antler A, Romanini CEB, Viazzi S, Bahr C, et al. Lameness detection based on multivariate continuous sensing of milk yield, rumination, and neck activity. *J Dairy Sci*. (2013) 96:4286–98. doi: 10.3168/jds.2012-6188

24. Arcidiacono C, Porto SMC, Mancino M, Cascone G. Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data. *Comput. Electron. Agric.* (2017) 134:124–34. doi: 10.1016/j.compag.2017.01.021
25. Taneja M, Byabazaire J, Jalodia N, Davy A, Olariu C, Malone P. Machine learning based fog computing assisted data-driven approach for early lameness detection in dairy cattle. *Comput. Electron. Agric.* (2020) 171:105286. doi: 10.1016/j.compag.2020.105286
26. Neethirajan S. The role of sensors, big data and machine learning in modern animal farming. *Sens. Bio-Sens. Res.* (2020) 29:100367. doi: 10.1016/j.sbsr.2020.100367
27. Casella E, Cantor MC, Setser MMW, Silvestri S, Costa JHC. A machine learning and optimization framework for the early diagnosis of bovine respiratory disease. *IEEE Access.* (2023) 11:71164–79. doi: 10.1109/ACCESS.2023.3291348
28. Cernek P, Bollig N, Anklam K, Dopfer D. Hot topic: detecting digital dermatitis with computer vision. *J. Dairy Sci.* (2020) 103:9110–5. doi: 10.3168/jds.2019-17478
29. Kang X, Zhang XD, Liu G. Accurate detection of lameness in dairy cattle with computer vision: a new and individualized detection strategy based on the analysis of the supporting phase. *J. Dairy Sci.* (2020) 103:10628–38. doi: 10.3168/jds.2020-18288
30. Vanhoudt A, Jacobs C, Caron M, Barkema HW, Nielen M, van Werven T, et al. Broad-spectrum infrared thermography for detection of M2 digital dermatitis lesions on hind feet of standing dairy cattle. *PLoS One.* (2023) 18:e0280098. doi: 10.1371/journal.pone.0280098
31. Borchers MR, Chang YM, Tsai IC, Wadsworth BA, Bewley JM. A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors. *J. Dairy Sci.* (2016) 99:7458–66. doi: 10.3168/jds.2015-10843
32. Pavlenko A, Bergsten C, Ekesbo I, Kaart T, Aland A, Lidfors L. Influence of digital dermatitis and sole ulcer on dairy cow behaviour and milk production. *Animal.* (2011) 5:1259–69. doi: 10.1017/S1751731111000255
33. Thomas AD, Orsel K, Cortes JA, Pajor EA. Impact of digital dermatitis on feedlot cattle behaviour. *Appl. Anim. Behav. Sci.* (2021) 244:105468. doi: 10.1016/j.applanim.2021.105468
34. Bikker JP, van Laar H, Rump P, Doorenbos J, van Meurs K, Griffioen GM, et al. Technical note: evaluation of an ear-attached movement sensor to record cow feeding behavior and activity. *J. Dairy Sci.* (2014) 97:2974–9. doi: 10.3168/jds.2013-7560
35. Dopfer D, Koopmans A, Meijer FA, Schukken YH, Szakall I, Klee W, et al. Histological and bacteriological evaluation of digital dermatitis in cattle, with special reference to spirochaetes and *Campylobacter faecalis*. *Vet. Rec.* (1997) 140:620–3. doi: 10.1136/vr.140.24.620
36. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI.* (1995) 14:1137–45.
37. Olson RS, Moore JH. TPOT: a tree-based pipeline optimization tool for automating machine learning. In: Hutter F, Kotthoff L, and Vanschoren J. Eds. *Workshop on automatic machine learning PMLR*. (Proceedings of the 2016 Workshop on Automatic Machine Learning, AutoML 2016, co-located with 33rd International Conference on Machine Learning (ICML 2016) New York City, NY, USA) (2016). 66–74. Available at: <https://dblp.org/rec/conf/icml/2016autml.html>
38. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *Machine Learning Res.* (2011) 12:2825–30.
39. Lazebnik T, Fleischer T, Yaniv-Rosenfeld A. Benchmarking biologically-inspired automatic machine learning for economic tasks. *Sustainability.* (2023) 15:11232. doi: 10.3390/su151411232
40. Liu R, Liu E, Yang J, Li M, Wang F. Optimizing the hyper-parameters for SVM by combining evolution strategies with a grid search. In: Huang DS, Li K, Irwin GW and Eds. *Intelligent control and automation: International conference on intelligent computing, ICIC Kunning*. Berlin: Springer (2006). 712–21.
41. Lazebnik T, Somech A, Weinberg AI. SubStrat: a subset-based optimization strategy for faster AutoML. *Proc. VLDB Endowment.* (2022) 16:772–80. doi: 10.14778/3574245.3574261
42. Vidal G, Sharpnack J, Pinedo P, Tsai IC, Lee AR, Martinez-Lopez B. Impact of sensor data pre-processing strategies and selection of machine learning algorithm on the prediction of metritis events in dairy cattle. *Prev. Vet. Med.* (2023) 215:105903. doi: 10.1016/j.prevetmed.2023.105903
43. Likas A, Vlassis N, Verbeek JJ. The global k-means clustering algorithm. *Pattern Recogn.* (2003) 36:451–61. doi: 10.1016/S0031-3203(02)00060-2
44. Srinath KR. Python—the fastest growing programming language. *Int. Res. J. Eng. Technol.* (2017) 4:354–7.
45. Havlicek LL, Peterson NL. Robustness of the Pearson correlation against violations of assumptions. *Percept. Mot. Skills.* (1976) 43:1319–34. doi: 10.2466/pms.1976.43.3f.1319
46. Shami L, Lazebnik T. Implementing machine learning methods in estimating the size of the non-observed economy. *Comput. Econ.* (2023). 1–18. doi: 10.1007/s10614-023-10369-4
47. Rokach L. Decision Forest: twenty years of research. *Information Fusion.* (2016) 27:111–25. doi: 10.1016/j.inffus.2015.06.005
48. Belgiu M, Dragut L. Random forest in remote sensing: a review of applications and future directions. *ISPRS J.* (2016) 114:24–31. doi: 10.1016/j.isprsjprs.2016.01.011
49. Jiang L, Cai Z, Wang D, Jiang S. Survey of improving k-nearest-neighbor for classification. In fourth international conference on fuzzy systems and knowledge discovery. *IEEE.* (2007) 1:679–83. doi: 10.1109/FSKD.2007.552
50. Patro S, Sahu KK. Normalization: a preprocessing stage. *arXiv.* (2015). doi: 10.48550/arXiv.1503.06462
51. Kalousis A, Prados J, Hilario M. Stability of feature selection algorithms, Fifth IEEE international conference on data mining, Houston, TX, USA (2005), pp.8.
52. Tsai IC, Mayo LM, Jones BW, Stone AE, Janse SA, Bewley JM. Precision dairy monitoring technologies use in disease detection: differences in behavioral and physiological variables measured with precision dairy monitoring technologies between cows with or without metritis, hyperketonemia, and hypocalcemia. *Livestock Sci.* (2021) 244:104334. doi: 10.1016/j.livsci.2020.104334
53. Soriani N, Trevisi E, Calamari L. Relationships between rumination time, metabolic conditions, and health status in dairy cows during the transition period. *J. Anim. Sci.* (2012) 90:4544–54. doi: 10.2527/jas.2011-5064
54. Barker ZE, Vazquez Diosdado JA, Codling EA, Bell NJ, Hodges HR, Croft DP, et al. Use of novel sensors combining local positioning and acceleration to measure feeding behavior differences associated with lameness in dairy cattle. *J. Dairy Sci.* (2018) 101:6310–21. doi: 10.3168/jds.2016-12172
55. Harris-Bridge G, Young L, Handel I, Farish M, Mason C, Mitchell MA, et al. The use of infrared thermography for detecting digital dermatitis in dairy cattle: what is the best measure of temperature and foot location to use? *Vet. J.* (2018) 237:26–33. doi: 10.1016/j.tvjl.2018.05.008
56. Pirkkalainen H, Talvio I, Kujala-Wirth M, Soveri T, Orro T. Acute phase response of sole ulcer, white line disease and digital dermatitis in dairy cows. *Vet. Anim. Sci.* (2022) 17:100253. doi: 10.1016/j.vas.2022.100253
57. Benaissa S, Tuytens FAM, Plets D, Martens L, Vandaele L, Joseph W, et al. Improved cattle behaviour monitoring by combining ultra-wideband location and accelerometer data. *Animal.* (2023) 17:100730. doi: 10.1016/j.animal.2023.100730
58. Hosseininorbin S, Layeghy S, Kusy B, Jurdak R, Bishop-Hurley GJ, Greenwood PL, et al. Deep learning-based cattle behaviour classification using joint time-frequency data representation. *Comput. Electr. Agric.* (2021) 187:106241. doi: 10.1016/j.compag.2021.106241
59. Cabezas J, Yubero R, Visitacion B, Navarro-Garcia J, Algar MJ, Cano EL, et al. Analysis of accelerometer and GPS data for cattle behaviour identification and anomalous events detection. *Entropy.* (2022) 24:336. doi: 10.3390/e24030336
60. Gibbons JM, Lawrence AB, Haskell MJ. Measuring sociability in dairy cows. *Appl. Anim. Behav. Sci.* (2010) 122:84–91. doi: 10.1016/j.applanim.2009.11.011
61. Proudfoot KL, Jensen MB, Weary DM, von Keyserlingk MAG. Dairy cows seek isolation at calving and when ill. *J. Dairy Sci.* (2014) 97:2731–9. doi: 10.3168/jds.2013-7274
62. Ruuska S, Kajava S, Mughal M, Zehner N, Mononen J. Validation of a pressure sensor-based system for measuring eating, rumination and drinking behaviour of dairy cattle. *Appl. Anim. Behav. Sci.* (2016) 174:19–23. doi: 10.1016/j.applanim.2015.11.005
63. Achour B, Belkadi M, Filali I, Laghrouche M, Lahdir M. Image analysis for individual identification and feeding behaviour monitoring of dairy cows based on convolutional neural networks (CNN). *Biosyst. Eng.* (2020) 198:31–49. doi: 10.1016/j.biosystemseng.2020.07.019
64. Ren K, Bernes G, Hetta M, Karlsson J. Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning. *J. Syst. Archit.* (2021) 116:102139. doi: 10.1016/j.sysarc.2021.102139



OPEN ACCESS

EDITED BY

Flaviana Gottardo,
University of Padua, Italy

REVIEWED BY

Matthew R. Beck,
United States Department of Agriculture
(USDA), United States
Temple Grandin,
Colorado State University, United States

*CORRESPONDENCE

Dinu Gavojdian
✉ gavojdian_dinu@animalsci-tm.ro

RECEIVED 17 December 2023

ACCEPTED 19 January 2024

PUBLISHED 01 February 2024

CITATION

Gavojdian D, Mincu M, Lazebnik T, Oren A,
Nicolae I and Zamansky A (2024) BovineTalk:
machine learning for vocalization analysis of
dairy cattle under the negative affective state
of isolation.

Front. Vet. Sci. 11:1357109.

doi: 10.3389/fvets.2024.1357109

COPYRIGHT

© 2024 Gavojdian, Mincu, Lazebnik, Oren,
Nicolae and Zamansky. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

BovineTalk: machine learning for vocalization analysis of dairy cattle under the negative affective state of isolation

Dinu Gavojdian^{1*}, Madalina Mincu¹, Teddy Lazebnik^{2,3},
Ariel Oren⁴, Ioana Nicolae¹ and Anna Zamansky⁴

¹Cattle Production Systems Laboratory, Research and Development Institute for Bovine, Balotesti, Romania, ²Department of Mathematics, Ariel University, Ariel, Israel, ³Department of Cancer Biology, University College London, London, United Kingdom, ⁴Tech4Animals Laboratory, Information Systems Department, University of Haifa, Haifa, Israel

There is a critical need to develop and validate non-invasive animal-based indicators of affective states in livestock species, in order to integrate them into on-farm assessment protocols, potentially via the use of precision livestock farming (PLF) tools. One such promising approach is the use of vocal indicators. The acoustic structure of vocalizations and their functions were extensively studied in important livestock species, such as pigs, horses, poultry, and goats, yet cattle remain understudied in this context to date. Cows were shown to produce two types of vocalizations: low-frequency calls (LF), produced with the mouth closed, or partially closed, for close distance contacts, and open mouth emitted high-frequency calls (HF), produced for long-distance communication, with the latter considered to be largely associated with negative affective states. Moreover, cattle vocalizations were shown to contain information on individuality across a wide range of contexts, both negative and positive. Nowadays, dairy cows are facing a series of negative challenges and stressors in a typical production cycle, making vocalizations during negative affective states of special interest for research. One contribution of this study is providing the largest to date pre-processed (clean from noises) dataset of lactating adult multiparous dairy cows during negative affective states induced by visual isolation challenges. Here, we present two computational frameworks—deep learning based and explainable machine learning based, to classify high and low-frequency cattle calls and individual cow voice recognition. Our models in these two frameworks reached 87.2 and 89.4% accuracy for LF and HF classification, with 68.9 and 72.5% accuracy rates for the cow individual identification, respectively.

KEYWORDS

cattle, animal communication, affective states, vocal parameters, welfare indicators

Introduction

Farm animal welfare is commonly defined as the balance between positive and negative emotions, where positive emotions are considered as the main indicators of a moral animal life (“a life worth living” concept), with most of the recent research body of literature outlining the importance of affective states in farmed animals’ health and wellbeing (1, 2). Non-human mammals’ affective states might vary in valence (positive to negative) and arousal levels (high

to low), having functional adaptations linked to behavioral decisions that facilitate individual survival and reproduction while promoting approaches toward rewards and avoidance (3, 4).

There is an evident need to develop valid non-invasive animal-based indicators of emotions in domestic animals in order to integrate them into future on-farm assessment protocols, potentially via the use of precision livestock farming (PLF) tools, such as novel sensors (5, 6). To this end, the use of bioacoustics to evaluate health, emotional states, and stress responses has been validated for some of the most important livestock species such as pigs (*Sus scrofa domestica*), goats (*Capra hircus*), horses (*Equus caballus*), and poultry (*Gallus gallus domesticus*). The research findings consistently show that vocal parameters differ substantially during positive and negative experiences (7–15). Consequently, these developments started to be implemented and used in commercial settings in order to automatically classify animal vocalizations and identify health issues. For instance, the AI-based solution SoundTalks® was introduced in pig farms to detect respiratory diseases. However, compared to the aforementioned species, there is a significant knowledge gap regarding cattle communication behavior (16, 17). A potential explanation of this might be that cattle have a lower incidence of emitting vocalizations (16, 18), especially alarm and pain-specific vocalizations, developed as an adaptive response of the species as prey animals in order to avoid the risk of alarming potential predators.

Domestic cattle vocalizations were shown to contain information on individuality, given the high levels of inter-cow variability in the acoustic characteristic of the vocalizations emitted under various contexts, as well as allowing facilitation of short- and long-distance interactions with herd-mates. This variability found in vocalizations produced by cattle allows for each animal to be identified by the “fingerprint” of their call (19–22). Cattle are highly gregarious and form complex social relationships, having a strong innate motivation for continuous social contact (23, 24), with isolation from conspecifics resulting in physiological changes such as increased heart rates, cortisol levels, ocular and nasal temperature, and an increase in vocalization production (22, 25). Furthermore, it was suggested that individual cattle vary in susceptibility to emotional stressors and challenges (26–28), with limited research being undertaken to evaluate the effects that isolation over prolonged periods of time has on vocalization response in adult cattle. Throughout a typical production cycle, dairy cows face a series of negative emotional challenges and stressors, such as separation from a calf immediately after calving, frequent regrouping based on production levels and lactation phase, re-establishing social hierarchy and dominance, frequent milking, isolation from herd-mates for insemination, pregnancy check-ups, being at high risks of developing metabolic disorders, isolation in sickness pens, etc.

Cattle are known to produce two types of vocalizations, which are modulated by the configuration of the supra-laryngeal vocal tract (21). The first type is low-frequency calls (LF), produced by the animal with the mouth closed or partially closed, used for close distance contact, and regarded as indicative of lower distress or positive emotions. The second type is open-mouth emitted high-frequency calls (HF), produced for long-distance communication, and indicating higher arousal emotional states, generally associated with negative affective states (22, 29).

In domestic ungulates, individuality was proven to be encoded in a wide range of vocal parameters, most evidently in the F0-contour (15, 30), amplitude contour and call duration (31), as well as in filter-related vocal parameters including formant frequencies (21). Individuality expression was shown to be distinct for each call type (22, 30), with individual differences in cattle high-frequency calls being attributed mainly to sound formants (21, 22), while vocalization formants are being modulated in turn by the caller vocal tract morphology (32). Given that cattle can produce vocalizations with fundamental frequencies of over 1,000 Hz (30), which are more likely to occur during times of higher arousal affective states, it was hypothesized by the authors that high-frequency calls encode a larger amount of individuality information, than their low-frequency equivalents, due to their propagation over longer distances where vision and/or olfactory signaling are not possible.

Methods of studying animal vocal communication are becoming increasingly automated, with a growing body of research validating the use of both hardware and software that are capable of automatically collecting and processing bioacoustics data [reviewed by McLoughlin et al. (18)]. In this vein, Shorten and Hunter (33) found significant variability in cattle vocalization parameters, and suggested that such traits can be monitored using animal-attached acoustic sensors in order to provide information on the welfare and emotional state of the animal. Therefore, automated vocalization monitoring could prove to be a useful tool in precision livestock farming (18, 34, 35), especially as dairy farming systems become increasingly automated with wide-scale use of milking and feeding robots, all this having the potential to dynamically adjust the management practices while the number of animals per farm unit tends to increase.

Machine learning techniques are therefore increasingly applied in the study of cattle vocalizations. Some tasks that have been addressed to date include the classification of high vs. low frequency calls (33), ingestive behavior (35), and categorization of calls such as oestrus and coughs (34).

This study makes the following contributions to the investigation of cattle vocalizations using machine learning techniques. First of all, we present the largest dataset to date of ($n = 20$) cows' vocalizations collected under a controlled “station” setting, exclusively for negative affective states. Furthermore, we develop two types of AI models: deep-learning-based and explainable machine-learning-based for two tasks: (1) classification of high and low-frequency calls, and (2) individual cow identification. Finally, we investigated the feature importance of the explainable models.

Materials and methods

Ethical statement

All experiments were performed in accordance with relevant guidelines and regulations. The experimental procedures and protocols were reviewed and approved by the Ethical Committee from the Research and Development Institute for Bovine, Balotesti, Romania (approval no. 0027, issued on July 11, 2022), with the isolation challenge producing exclusively temporary distress to cows.

Subjects and experimental approach

The study was carried out at the experimental farm of the Research and Development Institute for Bovine in Balotesti, Romania. At the experimental facilities, cattle were managed indoors year-round (zero-grazing system), being housed under tie-stall conditions (stanchion barn) in two identical animal barns with a housing capacity of 100 heads/barn, having access to outdoor paddocks (14–16 m²/head) 10 h/day, between milkings (7:00–17:00). Cows had *ad libitum* access to water and mineral blocks, receiving a daily feed ration of 30 kg corn silage (37% dry matter, DM), 6 kg of alfalfa hay (dehydrated whole plant, 90.5% DM), and 6 kg of concentrates (88.3% DM). Concentrates were given immediately after morning and evening milkings (3 kg/feeding session); while silage and hay were offered at the feeding rails in the outside paddocks, assuring a feeding space of 0.7 m/cow. In total, 20 lactating adult multiparous cows of the Romanian Holstein breed were tested between August and September 2022. The Romanian Holstein breed (RH, national name Bălțată cu Negru Românească) belongs to the dairy Holstein-Friesian strain, with a current census of 264,000 cows, representing 22% of the breed structure in Romania (36). The RH originates in the 19th century, being the result of systematic crossbreeding between Friesian bulls imported from Denmark and Germany and local cattle, such as Romanian Spotted and Dobruja Red. The average milk yield for the RH breed ranges between 6,000 and 8,500 kg of milk/lactation, with adult body weights of cows ranging between 550 and 650 kg. The selection index for the RH breed is focused on milk yield (90%) and fertility related traits (10%) (37). In order to avoid bias and to have a homogeneous study group, cattle included in our research were of similar age (lactations II & III), were habituated previously to the housing system (min. 40 days in milk), and were comparable for body weight (619.5 ± 17.40 kg, mean \pm SEM). Cows were individually isolated inside the two identical barns, remaining tethered at the neck to their stall (stall dimensions of 1.7/0.85 m), starting at 7:00 AM for 240 consecutive minutes post-milking when the rest of the herd members were moved to the outside paddocks. The two paddocks were in the immediate vicinity of the barns, having one lateral shared concrete compact wall, thus the animal that remained inside the barn was visually isolated, while being able to hear and communicate vocally with their herd-mates. The isolated cows had *ad-libitum* access to water throughout individual drinkers and fresh wheat straws bedding was provided for animal comfort. After the commencement of recordings, animal caretakers were restricted from access to barns, and human traffic and machinery noise production outside the two barns were limited as much as possible.

Vocalization recordings

The vocalizations for this study were obtained using two identical directional microphones (Sennheiser MKH416-P48U3, frequency response 40–20,000 Hz, max. sound pressure level 130 dB at 1 kHz, producer Sennheiser Electronic®, Wede-mark, Germany) attached to Marantz PMD661 MKIII digital solid-state recorders (with file encryption, WAV recording at 44.1/48/96 kHz, 16/24-bit, recording bit rates 32–320 kbps, producer Marantz Professional®, United Kingdom). The microphones were directed toward the animal using tripods placed on the central feeding alleys at a

distance of 5–6 m from the cows. For shock and noise reduction, Sennheiser MZW 415 ANT microphone windshields were used. After the end of each experimental day, vocal recordings were saved as separate files in the WAV uncompressed format, at 44.1 kHz sampling rate and a 16-bit amplitude resolution. Despite the fact that all 20 multiparous lactating cows were isolated and recorded for 240 min post-milking, under identical conditions, not all cows vocalized with a similar frequency during the trials, resulting in the analysis of 1,144 vocalizations (57.2 vocalizations per cow, ranging between 33 and 90 vocalizations per cow), out of which 952 were high-frequency vocalizations (HF) and 192 low-frequency vocalizations (LF). All sounds included in our investigation had undergone quality control check, while looking for clear, under- and un-saturated vocalizations, without combined environmental noises such as rattling equipment, chains clanging, or wind. Vocalizations were visualized on spectrograms using the fast Fourier transform method, at window lengths of 0.03 s, time steps of 1,000, frequency steps of 250, dynamic range of 60 dB, and a view range between 0 and 5,000 Hz (Figure 1).

Vocalization recordings were then analyzed using Praat DSP package v.6.0.31 (38), as well as previously developed custom-built scripts (10, 15, 39–41), for the automatic extraction of the 23 acoustic features of each vocalization, with the vocal parameters studied and their definitions being presented in Table 1, the output data being exported to Microsoft Excel for further analysis.

Classification models

We developed two different computational frameworks of the following types:

- i Explainable model—a pipeline that uses as features the 23 vocal parameters described in Table 1, which have been studied in the context of cattle vocalizations. By using features that are highly relevant to our domain, we increase the explainability of our pipeline, allowing for the study of the feature importance of our model.
- ii Deep learning model, which uses learned features and operates as a “black box” that is not explainable. This model was expected to be more flexible and to have increased performance.

The explainable framework was based on the TPOT (42), AutoSklearn (43), and H2O (44) automatic machine learning libraries. Namely, we assumed a dataset represented by a matrix $x \in R^{n \times m}$ and a vector $y \in R^n$, where n is the number of rows and m is the number of features in the dataset. Notably, we used the features described in Table 1, which made the model more explainable, as the contribution of each feature to the model's prediction could be computed. This dataset was divided into training and testing sets, such that the first has 80% of the data and the latter the remaining 20%, divided randomly. The training cohort was used to train the model and the testing cohort was used to evaluate its performance. Moreover, we randomly picked 90% of the training dataset each time for $r = 50$ times, making sure each value was picked at least half of the times. For each of these cohorts, we first obtained a machine learning pipeline from TPOT, AutoSklearn, and H2O aiming to optimize the following loss function: $\sum_{i=1}^k \left(\frac{a^i + f^i}{2k} \right)$ where a^i and f^i are the i_{th} model's

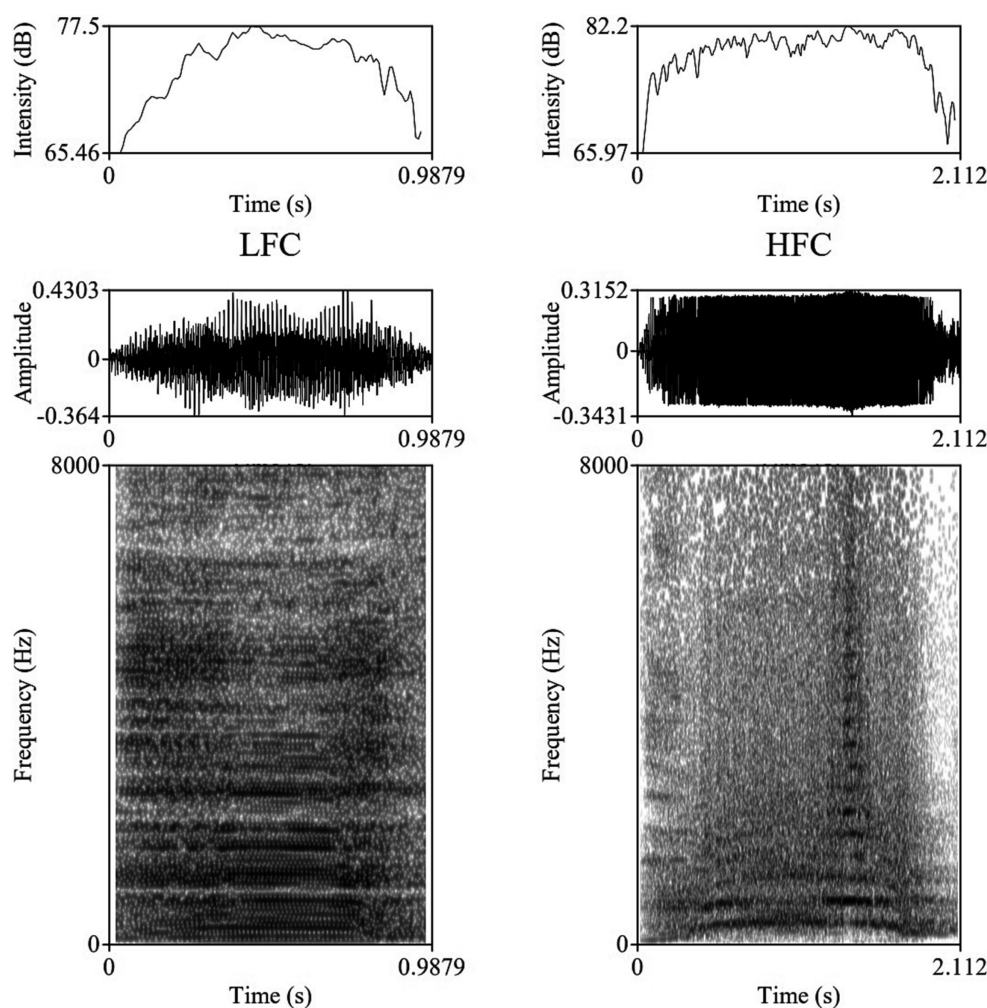


FIGURE 1

Example of a Low Frequency Call (LFC) left-side and a High Frequency Call (HFC) right-side with intensity contours (above), oscillograms (middle), and spectrograms [below; fast Fourier transform (FFT) method, at window lengths of 0.03 s, time steps of 1,000, frequency steps of 250, dynamic range of 60 dB, and a view range between 0 and 5,000 Hz] of typical vocalizations produced by cows during the isolation challenge.

instance accuracy and F_1 scores, respectively, where k was the number of k -folds in the cross-validation analysis (45). Once all three models were obtained, we used all three of them to generate another cohort containing their predictions and the corresponding y value. These were then used to train an XGboost (46) model for the final prediction. For the hyperparameter tuning of the XGboost, we took advantage of the grid-search method. Finally, a majority vote between the r instances was used to determine the final model's prediction. We reported the results for $k=5$ fold cross-validation over the entire dataset.

The DL framework was adopted from Ye and Yang (47) which proposed a deep-gated recurrent unit (GRU) neural network (NN) model, combining a two-dimensional convolution NN and recurrent NN based on the GRU cell unit that gets as input the spectrogram of the audio signal. Generally, the two-dimensional convolution NN is used as a feature extraction component, finding spatio-temporal connections in the signal which than is being fed into the recurrent NN that operates as a temporal model able to detect short-and long-term connections in this feature space over time, these being effectively the rules for the voice identifications. For the

hyperparameters of the model such as batch size, learning rate, optimization, etc., we adopted the values from Ye and Yang (47).

Results

In this section, we examined the data obtained and outlined the performance of the proposed explainable and DL models for the two different tasks based on the collected dataset. First, we provided a descriptive statistical analysis of the obtained dataset and its properties. Secondly, we present the performance of the models in classification between high and low-frequency calls. Finally, we present the models' ability to identify each cow according to its vocalizations, divided into low, high, and all low + high vocalizations. Table 2 summarizes the results of the explainable and DL models' performance in separating between the high and low-frequency calls. The results are shown as mean \pm standard deviation for $k=5$ fold. Importantly, we made sure that the train and test cohorts had vocalizations from both classes at each fold and that the ratio between the classes was kept between folds. Both

TABLE 1 Abbreviations and definitions of the 23 vocal parameters studied (21, 29).

Abbreviation/Unit of measure	Definition
F0Mean (Hz)	Mean F0* frequency value across the call
F0Max (Hz)	Maximum F0 frequency value across the call
F0Min (Hz)	Minimum F0 frequency value across the call
F0Range (Hz)	Difference between minimum and maximum F0
Q25 (Hz)	Frequency value at the upper limit of the first quartiles of energy (below 25%)
Q50 (Hz)	Frequency value at the upper limit of the second quartiles of energy (below 50%)
Q75 (Hz)	Frequency value at the upper limit of the third quartiles of energy (below 75%)
Fpeak (Hz)	Frequency of peak amplitude
Sound duration (s)	Duration of the call from start to end, measured on the oscillogram
AMVar (dB/s)	Cumulative variation in amplitude divided by the total call duration
AMRate (s ⁻¹)	Number of amplitude modulations in a certain time frame
AMExtent (dB)	Mean-to-mean peak variation of each amplitude modulation
Harmonicity (dB)	Degree of acoustic periodicity, also called harmonic-to-noise ratio—higher values indicate more tonal voice
F1Mean (Hz)	Mean frequency value of the first formant
F2Mean (Hz)	Mean frequency value of the 2nd formant
F3Mean (Hz)	Mean frequency value of the 3rd formant
F4Mean (Hz)	Mean frequency value of the 4th formant
F5Mean (Hz)	Mean frequency value of the 5th formant
F6Mean (Hz)	Mean frequency value of the 6th formant
F7Mean (Hz)	Mean frequency value of the 7th formant
F8Mean (Hz)	Mean frequency value of the 8th formant
Formant dispersal (Hz)	Minimum spacing of the formants
Wiener entropy mean	Spectral flatness of a sound, calculated as the ratio of a power spectrum's geometric mean to its arithmetic mean measured on a logarithmic scale

TABLE 2 The high- and low-frequency calls (LF and HF) classifier models' performances.

Model	Train set accuracy	Test set accuracy
Explainable	89.9 ± 2.2%	87.2 ± 4.1%
Deep learning	91.5 ± 2.6%	89.4 ± 3.8%

The results are shown as mean standard deviation for $k = 5$ fold.

models achieved good results with almost nine out of 10 correct detections. One can notice that the DL model outperforms the explainable model. One explanation for this is that the DL is more expressive and therefore captures more complex dynamics, which are not necessarily expressed by the features provided to the explainable model (see Table 2). Figure 2 presents the features' importance of the explainable model for the high and low-frequency calls calculated by reducing one feature from the input and calculating its influence on the model's performance. One can notice that AMvar, AMrate, AMExtent, Formant dispersal, and the Wiener entropy mean are the most important features, with a joint importance of 55.36%.

Table 3 summarizes the results of the explainable and DL models' individual cow identification accuracy. The results are shown as mean ± standard deviation for $k = 5$ fold. For this case, we made sure

that the train and test cohorts had vocalizations from all cows such that the proportion of the vocalizations of each individual cow was present both in the train and test cohorts at each fold. The models obtained around 70% accuracy, with a relatively low standard deviation. Like the previous experiment, the DL model outperformed the explainable model.

Figure 3 presents the features' importance of the explainable model for the cow identification task, calculated by reducing one feature from the input and calculating its influence on the model's performance. The sound duration played a critical role with 14.27% importance, indicating that different cows have a significant pattern in their vocal duration, or at least a non-linear connection between the vocal duration to other features that allows for capturing unique identification patterns. The Wiener Entropy mean is the second best, with 11.65% importance.

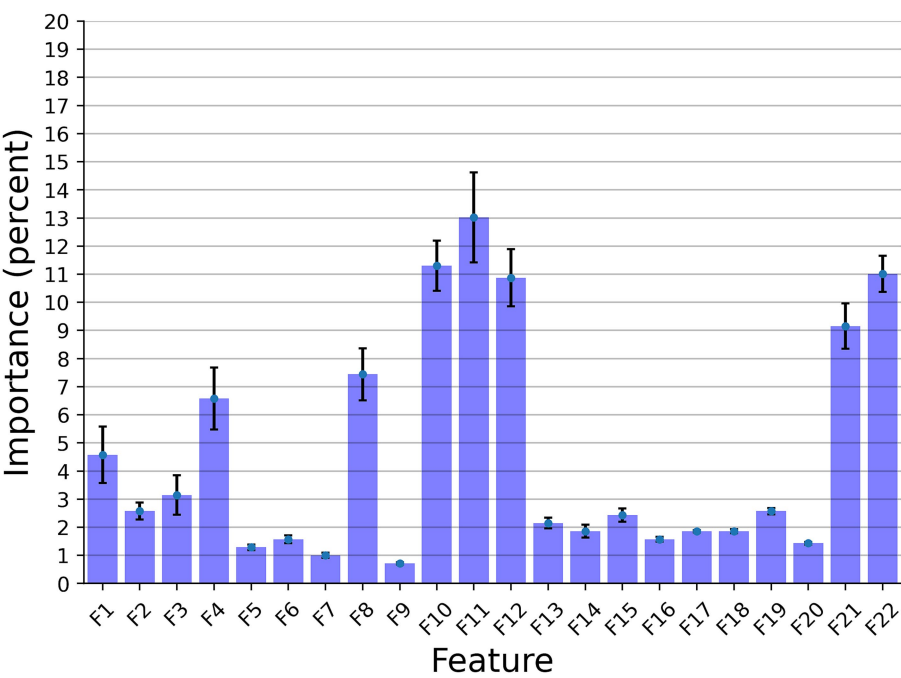


FIGURE 2 The distribution of the features' importance for the high and low frequency calls (LF and HF) explainable classifier model. The results are shown as the average of a $k = 5$ fold cross-validation where the error bars indicate one standard deviation.

TABLE 3 The individual identification classifier models' performances.

Dataset	Model	Train set accuracy	Test set accuracy
Low Frequency + High Frequency	Explainable	73.0 ± 3.3%	68.9 ± 5.1%
	Deep learning	76.3 ± 4.2%	72.5 ± 4.7%
Low Frequency	Explainable	58.2 ± 1.3%	50.9 ± 2.8%
	Deep learning	65.5 ± 1.8%	46.8 ± 3.3%
High Frequency	Explainable	79.6 ± 2.6%	68.4 ± 3.2%
	Deep learning	74.9 ± 3.0%	70.8 ± 3.4%

The results are shown as mean standard deviation for $k = 5$ fold.

Discussion

In this study, we present a dataset of cattle vocal recordings during negative affective states caused by isolation of the cows, which is, to the best of our knowledge, the largest dataset collected to date. The data from $n = 20$ cows has been manually cleaned from background noises and trimmed to contain only the low-frequency (LF) and high-frequency (HF) calls, to ensure as high quality of data as possible. The resulting dataset comprises 1,144 records in total. Based on this data, we conducted two sets of tasks. Firstly, we provided a classifier for separating between low and high-frequency calls. Secondly, we provided a classifier for identifying individual cows based on their high-, low-, or high + low-frequency vocalizations produced.

As shown in Table 2, both the explainable and DL models were able to accurately classify between the low- and high-frequency calls, with 87.2 and 89.4% accuracy, respectively. This outcome slightly outperforms (2%, 4.4%) the current state-of-the-art model (33), which used a smaller dataset of $n = 10$ individuals. Notably,

the differences between the models' performances between the training and testing cohorts was around 2.5% toward the training cohort, compared to the state-of-the-art, which reports a 14.2% difference. As such, our model resulted in less over-fitting, if any at all, than the previous model. In addition, as the standard deviations of both models were 4.1 and 3.8%, this indicates that both models are robust.

For the individual cow identification task for both the LF and HF data, the explainable and DL models obtained 68.9 and 72.5% accuracy, respectively. When focusing only on the HF calls, the results were similar, with only 0.5 and 1.7% decrease in performance. On the other hand, when using only the LF samples, the accuracy sharply dropped to 50.9 and 46.8%, respectively, while also revealing overfit over the training dataset. This may be an indication that high-frequency calls contain more individuality information than low-frequency calls in cattle. These results are in accordance with previous findings across non-human mammals (29, 48), where an increase in the arousal states was shown to lead to higher frequency vocalizations for both F0 and formant-related

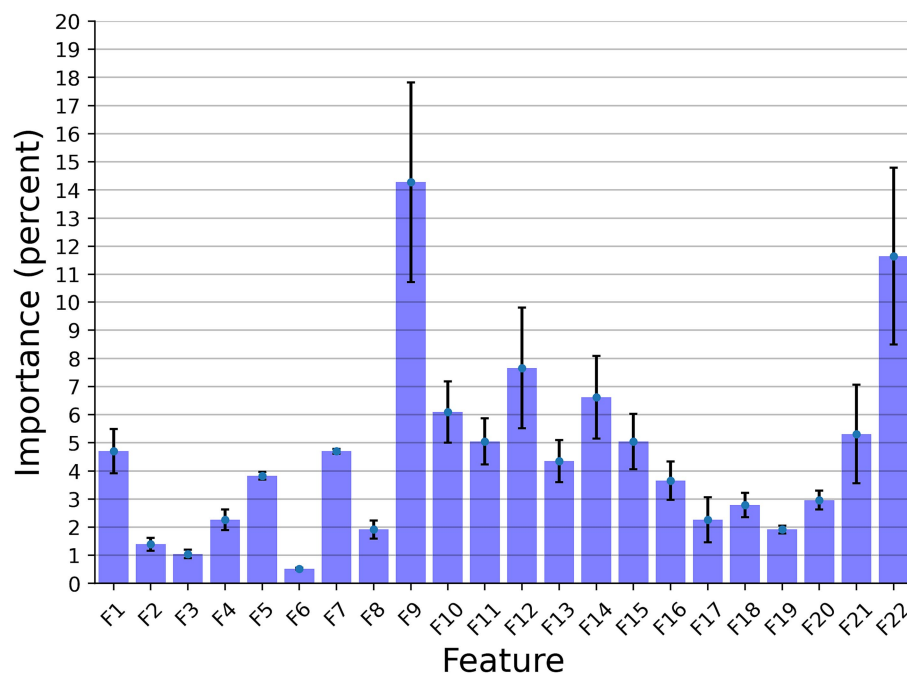


FIGURE 3

The distribution of the features' importance for identification explainable classifier model for the Low Frequency + High Frequency dataset. The results are shown as the average of a $k = 5$ fold cross-validation, where the error bars indicate one standard deviation.

features, with vocalization parameters being more variable in negative-high arousal states. An alternative explanation for this might be attributed to the reduced amount of LF data, which contained 192 samples (i.e., 16.8% of the entire dataset). While the performance of the model of Shorten and Hunter (33) was better, this study worked with a reduced dataset for LF calls. In addition, their results may be indicative of overfitting, while explainable frameworks were not considered.

Considering homologies in the physiology of vocalization production and the commonalities found across species (48), the current findings could be extrapolated to other European cattle (*Bos taurus*), both dairy and beef specialized breeds. In a comparative study conducted on the two cattle sub-species (*B. taurus* and *B. indicus*), Moreira et al. (49) found *B. indicus* animals to be more reactive to both low and high frequencies sounds, which the authors attributed to the smaller auricle and greater interaural distance found in *B. taurus*, when compared to the indicine cattle. Although the hearing range might differ among closely related species, Maigrot et al. (14) found the functions of vocalizations to exceed intraspecies exchanges of information in domestic horses and Przewalski's horses, wild boars, and domestic pigs, these species being able to discriminate among positive and negative vocalizations produced by heterospecifics, including humans. Moreover, another potential contribution of the current research becomes apparent based on the experimental design and data collection. Whereas the studies conducted on cattle communication behavior to-date analyzed predominantly vocalizations emitted by cows either in an un-controlled setting (e.g., mob on pasture or inside the barn), or assessed and compared calls among a wider set of contexts (e.g., positive and/or negative, with different putative valences and arousal levels) (21, 22, 33, 35).

Conversely, our experimental setting was exclusively focused on a single negative context, while changes in affective states of the same animal being proven previously to result in modulations of the vocal parameters and behavior, e.g., during dam-calf separation and reunion in beef cattle (50). However, it is worth pointing out that our isolation challenge replicates an event that is occurring frequently under production conditions, with cows being individually isolated for health (e.g., sickness and veterinary visits) and reproduction (e.g., artificial insemination, fertility treatments, and pregnancy diagnosis) reasons.

Moreover, the use of vocalization scoring as an indicator of welfare during cattle handling at slaughter was shown to be a feasible approach, with vocalizations frequency and cortisol levels being influenced by the use of electric prods, deficient stunning and aversive handling (51, 52).

Our results are in alignment with previous research which showed that isolation from herd-mates induces a wide range of behavioral and physiological responses in cattle (22, 25), given the much higher incidence of HF calls observed during the isolation challenge, and taking into account the previous research results which suggested that the production and broadcasting of a repetitive single call type is indicative of persistent negative affective states (53), while reflecting a high urgency for the animal itself (54).

This research is, however, not without its limitations. Factors such as emotional contagion among herd-mates, and thus the potential biological role of the distress vocalizations emitted by cows during the isolation challenge were not studied in the current trial. For instance, Rhim (55) found that vocalization and behavior of Holstein cows and calves during partial separation (with vocal and olfactory communication) has led to significantly higher vocalization rates in

both cows and calves, when compared to complete separation. To address this, in our future research, we plan to include the use of additional sensors such as heart rate monitors, infrared thermography, and stress-related biomarkers, to have a more generalized approach when evaluating emotional responses to negative contexts. Moreover, considering the psychology and behavioral patterns of the species, mental processes such as learned helplessness could have contributed to the time-modulation of the vocal parameters following herd isolation, with animals abandoning their attempts to signal the negative event due to a perceived lack of control, which, however, does not mean that the negative event is being perceived as neutral by the animals. Additionally, the study herd consisted of multiparous adult cows, with various degrees of existing habituation to social isolation being presumed.

To summarize, cattle vocalizations can be seen as commentaries emitted by an individual on their own internal affective state, with the challenges lying in understanding and deciphering those commentaries. Looking forward, significantly more work needs to be done, taking into account a wider range of contexts and potential influencing factors on the vocal cues, in order to be able to draw strong conclusions regarding arousal or valence in cattle bioacoustics. Our study highlights the promising applications of machine learning approaches in cattle vocalization behavior. In order for such approaches to be validated for commercial use and adopted at farm-level, further important developments need to occur, such as designing of hardware and software capable to filter and limit external farm noises, and to process vocalizations automatically. Furthermore, in order for vocalizations to be reliable welfare and health predictors, the training and testing vocalization sets need to include a significant wider variety of negative affective states (e.g., sickness, cow-calf separation, weaning, estrous, and pain inducing vocalizations), all while keeping a special focus on changes in individual animal vocalizations.

Conclusion

In this study, we compiled a data set of dairy cattle vocal recordings during the negative affective state of isolation, which is one of the largest and cleanest datasets of its kind. Through our experiments using explainable and DL models, we have demonstrated the effectiveness of these models in classifying high- and low-frequency calls, as well as for identifying individual cows based on their vocalization productions. These results highlight the future potential of vocalization analysis as a valuable tool for assessing the emotional valence of cows and for providing new insights into promoting precision livestock farming practices. By monitoring cattle vocalizations, animal scientists could gain crucial insights into the emotional states of the animals, empowering them to make informed decisions to improve the overall farm animal welfare. Future work endeavors can take these results a step forward, gathering cattle vocalizations at different critical affective states to identify possible health risks or early real-time disease diagnostics.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession

number(s) can be found here: <https://gitlab.com/is-annazam/bovinetalk>.

Ethics statement

The experimental procedures and protocols were reviewed and approved by the Ethical Committee from the Research and Development Institute for Bovine, Balotesti, Romania (approval no. 0027, issued on July 11, 2022). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

DG: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. MM: Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. TL: Data curation, Formal Analysis, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AO: Data curation, Formal Analysis, Investigation, Software, Validation, Writing – original draft, Writing – review & editing. IN: Formal Analysis, Investigation, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. AZ: Data curation, Formal Analysis, Investigation, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS—UEFISCDI, project number PN-III-P1-1.1-TE-2021-0027, within PNCDI III.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Webster J. Animal welfare: freedoms, dominions and “a life worth living”. *Animals*. (2016) 6:35. doi: 10.3390/ani6060035
- Webb LE, Veenhoven R, Harfeld JL, Jensen MB. What is animal happiness? *Ann N Y Acad Sci*. (2019) 1438:62–76. doi: 10.1111/nyas.13983
- Kremer L, Holkenborg SK, Reimert I, Bolhuis J, Webb L. The nuts and bolts of animal emotion. *Neurosci Biobehav Rev*. (2020) 113:273–86. doi: 10.1016/j.neubiorev.2020.01.028
- Laurijs KA, Briefer EF, Reimert I, Webb LE. Vocalisations in farm animals: a step towards positive welfare assessment. *Appl Anim Behav Sci*. (2021) 236:105264. doi: 10.1016/j.applanim.2021.105264
- Ben Sassi N, Averós X, Estevez I. Technology and poultry welfare. *Animals*. (2016) 6:62. doi: 10.3390/ani6100062
- Matthews SG, Miller AL, Clapp J, Plötz T, Kyriazakis I. Early detection of health and welfare compromises through automated detection of behavioural changes in pigs. *Vet J*. (2016) 217:43–51. doi: 10.1016/j.tvjl.2016.09.005
- Silva M, Ferrari S, Costa A, Aerts JM, Guarino M, Berckmans D. Cough localization for the detection of respiratory diseases in pig houses. *Comput Electron Agric*. (2008) 64:286–92. doi: 10.1016/j.compag.2008.05.024
- Tallet C, Linhart P, Policht R, Hammerschmidt K, Šimeček P, Kratinova P, et al. Encoding of situations in the vocal repertoire of piglets (*sus scrofa*): a comparison of discrete and graded classifications. *PLoS One*. (2013) 8:e71841. doi: 10.1371/journal.pone.0071841
- Whitaker B. M., Carroll B. T., Daley W., Anderson D. V. (2014). “Sparse decomposition of audio spectrograms for automated disease detection in chickens” in IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, 2014. 1122–1126.
- Briefer EF, Tettamanti F, McElligott AG. Emotions in goats: Mapping physiological, behavioural and vocal profiles. *Anim Behav*. (2015) 99:131–43. doi: 10.1016/j.anbehav.2014.11.002
- Linhart P, Ratcliffe VF, Reby D, Špinková M. Expression of emotional arousal in two different piglet call types. *PLoS One*. (2015) 10:e0135414. doi: 10.1371/journal.pone.0135414
- Briefer EF, Mandel R, Maigrot AL, Briefer Freymond S, Bachmann I, Hillmann E. Perception of emotional valence in horse whinnies. *Front Zool*. (2017) 14:1–12. doi: 10.1186/s12983-017-0193-1
- McGrath N, Dunlop R, Dwyer C, Burman O, Phillips CJ. Hens vary their vocal repertoire and structure when anticipating different types of reward. *Anim Behav*. (2017) 130:79–96. doi: 10.1016/j.anbehav.2017.05.025
- Maigrot A-L, Hillmann E, Briefer EF. Encoding of emotional valence in wild boar (*sus scrofa*) calls. *Animals*. (2018) 8:85. doi: 10.3390/ani8060085
- Briefer EF, Sypherd CCR, Linhart P, Leliveld LMC, Padilla de la Torre M, Read ER, et al. Classification of pig calls produced from birth to slaughter according to their emotional valence and context of production. *Sci Rep*. (2022) 12:3409. doi: 10.1038/s41598-022-07174-8
- Ede T, Lecorps B, von Keyserlingk MA, Weary DM. Symposium review: scientific assessment of affective states in dairy cattle. *J Dairy Sci*. (2019) 102:10677–94. doi: 10.3168/jds.2019-16325
- Green AC, Lidfors LM, Lomax S, Favaro L, Clark CE. Vocal production in postpartum dairy cows: temporal organization and association with maternal and stress behaviors. *J Dairy Sci*. (2021) 104:826–38. doi: 10.3168/jds.2020-18891
- McLoughlin MP, Stewart R, McElligott AG. Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring. *J Royal Soc Interface*. (2019) 16:20190225. doi: 10.1098/rsif.2019.0225
- Watts JM, Stookey JM. The propensity of cattle to vocalise during handling and isolation is affected by phenotype. *Appl Anim Behav Sci*. (2001) 74:81–95. doi: 10.1016/S0168-1591(01)00163-0
- Yajuvendra S, Lathwal SS, Rajput N, Raja TV, Gupta AK, Mohanty TK, et al. Effective and accurate discrimination of individual dairy cattle through acoustic sensing. *Appl Anim Behav Sci*. (2013) 146:11–8. doi: 10.1016/j.applanim.2013.03.008
- de la Torre MP, Briefer EF, Reader T, McElligott AG. Acoustic analysis of cattle (*Bos taurus*) mother–offspring contact calls from a source–filter theory perspective. *Appl Anim Behav Sci*. (2015) 163:58–68. doi: 10.1016/j.applanim.2014.11.017
- Green A, Clark C, Favaro L, Lomax S, Reby D. Vocal individuality of Holstein-Friesian cattle is maintained across putatively positive and negative farming contexts. *Sci Rep*. (2019) 9:18468. doi: 10.1038/s41598-019-54968-4
- Boissy A, Le Neindre P. Behavioral, cardiac and cortisol responses to brief peer separation and Reunion in cattle. *Physiol Behav*. (1997) 61:693–9. doi: 10.1016/S0031-9384(96)00521-5
- Holm L, Jensen MB, Jeppesen LL. Calves’ motivation for access to two different types of social contact measured by operant conditioning. *Appl Anim Behav Sci*. (2002) 79:175–94. doi: 10.1016/S0168-1591(02)00137-5
- Müller R, Schrader L. Behavioural consistency during social separation and personality in dairy cows. *Behaviour*. (2005) 142:1289–306. doi: 10.1163/156853905774539346
- van Reenen CG, O’Connell NE, van der Werf JTN, Korte SM, Hopster H, Jones RB, et al. Responses of calves to acute stress: individual consistency and relations between behavioral and physiological measures. *Physiol Behav*. (2005) 85:557–70. doi: 10.1016/j.physbeh.2005.06.015
- Lecorps B, Kappel S, Weary DM, von Keyserlingk MA. Dairy calves’ personality traits predict social proximity and response to an emotional challenge. *Sci Rep*. (2018) 8:16350. doi: 10.1038/s41598-018-34281-2
- Nogues E, Lecorps B, Weary DM, von Keyserlingk MA. Individual variability in response to social stress in dairy heifers. *Animals*. (2020) 10:1440. doi: 10.3390/ani10081440
- Briefer EF. Vocal expression of emotions in mammals: mechanisms of production and evidence. *J Zool*. (2012) 288:1–20. doi: 10.1111/j.1469-7998.2012.00920.x
- Volodin IA, Lapshina EN, Volodina EV, Frey R, Soldatova NV. Nasal and oral calls in juvenile goitred gazelles (*Gazella subgutturosa*) and their potential to encode sex and identity. *Ethology*. (2011) 117:294–308. doi: 10.1111/j.1439-0310.2011.01874.x
- Sèbe F, Poindron P, Ligout S, Sèbe O, Aubin T. Amplitude modulation is a major marker of individual signature in lamb bleats. *Bioacoustics*. (2018) 27:359–75. doi: 10.1080/09524622.2017.1357146
- Taylor AM, Charlton BD, Reby D. Vocal production by terrestrial mammals: source, filter, and function. *Vertebrate Sound Product Acoustic Commun*. (2016) 53:229–59. doi: 10.1007/978-3-319-27721-9_8
- Shorten P, Hunter L. Acoustic sensors for automated detection of cow vocalization duration and type. *Comput Electron Agric*. (2023) 208:107760. doi: 10.1016/j.compag.2023.107760
- Jung DH, Kim NY, Moon SH, Jhin C, Kim HJ, Yang JS, et al. Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering. *Animals*. (2021) 11. doi: 10.3390/ani11020357
- Li G, Xiong Y, Du Q, Shi Z, Gates RS. Classifying ingestive behavior of dairy cows via automatic sound recognition. *Sensors*. (2021) 21:5231. doi: 10.3390/s21155231
- Ilie DE, Gao Y, Nicolae I, Sun D, Li J, Han B, et al. Evaluation of single nucleotide polymorphisms identified through the use of SNP assay in Romanian and Chinese Holstein and Simmental cattle breeds. *Acta Biochim Pol*. (2020) 67:341–6. doi: 10.18388/abp.2020_5080
- Mincu M, Gavojdian D, Nicolae I, Olteanu AC, Vlagioiu C. Effects of milking temperament of dairy cows on production and reproduction efficiency under tied stall housing. *J Vet Behav*. (2021) 44:12–7. doi: 10.1016/j.jveb.2021.05.010
- Boersma P., Praat D.W. (2022). Doing phonetics by computer [computer program]. Available at: <http://www.praat.org/>
- Reby D, McComb K. Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Anim Behav*. (2003) 65:519–30. doi: 10.1006/anbe.2003.2078
- Gabriël J.L.B. (2004). Wiener entropy, script developed in praat v. 4.2.06. Available at: <https://gbeckers.nl/pages/phonetics.html>
- Briefer EF, Vizier E, Gyga L, Hillmann E. Expression of emotional valence in pig closed-mouth grunts: involvement of both source- and filter-related parameters. *J Acoust Soc Am*. (2019) 145:2895–908. doi: 10.1121/1.5100612
- Le TT, Fu W, Moore JH. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics*. (2020) 36:250–6. doi: 10.1093/bioinformatics/btz470
- Feurer M, Eggensperger K, Falkner S, Lindauer M, Hutter F. Auto-sklearn 2.0: hands-free automl via meta-learning. *J Mach Learn Res*. (2022) 23:11936–96. doi: 10.48550/arXiv.2007.04074
- Daugela K., Vaiciukynas E. (2022). “Real-time anomaly detection for distributed systems logs using apache kafka and h2o AI” in *International Conference on Information and Software Technologies*. Springer, 33–42.
- Fushiki T. Estimation of prediction error by using k-fold cross-validation. *Stat Comput*. (2011) 21:137–46. doi: 10.1007/s11222-009-9153-8
- Chen T., Guestrin C. (2016). “Xgboost: a scalable tree boosting system” in *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Ye F, Yang J. A deep neural network model for speaker identification. *Appl Sci*. (2021) 11:3603. doi: 10.3390/app11083603
- Briefer EF. Coding for ‘dynamic’ information: vocal expression of emotional arousal and valence in non-human animals In: T Aubin and N Mathevon, editors. *Animal signals and communication: coding strategies in vertebrate acoustic communication*. Springer, Cham. (2020) 7:137–62. doi: 10.1007/978-3-030-39200-0_6

49. Moreira SM, Barbosa Silveira ID, da Cruz LAX, Minello LF, Pinheiro CL, Schwengber EB, et al. Auditory sensitivity in beef cattle of different genetic origins. *J Vet Behav.* (2023) 59:67–72. doi: 10.1016/j.jveb.2022.10.004
50. Schnaider MA, Heidemann MS, Silva AHP, Taconeli CA, Molento CFM. Vocalization and other behaviors as indicators of emotional valence: the case of cow-calf separation and Reunion in beef cattle. *J Vet Behav.* (2022) 49:28–35. doi: 10.1016/j.jveb.2021.11.011
51. Grandin T. The feasibility of using vocalization scoring as an indicator of poor welfare during cattle slaughter. *Appl Anim Behav Sci.* (1998) 56:121–8. doi: 10.1016/S0168-1591(97)00102-0
52. Hemsworth PH, Rice M, Karlen MG, Calleja L, Barnett JL, Nash J, et al. Human–animal interactions at abattoirs: relationships between handling and animal stress in sheep and cattle. *Appl Anim Behav Sci.* (2011) 135:24–33. doi: 10.1016/j.applanim.2011.09.007
53. Collier K, Townsend SW, Manser MB. Call concatenation in wild meerkats. *Anim Behav.* (2017) 134:257–69. doi: 10.1016/j.anbehav.2016.12.014
54. Kershenbaum A, Blumstein DT, Roch MA, Akçay Ç, Backus G, Bee MA, et al. Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biol Rev.* (2016) 91:13–52. doi: 10.1111/brv.12160
55. Rhim SJ. Vocalization and behavior of Holstein cows and calves after partial and complete separation. *Rev Colomb Cienc Pecuarias.* (2013) 26:24–9.



OPEN ACCESS

EDITED BY

Jasmin Nicole Nessler,
University of Veterinary Medicine Hannover,
Germany

REVIEWED BY

James Edward Brereton,
Sparsholt College, United Kingdom
Yingxin Wang,
Chinese Academy of Forestry, China

*CORRESPONDENCE

Xiao-Bo Wang
✉ wangxiaobo@lzu.edu.cn

RECEIVED 28 February 2024

ACCEPTED 06 May 2024

PUBLISHED 11 June 2024

CITATION

Yang X, Daraz U, Ma J, Lu X, Feng Q,
Zhu H and Wang X-B (2024) Temporal–spatial
variability of grazing behaviors of yaks and the
drivers of their intake on the eastern
Qinghai-Tibetan Plateau.
Front. Vet. Sci. 11:1393136.
doi: 10.3389/fvets.2024.1393136

COPYRIGHT

© 2024 Yang, Daraz, Ma, Lu, Feng, Zhu and
Wang. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Temporal–spatial variability of grazing behaviors of yaks and the drivers of their intake on the eastern Qinghai-Tibetan Plateau

Xiaoqian Yang, Umar Daraz, Jianguo Ma, Xingxin Lu,
Qingshan Feng, Huaide Zhu and Xiao-Bo Wang*

State Key Laboratory of Herbage Improvement and Grassland Agro-Ecosystems, College of Pastoral Agriculture Science and Technology, Center for Grassland Microbiome, Lanzhou University, Lanzhou, China

Introduction: Grassland-livestock balance is an important principle of sustainable development of grassland livestock production and grassland ecosystem health. Grassland degradation becomes more serious at global scales and especially at the area that is sensitive to climate change and human activities. Decreases in pasture biomass and shifts in plant community composition in degraded grasslands can largely affect grazing behaviors of livestock. Up to date, however, it is unclear that whether livestock behaviors change across spatial and temporal scales and what key factors are to shape observed behavioral patterns of livestock.

Methods: Here, yak behaviors including grazing, rumination and walking on the eastern Qinghai-Tibetan Plateau (QTP) were monitored by a continuous visual observation, to investigate temporal and spatial variations of grazing behavior of yaks (*Bos grunniens*); based on the data from public database in the past 18 years, a meta-analysis was conducted to examine the main factors that affect grazing behaviors and intake of yaks.

Results: We showed that grazing behaviors of yaks differed significantly within hours, among hours of each day and among days as well as across different observation sites. Intake rate of yaks was higher in the morning than in the afternoon, but walking speed showed an inverse trend compared with intake rate. Resting, altitude, the mean annual precipitation (MAP), the mean annual temperature (MAT), forage ash, yak age and season were the main predictors for yak intake, and forage and yak individual characteristics had direct effects on grazing behaviors and intake of yaks.

Discussion: The findings confirm that grazing behaviors of yaks can vary even at small temporal scales and regional scales, which is closely related to the shift in forage quality and biomass caused by environmental changes. The study suggests that multiple factors can be responsible for the variation in livestock behaviors and shifts in behavioral patterns may consequently lead to positive or negative feedback to grassland ecosystems through plant-animal interactions.

KEYWORDS

grazing behaviors, intake, yak, Qinghai-Tibet Plateau, grassland

1 Introduction

Livestock are key components in a natural grassland ecosystem and play essential roles in regulating grassland ecosystem health and services. Traditionally, since the land use in the grassland

is associated closely with animal husbandry, many grasslands worldwide have heavily relied on grazing for hundreds of years for the purpose of satisfying increasing demands for products (1, 2). However, negative effects of anthropogenic activities on rangeland ecosystems are being intensified due to grazing-induced shifts in ecosystem structure, functioning and stability (3, 4). Particularly, at local and regional scales, changes in plant community composition such as decreased numbers of palatable forage in degraded grasslands, have affected heavily on the behavior of livestock, especially for large herbivores (5). Livestock behaviors can also produce an impact on above-belowground ecological processes such as plant succession and nutrient cycling via grazing, treading and excreta return (4, 6). Understanding the changes of livestock behaviors can provide important parameters for modeling livestock intake and improve predictions of grassland ecosystem health, and consequently achieving sustainable management of livestock grazing in grassland ecosystems (6, 7).

Grazing, ruminating, and walking are the three primary activities carried out by livestock. Generally, free-ranging livestock spend much energy on grazing and walking. It is reported that ruminants spend 90–95% of their daily time grazing, ruminating, and resting in the pasture-based system (8). This may result in a significant increase in the amount of energy that is consumed (6, 9). The higher consumption of energy associated with physical activity may raise animals' maintenance energy requirements and reduce the energy availability for growth and development (10). The daily intake capability of forage is thus dependent greatly on the amount of time spent grazing and the rate of forage consumption throughout that period. Daily consumption of forage is proportional to the number of bites per unit time (bite rate) and the mass of forage consumed per bite (bite mass) (11). Rumination is the behavior utilized by ruminants after grazing, which is crucial for feed breakdown because it raises the specific gravity of forages, shreds plant tissue coverings, and provides more of the forage surface area to the rumen microbiota. Some studies have shown that grazing and ruminating behaviors of ruminants are cyclical (12) and can change based on the forage quality and types, environmental conditions, individual characteristics of livestock, and different grazing intensities (13–16).

Livestock behaviors in natural grassland vary temporally and spatially, depending strongly on resource availability and changing environments that they live (17, 18). Due to the energy and metabolic demands, vegetation dynamics is mainly responsible for livestock behaviors. Plant composition and distribution have been well demonstrated to vary largely over space and time (5, 19). One of the important consequences of such variations is to lead to temporal and spatial variations in livestock behaviors. For example, ruminants' grazing behavior often changes with a shift in herbage biomass and pasture nutritional quality (20). In a ranch with abundant vegetation, ruminants generally gather around the areas that have good quality forage (21, 22). In a ranch with spatially homogenous resources, the herbage resources are often utilized through selective grazing by ruminants to meet their nutrient demand and energy supply so that plant community shows a mosaic pattern (23). As a result, the factors that influence plant growth and physiological activities, including climate, altitude and soil conditions etc. can affect directly and indirectly the grazing livestock behaviors. Soil spatial heterogeneity strongly influences the growth and physiology of individual plants (24, 25), the dynamics of plant populations (26) and interspecific interactions (25), and plant community composition (27). Alternatively, temporal and spatial variations in climatic conditions,

such as the inter- and intra-annual variability of precipitation and temperature have also led to significant shifts in the plant community characteristics (28, 29). However, up to date, our understanding about how livestock behaviors change across spatial and temporal scales and what factors are main drivers that shape these behavioral patterns of livestock remains very limited.

The Qinghai-Tibetan Plateau (QTP) is the largest grassland ecosystem in Eurasia, where the yak is the most important livestock grazing on the highlands (30). It is estimated that there are over 13.3 million domesticated yaks (*Bos grunniens*) that freely graze in this area (31, 32), ranging from the extensive grassland of the QTP to regions surrounding the Himalayan Mountains. Throughout the year, yaks are grazed on natural pastures of traditional ranches without the need for supplementary feeding (33). Therefore, the yak can adapt well to variations of plant community composition, biomass, and abiotic environments induced by seasons via an adjustment of their own grazing behaviors (34, 35). However, only a few studies have examined the behavioral patterns of grazing yaks of the QTP (36, 37). A recent study on the QTP has even shown that yak behaviors can vary throughout a day (38), but more field-based survey is needed to understand whether the yak behavior can vary at temporal and spatial scales and what factors are crucial to affect yak behaviors.

In this study, we hypothesized that (i) yak behavior would vary at both short temporal scales (hour, day and week) and at ranch scales with similar grazing intensity; and (ii) grazing behaviors of yak would be also strongly influenced by multiple factors including climate, season, altitude and forage characteristics, which would consequently affect intake of yak on the QTP. To test these hypotheses, first, three main yak behaviors including grazing, rumination and walking were continually monitored for 1 week at Maqu Research Station on the QTP; secondly, we selected four ranches nearby Maqu with similar grazing intensity to investigate spatial variations in grazing behavior; thirdly, we conducted a meta-analysis for yak behaviors based on a search of papers that were achieved in public database over the past 18 years, to examine the main factors that affect yak behaviors and intake. We aim to reveal the temporal and spatial variations of yak behaviors and figure out what factors are main drivers that shape observed behavior patterns of yak on the QTP.

2 Materials and methods

2.1 Study areas and animals

All trial procedures strictly followed rules and regulations of the Experimental Field Management protocols (files 2010–1 and 2010–2) of Lanzhou University and were approved by the Animal Ethics Committee of Lanzhou University.

Yak behaviors were monitored by visual observation during 12 July to 20 July near the peak of the plant growth at Maqu, Jiuzhi and Gande County on the eastern QTP of China (Figure 1). The three locations are typical areas where yaks are raised on the QTP, where the mean annual temperature (MAT) is 1.2°C and mean annual precipitation (MAP) is 620 mm. A total of five sites with yak grazing in the three locations were selected; one site (Maqu) was located in Maqu Country (101°52'E, 33°40'N) with an altitude of 3,536 m.a.s.l., three sites (JZ_SO, JZ_NBYZ and JZ_MT) were located in Jiuzhi Country (101°29'E, 33°25'N; 101°16'E, 33°26'N; 101°3'E, 33°46'N)

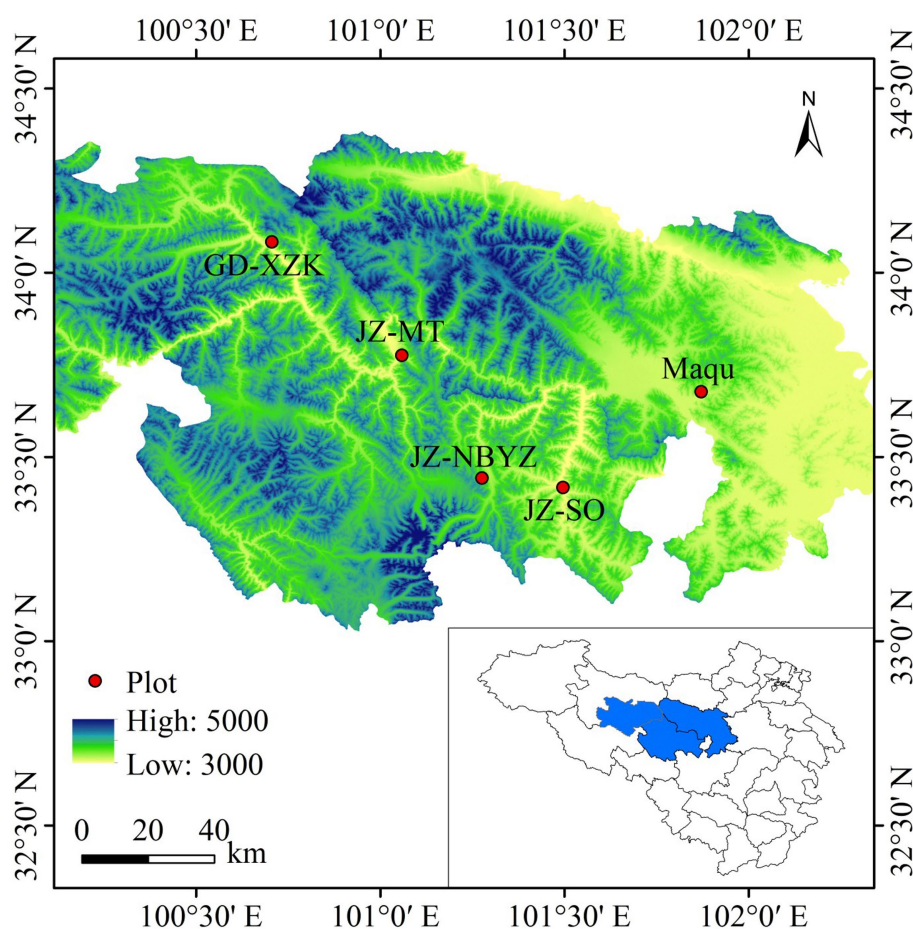


FIGURE 1

The location of five survey sites with showing geographical coordinates and altitude on the eastern Qinghai-Tibetan Plateau, China. The sites of Maqu and JZ_MT are located in Maqu and GD_XZK County, respectively. The sites of JZ_SO, JZ_NBYZ are located in Jiuzhi County. The graph was generated using ArcGIS (version 8.0).

with an altitude of 3,650 m, 3,888 m and 3,782 m a.s.l., and one site (GD_XZK) was located in Gande Country (100°42'E, 34°5'N) with an altitude of 3,853 m a.s.l. The main grassland type in these regions is typical alpine meadow, with *Kobresia capillifolia*, *Carex thibetica*, *Elymus nutans*, *Poa pratensis*, *Stipa aliena*, *Anemone rivularis* var. *flore-minore*, *Halenia corniculata*, and *Ligularia virgaurea* being dominant plant species. In each site, yaks were free to be grazed with a grazing intensity of 1.5–2 head of yak/ha from 9:00–17:00 every day, housed in shelters overnight and did not receive any supplementary feed during the experimental period.

2.2 Grazing behaviors

We carried out a continuous 6-day monitor of yak behaviors at Maqu. To minimum the effect of differences in individual characteristics on the grazing behaviors, yaks were selected based on their body weight (BW) and age across all survey sites in this study. Thus, a total of 18 yaks with BW of around 200 kg and an age of 3 years were selected as survey objects. Yak behaviors including the intake rate (bites per min), walking speed (steps per min) and rumination frequency per min were visually recorded

by independent observers who were divided into two groups (three independent observers within each group). All participants have rich experiences in monitoring yak behaviors and were trained well before the data were collected. Yak behaviors including intake rate (bites per min), walking speed (steps per min), rumination frequency (per min) were recorded every 10 min within each hour between 9:00 and 17:00 (9:00–10:00, 10:00–11:00, 11:00–12:00, 14:00–15:00, 15:00–16:00, and 16:00–17:00). Each behavior was calculated based on the average of obtained data from all observations of two groups. Bite rate was determined as the time that animals spent taking 60 bites. If the time between bites was longer than 15 s the measurement was canceled and started over (39). To examine ranch-scale variations in grazing behavior, we also carried out a similar observation of yak behaviors for the other four sites from 18 July to 20 July.

2.3 Meta analysis

A meta-analysis was conducted to assess the main factors that affect grazing behaviors and intake of yaks on the QTP. The related references published on the QTP were identified based on a search of

keywords including “grazing behavior” or “behavior” or “feed intake” or “intake” and “yak” during the past 18 years (2004–2021), which was recorded in the online database of WoS (Web of Science, <http://www.webofknowledge.com/>) and CNKI (China National Knowledge Infrastructure, <https://www.cnki.net/>). The geographic coordinates in each reference were uniformly converted by the online software.¹ The data sets of each graph in the reference were extracted by online software Web Plot Digitizer.² Ultimately, a total of 15 articles were collected in which 43 observations were used for subsequent analysis (40) (Supplementary Figure S3 and Supplementary Table S1), including altitude, climate (MAP, MAT), grazing behavior [intake, walking, resting and rumination (Rum)], season (cold and warm), forage characteristics [ash (ASH), crude protein (CP), neutral detergent fiber (NDF) and acid detergent fiber (ADF)], individual characteristics of livestock (age and weight) and grazing intensity (GI).

2.4 Statistical analysis

All statistical analysis and figure generation were performed in R (v.4.0.3). The bar charts and box plots were generated using the `geom_boxplot`, `geom_bar`, and `geom_smooth` functions of the `ggplot2` package. One-way analysis of variance (ANOVA) with *post hoc* tests were used to test the significant differences in the intake rate, rumination and walking among different times and survey sites. A quantile-quantile plot (Q–Q plot) was carried out to assess whether the residuals of each variable in the collected data sets are normal distributed by `qqPlot` function. We used Random Forest model to determine which variables in the data sets are the main predictors of the intake rate. Since the `randomForest` package of R statistical software does not provide significance of the prediction variable, the significance of each prediction variable to the response variable was assessed by using the “`rfPermute`” package. Structural Equation Modelling (SEM) was performed using IBM SPSS Amos 24.0 software to further evaluate the direct and indirect relationships between prediction variables selected by Random Forest model and the intake rate. Before doing SEM, all the data were standardized and carried out principal component analysis (PCA) for each module including climate, forage characteristics, grazing season, grazing behavior and individual characteristics of yak with `vegan` package. We only chose the first principal component (PC1) of each module as the variables in the SEM. The best fit of SEM was assessed by the chi-square test ($p < 0.001$) and RMSEA.

3 Results

3.1 Temporal and spatial variations of grazing behavior

The number of bites was observed and counted three times in total within each hour (10 min per time in all animals). We found

large variations in the intake rate of yaks within hours, among hours of each day and among days (Figures 2, 3; Supplementary Figure S1). The intake rate differed significantly within each hour of each day during 12 July to 17 July ($p < 0.05$) (Figure 2). Likewise, intake rate also differed significantly among hours of each day ($p < 0.001$) (Figure 3) and among days ($p < 0.001$) (Supplementary Figure S1). The intake rate generally reached maximum at 11:00–12:00 and decreased to minimum at 14:00–15:00 in each day (Figure 3).

The intake rate was significantly higher in the morning than in the afternoon in July 12 ($p < 0.001$), July 13 ($p < 0.001$), July 15 ($p < 0.001$), July 16 ($p = 0.008$) and July 17 ($p = 0.009$), but was marginally significantly higher in July 14 ($p = 0.073$) (Figure 4). Walking speed of yaks showed an inverse trend compared with the intake rate, with showing higher speed in the afternoon than in the morning (Figure 5; Supplementary Figure S2). Walking speed of yaks also differed significantly among hours of each day ($p < 0.001$) (Figure 5) and among days ($p < 0.001$) (Supplementary Figure S2).

Rumination of yaks generally occurred at 14:00–16:00 in the afternoon with a frequency of 0.4–1.8 per minute, but rumination time was observed to vary in each day from July 12 to July 17 (Figure 6). Based on an observation across five survey sites, the intake rate, walking speed and rumination of yaks showed spatial variations and differed significantly among sites ($p < 0.001$) (Figure 7 and Supplementary Figure S4). The intake rate of yaks in GD_XZK and JZ_NBYZ was significantly higher than in JZ_SO and Maqu ($p < 0.05$) (Figure 7). The intake rate of yaks in JZ_MT was only significantly higher than in JZ_SO ($p < 0.05$) (Figure 7).

3.2 Meta-analysis for grazing yaks on the QTP

Among the variables we obtained from online database, resting, altitude, MAT, ASH, MAP, livestock age and season were found to be the most important predictors for the intake of yaks ($p < 0.05$) (Figure 8), while other variables including weight, grazing intensity (GI), walking, ADF, CP, rumination (Rum) and NDF were not significant predictors for the intake rate of yaks ($p > 0.05$) (Figure 8), based on the analysis of Random Forest model. Structural equation model (SEM) showed that climate, altitude, forage characteristics, season, grazing behaviors, and livestock individual characteristics had direct or indirect effects on the intake (Figure 9). Altitude did not significantly affect the intake directly ($\lambda = -0.33$; $p > 0.05$), but significantly positively affected forage characteristics ($\lambda = 0.65$; $p < 0.001$) (Figure 9). Climate did not significantly affect the forage characteristics, grazing behaviors, livestock individual characteristics and the intake ($\lambda = -0.04, 0.13$ and 0.07 , respectively; $p > 0.05$), but had a significantly negative effect on season ($\lambda = -0.28$; $p < 0.05$) (Figure 9). Forage characteristics had a significantly positive effect on grazing behavior ($\lambda = 0.82$; $p < 0.001$) and intake of yaks ($\lambda = 0.50$; $p < 0.05$), but did not significantly affect yak individual characteristics ($\lambda = 0.06$; $p > 0.05$) (Figure 9). Livestock individual characteristics significantly negatively affected grazing behaviors ($\lambda = -0.19$; $p < 0.05$) and the intake of yaks ($\lambda = -0.31$; $p < 0.05$) (Figure 9). We also found that compared with forage characteristics and grazing behaviors, only altitude had a strongly negative effect on yak intake (Figure 9).

¹ <http://gzhatu.com/du2dfm.html>

² https://apps.automeris.io/wpd/index.zh_cn.html

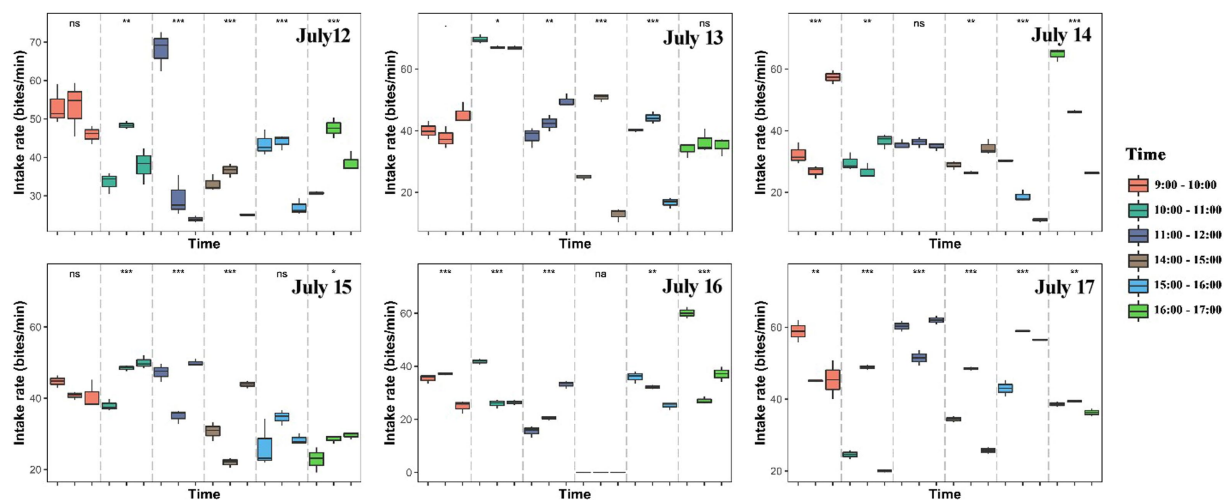


FIGURE 2

Differences in intake rate (bites/min) of yaks within hours (9:00–10:00, 10:00–11:00, 11:00–12:00, 14:00–15:00, 15:00–16:00, and 16:00–17:00) of each day. Asterisk indicates significant difference at * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

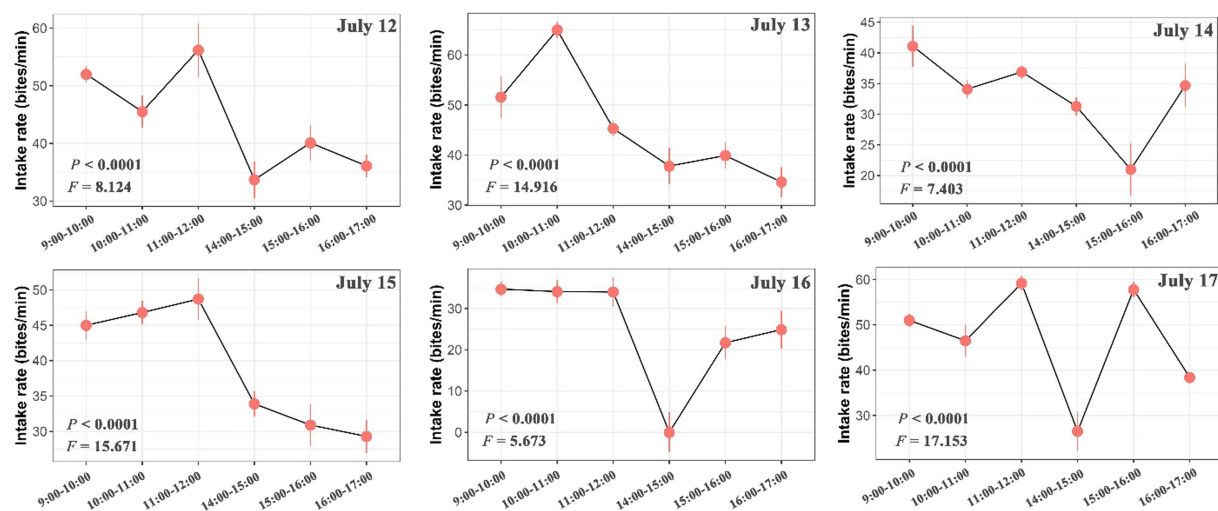


FIGURE 3

Differences in intake rate (bites/min) of yaks among hours of each day. The bars represent the standard errors. The statistically significance was tested by ANOVA at $p < 0.05$.

4 Discussion

Grassland systems play pivotal roles in ecosystem services such as livestock products and health through livestock grazing (41–43). However, global grasslands are undergoing serious degradation due to climate change and land use (44–48). The alpine grasslands on the QTP have been continuously used as pasturelands for millennia by herders for grazing of livestock (e.g., yak and Tibetan sheep) (49). Numerous evidence has reported that the QTP's grasslands have been overgrazed during past decades by rapidly increasing human and livestock populations (44). A direct consequence of grassland degradation is to lead to a decrease in pasture biomass and palatable grass

species and consequently has a great impact on grazing behaviors of livestock (50, 51). In grassland ecosystems, plant communities often show patchy distribution patterns over space due to soil spatial heterogeneity and increasing disturbance of human and animals (50) or variations of an environmental gradient (e.g., altitude). Also, plant community composition and diversity can vary temporally (52), depending greatly on the differences in soil and climatic factors induced by season or short-term variability in precipitation and temperature (53–56). In this case, grazing behaviors of livestock will thus probably change at spatial-temporal scales with the shift in plant communities. We did find in the study that grazing behaviors of yak varied at both temporal and ranch scales relying on a visual observation in the fields, and

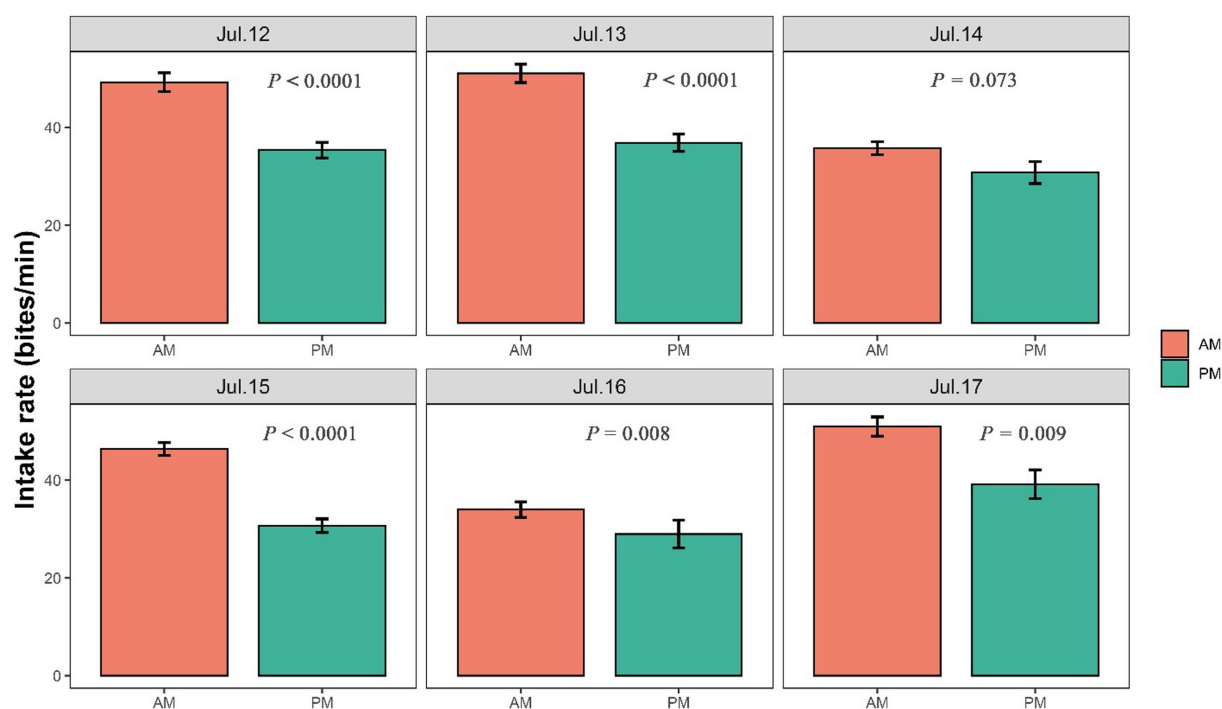


FIGURE 4

Differences in intake rate (bites/min) of yaks between in the morning and in the afternoon. The bar charts display standard errors. The statistically significance was tested by ANOVA at $p < 0.05$.

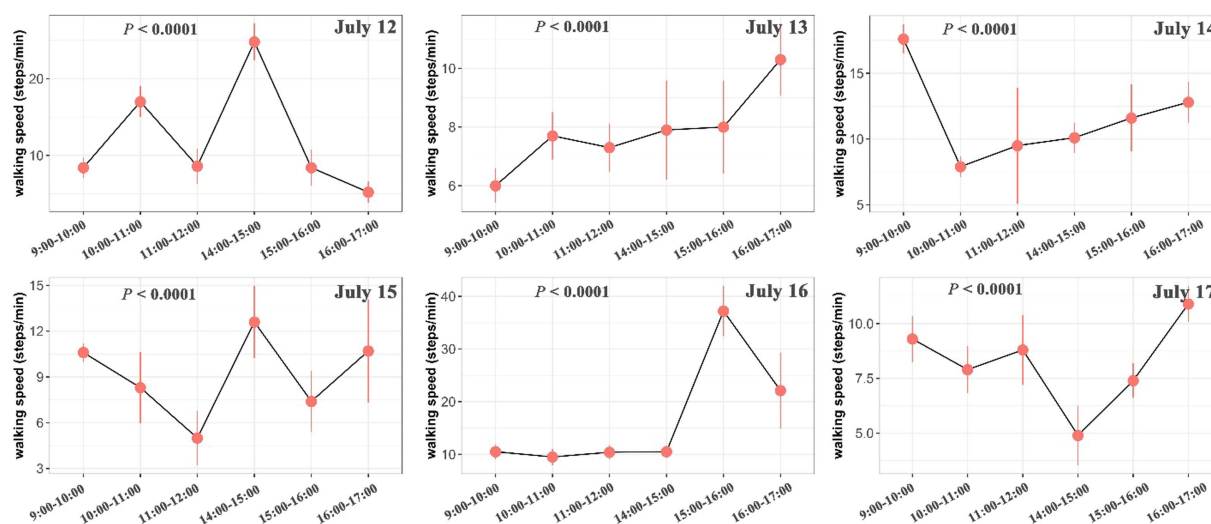


FIGURE 5

Differences in walking speed (step/min) of yaks among hours of each day. The bars represent the standard errors. The statistically significance was tested by ANOVA at $p < 0.05$.

importantly, intake of yaks was affected distinctly by multiple factors including climate, altitude, season, grazing behaviors, and forage and livestock individual characteristics based on a meta-analysis on the QTP. The findings verified our hypotheses about temporal and spatial variability in grazing behaviors of yaks and different factors that affecting yak intake at regional scales.

4.1 Temporal and spatial variations of grazing behavior

There were significant differences in grazing behaviors of yak including intake rates and walking speed within hours, among hours of each day and among days as well as across different observation

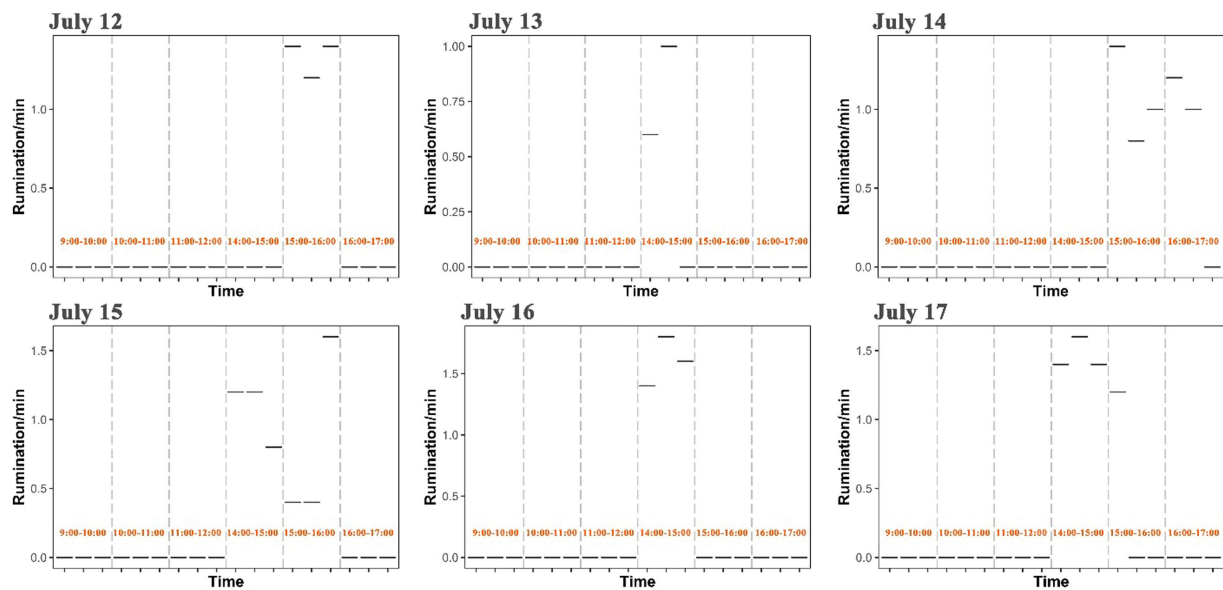


FIGURE 6
Distribution of rumination time of yaks across each day.

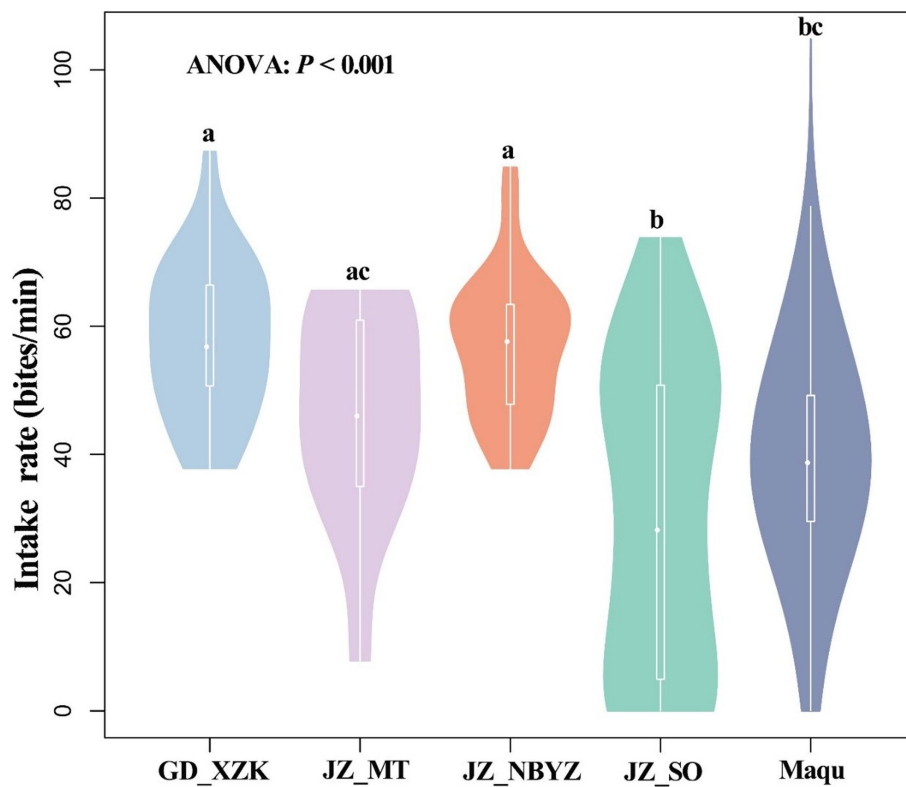


FIGURE 7
Violin diagram of the differences in feed intake (bites/min) of yaks across different observation sites of Qinghai-Tibetan Plateau. Lowercase indicates statistically significant differences at $p < 0.05$ by ANOVA.

sites. In this study, we only selected to carry out an observation for similar age of yaks, that is, the differences in individual characteristics such as the size and weight of yaks may be very small. Therefore, shifts

in yak behaviors may result mainly from the differences in composition and biomass of plant communities and the nutrients of pasture in the local ranch or among ranches (44, 57). It is well known

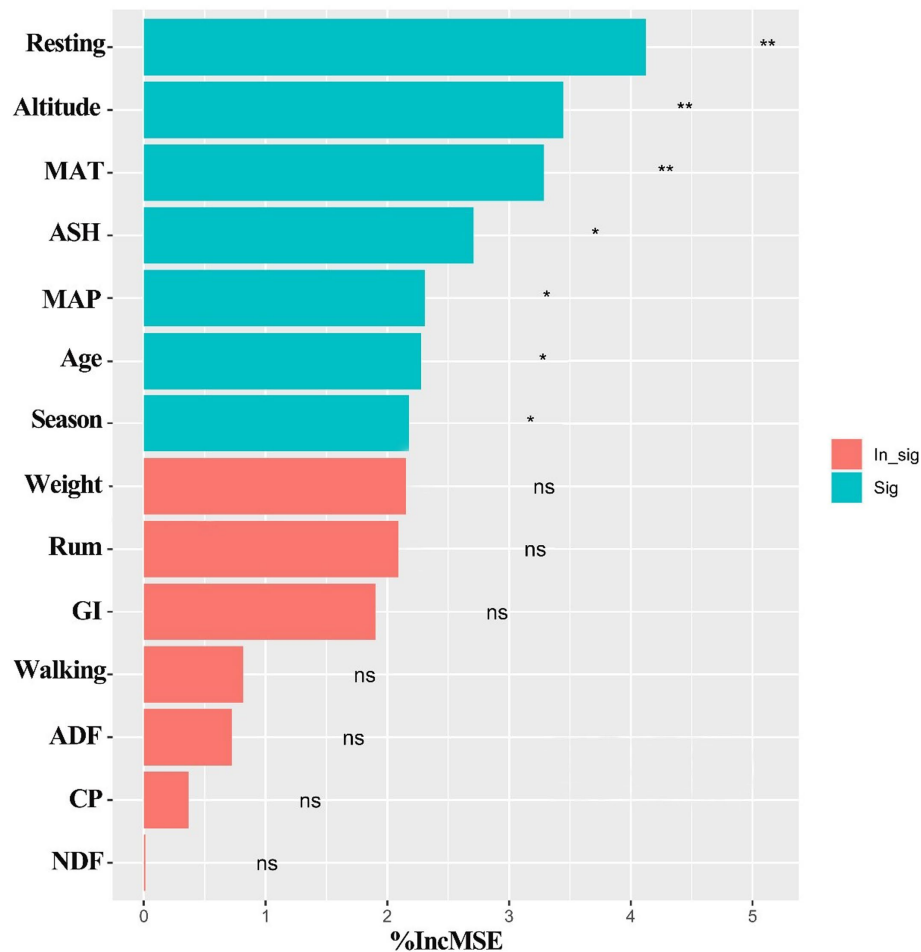


FIGURE 8

Random Forest model for evaluating significant factors affecting yak intake on the Qinghai-Tibetan Plateau, based on the data from public database during the past 18 years. Significance levels of each predictor are * $p < 0.05$ and ** $p < 0.01$. ASH, Ash; MAT, mean annual temperature; GI, grazing intensity; CP, crude protein; ADF, Acid detergent fiber; Rum, rumination; NDF, neutral detergent fiber; MAP, mean annual precipitation; Age, yak age; Season, livestock season; Weight, yak weight.

that plant communities are often found to have a mosaic distribution at different spatial scales (58). Even at a small local scale, the composition and biomass of palatable pasture may differ due to the effects of soil water content (54), grazing intensity (57, 59) and animal excrement (60). In fact, temporal variability in livestock behaviors has been reported in some previous studies. For example, several recent researches on the QTP have shown that grazing intensity may be the main factor affecting grazing behavior of yaks, and grazing behavior differed significantly between foraging time and distance traveled at light grazing (61, 62). Similarly, previous studies have reported that grazing season can affect heavily yak activities as well (63, 64). However, different from these reports, our findings provide evidence that even at a short temporal scale, such as within hours and among days, grazing behaviors of yaks also showed pronounced differences. In addition, we found that intake rate of yak was higher in the morning than in the afternoon. The finding might be related to the energy and metabolism of livestock. For instance, grazing yak generally grazing yaks generally need to acquire energy through daytime intake behavior after an overnight residence in the cattle sheds. The increased

intake rate in the morning can be also explained by the fact that the ruminal pool is usually at its smallest in the morning due to the body's natural process of expelling digesta (65, 66). The finding is consistent with prior reports in which showed that yak behaviors change with milking time and day/night cycles, which play major roles in the temporal distribution of intake rate (67, 68).

In contrast, we found a converse trend for walking speed in comparison to intake rate, that is, walking speed was relatively lower in the morning and higher in the afternoon. Such a discrepancy is reasonable because walking behavior generally increases energy consumption. Yaks need to save energy and allocate more time for foraging in the morning, but they commonly have higher rate of digestion through increasing walking and rumination in the afternoon. Actually, numerous studies have shown that forage acquisition is closely negatively correlated with walking speed in energy consumption of livestock (69–72). In addition to intake and walking behavior, rumination of yaks in the local ranch was mainly concentrated in the afternoon. The finding is in agreement with many previous studies showing that rumination activity in grazing yaks

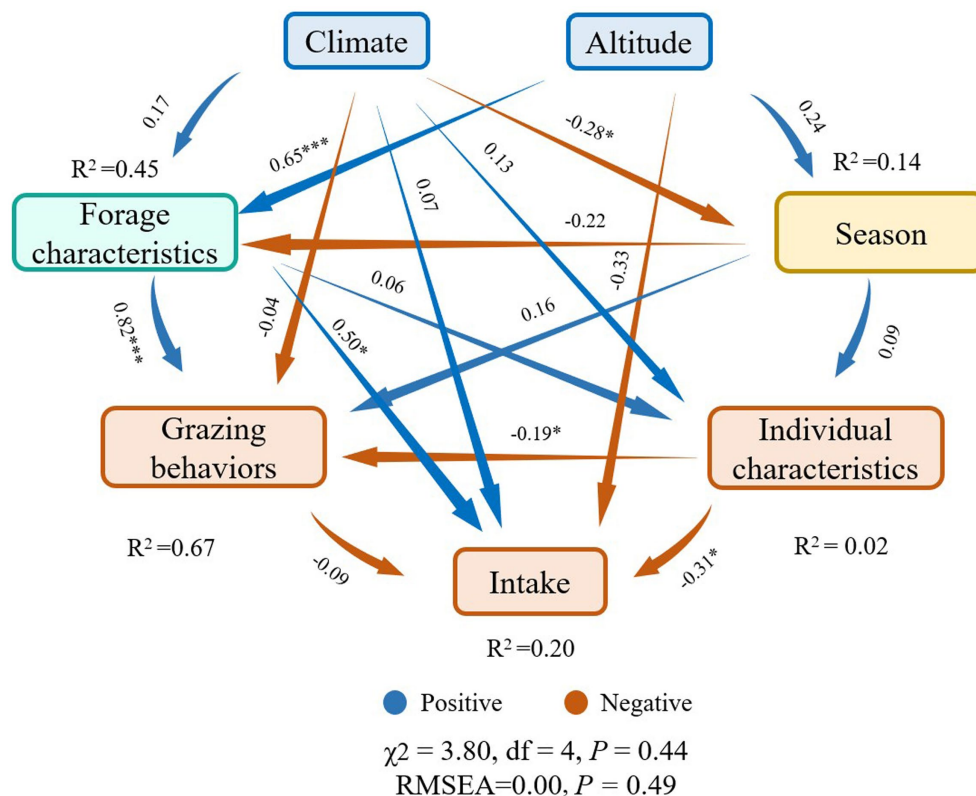


FIGURE 9

Structural equation model describing the effects of multiple factors including climate, altitude, season, forage characteristics, grazing behaviors and individual characteristics on yak intake. Numbers adjacent to arrows are indicative of the effect size of the relationship. Arrows reflect causality with blue (positive) and red (negative). R^2 denotes the proportion of variance explained. Significance levels of each predictor are $p < 0.10$, $*p < 0.05$, and $**p < 0.01$. In all cases, there was a non-significant deviation of the data from the model ($\chi^2 = 3.80$, $df = 4$; $p = 0.44$; RMSEA = 0.00; $p = 0.49$). $***p < 0.001$.

more frequently occurred in the afternoon (6, 33, 73, 74), probably because of the result of the high air temperature and solar radiation (7, 75).

4.2 The driving factors underlying the variations of yak intake based on meta-analysis

Despite observed temporal-spatial variability of grazing behaviors of yak, more focuses are needed to explore the consequence of varied grazing behaviors on foraging. For example, considering close associations between behaviors and feed intake of grazing livestock, we need clarify what factors affect livestock intake and which of them play direct or indirect roles in feed intake of livestock. As we can see from the Random Forest modeling base on a meta-analysis of the past 18-year dataset on the QTP, we indeed found that grazing behavior of yaks including resting, altitude, MAT, ASH, MAP, livestock age and season were the main predictors for yak intake in this area. The effects of these factors on yak intake have been reported in a recent study in which researchers found a close correlation between grazing behaviors of yaks and their intake (61). Some evidence has also shown that altitude would be an important predictor for grazing behaviors of yaks in summer if the pasture was utilized effectively (76). In addition, the

influences of season and climate (e.g., MAP and MAT) on feed intake of yaks is likely to be related to the changes in plant communities induced by changed hydrothermal conditions over space and time. For example, yaks on the QTP are usually grazed relying on seasonal migrations between summer and winter pasture, with migrating to higher altitude pastures in summer and then moving back to lower altitude winter pastures in order to maintain the energy consumption (74, 77–79). This can be further supported by the results from structural equation model, showing that forage characteristics were affected by altitude and resulted in direct effects on grazing behaviors and intake of yaks. Individual characteristics of yaks were also found to directly affect their grazing behaviors and intake. This suggests that the differences in yak age, size and weight may lead to totally different amounts of grasses they eat.

4.3 Implications

Our findings confirm that grazing behaviors of yaks on the QTP can vary even at both small temporal scales and regional scales. Temporal and spatial variations of yak behaviors are strongly affected by the environments including climate, altitude, season and forage quality and biomass. This implies that multiple factors can be responsible for the variations in livestock behaviors and shifts in

behavioral patterns may consequently lead to positive or negative feedback to QTP's grassland ecosystems through plant–animal interactions. Future work could focus on sustainable grassland management via modulating behaviors of grazing livestock on the QTP.

5 Conclusion

Altogether, we provide evidence that grazing behaviors of yaks on the QTP varied at small temporal and regional scales, and multiple factors involved in climate, altitude, season, forage and individual characteristics can have an effect on yak intake. Admittedly, some research limitations might exist in this study as only a visual observation was conducted for monitoring grazing behaviors of yaks, although a direct observation is generally reckoned as one of an effective way to investigate animal behaviors. It is necessary that more technologies such as developed wearable wireless sensors for continuously monitoring eating, rumination, laying and body temperature etc. in combination to visual observation should be widely utilized in the future studies, to provide more accurate parameters for modeling the relationships between grazing livestock and forages in rangeland ecosystems.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

Author contributions

XY: Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Writing – original draft. UD: Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing–original draft, Writing – review & editing. JM: Data curation, Investigation, Writing – review & editing. XL: Data curation, Investigation, Writing – review & editing. QF: Data curation, Investigation, Writing – review & editing. HZ: Data curation, Investigation, Writing – review & editing. X-BW: Conceptualization,

Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Software, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by the National Key Research and Development Program of China (2021YFD1300504), the General Program of the National Natural Science Foundation of China (32271716) and the Starting Research Fund for the Introduction of Talent of Lanzhou University (561120205).

Acknowledgments

We gratefully acknowledge Fujiang Hou for his support in field work in Maqu Research Station on the Qinghai-Tibetan Plateau.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2024.1393136/full#supplementary-material>

References

- Chen JQ, John R, Yuan J, Mack EA, Groisman P, Allington G, et al. Sustainability challenges for the social-environmental systems across the Asian Drylands Belt. *Environ Res Lett.* (2022) 17:023001. doi: 10.1088/1748-9326/ac472f
- Xin XP, Jin DY, Ge Y, Wang JH, Chen JQ, Qi JG, et al. Climate change dominated long-term soil carbon losses of inner Mongolian grasslands. *Glob Biogeochem Cycles.* (2020) 34, 1–16. doi: 10.1029/2020GB006559
- Cipriotti PA, Aguiar MR. Direct and indirect effects of grazing constrain shrub encroachment in semi-arid Patagonian steppes. *Appl Veg Sci.* (2012) 15:35–47. doi: 10.1111/j.1654-109X.2011.01138.x
- Hilario MC, Wrage-Monnig N, Isselstein J. Behavioral patterns of (co-)grazing cattle and sheep on swards differing in plant diversity. *Appl Anim Behav Sci.* (2017) 191:17–23. doi: 10.1016/j.applanim.2017.02.009
- Vandermeulen S, Ramirez-Restrepo CA, Marche C, Decruyenaere V, Beckers Y, Bindelle J. Behaviour and browse species selectivity of heifers grazing in a temperate silvopastoral system. *Agrofor Syst.* (2018) 92:705–16. doi: 10.1007/s10457-016-0041-x
- Liu PP, Ding LM, Zhou YQ, Jing XP, Degen AA. Behavioural characteristics of yaks grazing summer and winter pastures on the Qinghai-Tibetan plateau. *Appl Anim Behav Sci.* (2019) 218:104826. doi: 10.1016/j.applanim.2019.06.007
- Hejcmanova P, Stejskalova M, Pavlu V, Hejcman M. Behavioural patterns of heifers under intensive and extensive continuous grazing on species-rich pasture in the Czech Republic. *Appl Anim Behav Sci.* (2009) 117:137–43. doi: 10.1016/j.applanim.2009.01.003
- Kilgour RJ. In pursuit of “normal”: a review of the behaviour of cattle at pasture. *Appl Anim Behav Sci.* (2012) 138:1–11. doi: 10.1016/j.applanim.2011.12.002

9. Lin LJ, Dickhoefer U, Müller K, Wurina S, Susenbeth A. Grazing behavior of sheep at different stocking rates in the inner Mongolian steppe, China. *Appl Anim Behav Sci.* (2011) 129:36–42. doi: 10.1016/j.applanim.2010.11.002
10. Yang CT, Wang CM, Zhao YG, Chen TB, Aubry A, Gordon AW, et al. Updating maintenance energy requirement for the current sheep flocks and the associated effect of nutritional and animal factors. *Animal.* (2020) 14:295–302. doi: 10.1017/S1751731119002064
11. Hodgson J. *Grazing management. Science into practice.* Harlow, Essex, UK: Longman Group UK Ltd (1990).
12. Gregorini P, Tamminga S, Gunter SA. Behavior and daily grazing patterns of cattle. *Prof Anim Sci.* (2006) 22:201–9. doi: 10.15232/S1080-7446(15)31095-0
13. Gregorini P, DelaRue B, McLeod K, Clark CEF, Glassey CB, Jago J. Rumination behavior of grazing dairy cows in response to restricted time at pasture. *Livest Sci.* (2012) 146:95–8. doi: 10.1016/j.livsci.2012.02.020
14. Mangel M, Clark CW. Towards a unified foraging theory. *Ecology.* (1986) 67:1127–38. doi: 10.2307/1938669
15. Schirrmann K, Chapinal N, Weary DM, Heuwieser W, von Keyserlingk MAG. Rumination and its relationship to feeding and lying behavior in Holstein dairy cows. *J Dairy Sci.* (2012) 95:3212–7. doi: 10.3168/jds.2011-4741
16. Silva TPD, Marques CAT, Torreão JN, Bezerra LR, Araujo MJ, Lima LA. Effect of concentrate supplementation and time scales of evaluation on behavioral and physiological responses of pregnant ewes on grazing system. *Acta Scient Anim Sci.* (2016) 38:77–86. doi: 10.4025/actascianimsci.v38i1.28748
17. Chillo V, Ojeda R. Disentangling ecosystem responses to livestock grazing in drylands. *Agric Ecosyst Environ.* (2014) 197:271–7. doi: 10.1016/j.agee.2014.08.011
18. Rodrigues Silva R, Nunes do Prado I, Giordano Pinto de Carvalho G, Almeida de Santana Junior H, Ferreira da Silva F, Lucas Santos Dias D. Efeito da utilização de três intervalos de observações sobre a precisão dos resultados obtidos no estudo do comportamento ingestivo de vacas leiteiras em pastejo. *Ciência Anim Brasil.* (2008) 9:319–26. doi: 10.5216/cab.v9i2.1205
19. Lima SL, Zollner PA. Towards a behavioral ecology of ecological landscapes. *Trends Ecol Evol.* (1996) 11:131–5. doi: 10.1016/0169-5347(96)81094-9
20. Redbo I, Ehrlmark A, Redbo-Torstensson P. Behavioural responses to climatic demands of dairy heifers housed outdoors. *Can J Anim Sci.* (2001) 81:9–15. doi: 10.4141/A00-071
21. Evans SG, Pelster AJ, Leininger WC, Trlica MJ. Seasonal diet selection of cattle grazing a montane riparian community. *J Range Manag.* (2004) 57:539–45. doi: 10.2307/4003985
22. Pelster AJ, Evans S, Leininger WC, Trlica MJ, Clary WP. Steer diets in a montane riparian community. *J Range Manag.* (2004) 57:546–52. doi: 10.2307/4003986
23. Okayasu T, Okuro T, Jamsran U, Takeuchi K. An intrinsic mechanism for the co-existence of different survival strategies within mobile pastoralist communities. *Agric Syst.* (2010) 103:180–6. doi: 10.1016/j.agry.2009.12.006
24. Adomako MO, Xue W, Roiloa S, Zhang Q, Du DL, Yu FH. Earthworms modulate impacts of soil heterogeneity on plant growth at different spatial scales. *Front Plant Sci.* (2021) 12:735495. doi: 10.3389/fpls.2021.735495
25. Si C, Xue W, Guo ZW, Zhang JF, Hong MM, Wang YY, et al. Soil heterogeneity and earthworms independently promote growth of two bamboo species. *Ecol Indic.* (2021) 130:108068. doi: 10.1016/j.ecolind.2021.108068
26. Baer A, Wheeler JK, Pittermann J. Limited hydraulic adjustments drive the acclimation response of *Pteridium aquilinum* to variable light. *Ann Bot.* (2020) 125:691–700. doi: 10.1093/aob/mcaa006
27. Zhang XM, Cao XX, He LX, Xue W, Gao JQ, Lei NF, et al. Soil heterogeneity in the horizontal distribution of microplastics influences productivity and species composition of plant communities. *Front Plant Sci.* (2022) 13:1075007. doi: 10.3389/fpls.2022.1075007
28. Hopkins A, Del Prado A. Implications of climate change for grassland in Europe: impacts, adaptations and mitigation options: a review. *Grass Forage Sci.* (2007) 62:118–26. doi: 10.1111/j.1365-2494.2007.00575.x
29. Mundim FM, Bruna EM. Is there a temperate bias in our understanding of how climate change will alter plant-herbivore interactions? A meta-analysis of experimental studies. *Am Nat.* (2016) 188:S74–89. doi: 10.1086/687530
30. Joshi S, Shrestha L, Bisht N, Wu N, Ismail M, Dorji T, et al. Ethnic and cultural diversity amongst yak herding communities in the Asian highlands. *Sustain For.* (2020) 12:957. doi: 10.3390/su12030957
31. Yang CT, Gao P, Hou FJ, Yan TH, Chang SH, Chen XJ, et al. Relationship between chemical composition of native forage and nutrient digestibility by Tibetan sheep on the Qinghai-Tibetan plateau. *J Anim Sci.* (2018) 96:1140–9. doi: 10.1093/jas/sky002
32. Yang CT, Zhang Y, Hou FJ, Millner JP, Wang ZF, Chang SH. Grazing activity increases decomposition of yak dung and litter in an alpine meadow on the Qinghai-Tibet plateau. *Plant Soil.* (2019) 444:239–50. doi: 10.1007/s11104-019-04272-x
33. Ding LM, Long RJ, Yang YH, Xu SH, Wang CT. Behavioural responses by yaks in different physiological states (lactating, dry or replacement heifers), when grazing natural pasture in the spring (dry and germinating) season on the Qinghai-Tibetan plateau. *Appl Anim Behav Sci.* (2007) 108:239–50. doi: 10.1016/j.applanim.2006.12.004
34. Prendiville R, Lewis E, Pierce KM, Buckley F. Comparative grazing behavior of lactating Holstein-Friesian, Jersey, and Jersey x Holstein-Friesian dairy cows and its association with intake capacity and production efficiency. *J Dairy Sci.* (2010) 93:764–74. doi: 10.3168/jds.2009-2659
35. Vijayakumar M, Park JH, Ki KS, Lim DH, Kim SB, Park SM, et al. The effect of lactation number, stage, length, and milking frequency on milk yield in Korean Holstein dairy cows using automatic milking system. *Asian Australas J Anim Sci.* (2017) 30:1093–8. doi: 10.5713/ajas.16.0882
36. Jochims F, Soares EM, de Oliveira LB, Kuinchtner BC, Casanova PT, Marin L, et al. Timing and duration of observation periods of foraging behavior in natural grasslands. *Front Vet Sci.* (2020) 7:519698. doi: 10.3389/fvets.2020.519698
37. Sheahan AJ, Kolver ES, Roche JR. Genetic strain and diet effects on grazing behavior, pasture intake, and milk production. *J Dairy Sci.* (2011) 94:3583–91. doi: 10.3168/jds.2010-4089
38. Jing XP, Ding LM, Zhou JW, Huang XD, Degen A, Long RJ. The adaptive strategies of yaks to live in the Asian highlands. *Anim Nutr.* (2022) 9:249–58. doi: 10.1016/j.aninu.2022.02.002
39. Leaver J. D. (Ed.). (1982). *Herbage intake handbook (p. 143pp).* Wallingford, United Kingdom. 143.
40. Cleveland CC, Liptzin D. C:N:P stoichiometry in soil: is there a “Redfield ratio” for the microbial biomass? *Biogeochemistry.* (2007) 85:235–52. doi: 10.1007/s10533-007-9132-0
41. Isselstein J, Griffith BA, Pradel P, Venerus S. Effects of livestock breed and grazing intensity on biodiversity and production in grazing systems. 1. Nutritive value of herbage and livestock performance. *Grass Forage Sci.* (2007) 62:145–58. doi: 10.1111/j.1365-2494.2007.00571.x
42. Shang Z, Long R. Formation reason and recovering problem of the ‘black soil type’ degraded alpine grassland in Qinghai-Tibetan plateau. *Chinese Journal of Ecology.* (2005) 24:652–6.
43. Shang ZH, Gibb MJ, Leiber F, Ismail M, Ding LM, Guo XS, et al. The sustainable development of grassland-livestock systems on the Tibetan plateau: problems, strategies and prospects. *Rangel J.* (2014) 36:267–96. doi: 10.1071/RJ14008
44. Dong SK, Shang ZH, Gao JX, Boone RB. Enhancing sustainability of grassland ecosystems through ecological restoration and grazing management in an era of climate change on Qinghai-Tibetan plateau. *Agric Ecosyst Environ.* (2020) 287:106684. doi: 10.1016/j.agee.2019.106684
45. Li M, Zhang XZ, Wu JS, Ding QN, Niu B, He YT. Declining human activity intensity on alpine grasslands of the Tibetan plateau. *J Environ Manag.* (2021) 296:113198. doi: 10.1016/j.jenvman.2021.113198
46. Li XL, Gao J, Brierley G, Qiao YM, Zhang J, Yang YW. Rangeland degradation on the Qinghai-Tibet plateau: implications for rehabilitation. *Land Degrad Dev.* (2013) 24:72–80. doi: 10.1002/ldr.1108
47. Luo C, Wang S, Zhao L, Xu S, Xu B, Zhang Z, et al. Effects of land use and nitrogen fertilizer on ecosystem respiration in alpine meadow on the Tibetan plateau. *J Soils Sediments.* (2016) 17:1626–34. doi: 10.1007/s11368-016-1612-1
48. Zhang Y, Dong SK, Gao QZ, Liu SL, Zhou HK, Ganjurjav H, et al. Climate change and human activities altered the diversity and composition of soil microbial community in alpine grasslands of the Qinghai-Tibetan plateau. *Sci Total Environ.* (2016) 562:353–63. doi: 10.1016/j.scitotenv.2016.03.221
49. Yu HL, Ding QN, Meng BP, Lv YY, Liu C, Zhang XY, et al. The Relative Contributions of Climate and Grazing on the Dynamics of Grassland NPP and PUE on the Qinghai-Tibet Plateau. *Remote Sensing.* (2021) 13:3424. doi: 10.3390/rs13173424
50. Ma ZY, Liu HY, Mi ZR, Zhang ZH, Wang YH, Xu W, et al. Climate warming reduces the temporal stability of plant community biomass production. *Nat Commun.* (2017) 8:15378. doi: 10.1038/ncomms15378
51. Wu JS, Zhang XZ, Shen ZX, Shi PL, Xu XL, Li XJ. Grazing-exclusion effects on aboveground biomass and water-use efficiency of alpine grasslands on the northern Tibetan plateau. *Rangel Ecol Manag.* (2013) 66:454–61. doi: 10.2111/REM-D-12-00051.1
52. Niu Y, Yang S, Zhou J, Chu B, Ma S, Zhu H, et al. Vegetation distribution along mountain environmental gradient predicts shifts in plant community response to climate change in alpine meadow on the Tibetan plateau. *Sci Total Environ.* (2019) 650:505–14. doi: 10.1016/j.scitotenv.2018.08.390
53. Wang Y, Heberling G, Gorzen E, Miehe G, Seeber E, Wesche K. Combined effects of livestock grazing and abiotic environment on vegetation and soils of grasslands across Tibet. *Appl Veg Sci.* (2017) 20:327–39. doi: 10.1111/avsc.12312
54. Wu JS, Shen ZX, Shi PL, Zhou YT, Zhang XZ. Effects of grazing exclusion on plant functional group diversity alpine grasslands along a precipitation gradient on the northern Tibetan plateau. *Arct Antarct Alp Res.* (2014) 46:419–29. doi: 10.1657/1938-4246-46.2.419
55. Xu ZX, Gong TL, Li JY. Decadal trend of climate in the Tibetan plateau-regional temperature and precipitation. *Hydrol Process.* (2008) 22:3056–65. doi: 10.1002/hyp.6892
56. Yang J, Ding Y, Chen R. Spatial and temporal of variations of alpine vegetation cover in the source regions of the Yangtze and yellow Rivers of the Tibetan

plateau from 1982 to 2001. *Environ Geol.* (2006) 50:313–22. doi: 10.1007/s00254-006-0210-8

57. Zhou G, Zhou X, He Y, Shao J, Hu Z, Liu R, et al. Grazing intensity significantly affects belowground carbon and nitrogen cycling in grassland ecosystems: a meta-analysis. *Glob Chang Biol.* (2017) 23:1167–79. doi: 10.1111/gcb.13431

58. Xiao X, Zhang T, Angerer JP, Hou FJ. Grazing seasons and stocking rates affects the relationship between herbage traits of alpine meadow and grazing behaviors of Tibetan sheep in the Qinghai-Tibetan plateau. *Animals.* (2020) 10:488. doi: 10.3390/ani10030488

59. Feng XL, Tang JJ, Qiu HG. The effect of grassland transfer on herders' livestock production and grazing intensity in Inner Mongolia and Gansu, China. *China Agric Econ Rev.* (2022) 14:242–58. doi: 10.1108/CAER-09-2020-0221

60. Oudshoorn FW, Kristensen T, Nadimi ES. Dairy cow defecation and urination frequency and spatial distribution in relation to time-limited grazing. *Livest Sci.* (2008) 113:62–73. doi: 10.1016/j.livsci.2007.02.021

61. Hou LL, Xin XP, Shen BB, Qin Q, Altome AIA, Hamed YMZ, et al. Effects of long-term grazing on feed intake and digestibility of cattle in meadow steppe. *Agronomy.* (2023) 13:1760. doi: 10.3390/agronomy13071760

62. Zhuang XL, Kong WD. Grazing does not influence soil arbuscular mycorrhizal fungal diversity, but increases their interaction complexity with plants in dry grasslands on the Tibetan Plateau. *Ecol Indic.* (2023) 148:110065. doi: 10.1016/j.ecolind.2023.110065

63. Gusewell S, Pohl M, Gander A, Strehler C. Temporal changes in grazing intensity and herbage quality within a Swiss fen meadow. *Bot Helv.* (2007) 117:57–73. doi: 10.1007/s00035-007-0798-7

64. Linnane MI, Brereton AJ, Giller PS. Seasonal changes in circadian grazing patterns of Kerry cows (*Bos taurus*) in semi-feral conditions in Killarney National Park, co. Kerry, Ireland. *Appl Anim Behav Sci.* (2001) 71:277–92. doi: 10.1016/S0168-1591(00)00188-X

65. Fleming AE, Garrett K, Froehlich K, Beck MR, Mangwe MC, Bryant RH, et al. Rumen function and grazing behavior of early-lactation dairy cows supplemented with fodder beet. *J Dairy Sci.* (2021) 104:7696–710. doi: 10.3168/jds.2020-19324

66. Taweel HZ, Tas BM, Dijkstra J, Tamminga S. Intake regulation and grazing behavior of dairy cows under continuous stocking. *J Dairy Sci.* (2004) 87:3417–27. doi: 10.3168/jds.S0022-0302(04)73477-3

67. Brumby PJ. The grazing behaviour of dairy cattle in relation to milk production, live weight, and pasture intake. *N Z J Agric Res.* (2012) 2:797–807. doi: 10.1080/00288233.1959.10422839

68. Krysl LJ, Hess BW. Influence of supplementation on behavior of grazing cattle. *J Anim Sci.* (1993) 71:2546–55. doi: 10.2527/1993.7192546x

69. Berhan T, Puchala R, Goetsch AL, Merkel RC. Effects of walking speed and forage consumption on energy expenditure and heart rate by alpine does. *Small Rumin Res.* (2006) 63:119–24. doi: 10.1016/j.smallrumres.2005.02.012

70. Di Marco ON, Aello MS. Energy expenditure due to forage intake and walking of grazing cattle. *Arquivo Brasil Med Vet Zoot.* (2001) 53:105–10. doi: 10.1590/S0102-09352001000100017

71. Lachica M, Somlo R, Barroso FG, Boza J, Prieto C. Goats locomotion energy expenditure under range grazing conditions: seasonal variation. *J Range Manag.* (1999) 52:431–5. doi: 10.2307/4003768

72. Yang CT, Tsedan G, Fan QS, Wang SL, Wang ZF, Chang SH, et al. Behavioral patterns of yaks (*Bos grunniens*) grazing on alpine shrub meadows of the Qinghai-Tibetan plateau. *Appl Anim Behav Sci.* (2021) 234:105182. doi: 10.1016/j.applanim.2020.105182

73. Lindgren E. *Validation of rumination measurement equipment and the role of rumination in dairy cow time budgets [master's thesis]*. Uppsala, Sweden: Swedish University of Agricultural Sciences (2009).

74. Long RJ, Ding LM, Shang ZH, Guo XH. The yak grazing system on the Qinghai-Tibetan plateau and its status. *Rangel J.* (2008) 30:241–6. doi: 10.1071/RJ08012

75. Aublet JF, Festa-Bianchet M, Bergero D, Bassano B. Temperature constraints on foraging behaviour of male alpine ibex (*Capra ibex*) in summer. *Oecologia.* (2009) 159:237–47. doi: 10.1007/s00442-008-1198-4

76. Myrsetrud A, Iversen C, Austrheim G. Effects of density, season and weather on use of an altitudinal gradient by sheep. *Appl Anim Behav Sci.* (2007) 108:104–13. doi: 10.1016/j.applanim.2006.10.017

77. Baumont R, Prache S, Meuret M, Morand-Fehr P. How forage characteristics influence behaviour and intake in small ruminants: a review. *Livest Prod Sci.* (2000) 64:15–28. doi: 10.1016/S0301-6226(00)00172-X

78. Kharat M, Hassoun P, Aad PY, Hajj E, Bocquier F. Surveillance of baladi goat feeding behavior in response to a mediterranean natural rangeland evolution. *Rangel Ecol Manag.* (2020) 73:403–10. doi: 10.1016/j.rama.2020.01.005

79. Namgay K, Millar J, Black R, Samdup T. Transhumant agro-pastoralism in Bhutan: exploring contemporary practices and socio-cultural traditions. *Pastoralism: research. Policy Pract.* (2013) 3:13–26. doi: 10.1186/2041-7136-3-13



OPEN ACCESS

EDITED BY

Clara Mancini,
The Open University, United Kingdom

REVIEWED BY

Mark Hansen,
University of the West of England,
United Kingdom
Teddy Lazebnik,
University College London, United Kingdom

*CORRESPONDENCE

Nareed Farhat
✉ nhashe01@campus.haifa.ac.il

RECEIVED 01 March 2024

ACCEPTED 29 April 2024

PUBLISHED 14 June 2024

CITATION

Farhat N, van der Linden D, Zamansky A and
Assif T (2024) Automation in canine science:
enhancing human capabilities and
overcoming adoption barriers.
Front. Vet. Sci. 11:1394620.
doi: 10.3389/fvets.2024.1394620

COPYRIGHT

© 2024 Farhat, van der Linden, Zamansky and
Assif. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Automation in canine science: enhancing human capabilities and overcoming adoption barriers

Nareed Farhat^{1*}, Dirk van der Linden², Anna Zamansky¹ and
Tal Assif^{1,3}

¹Department of Information Systems, University of Haifa, Haifa, Israel, ²Department of Computer and
Information Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom, ³Lod Municipal
Shelter, Lod, Israel

The emerging field of canine science has been slow in adopting automated approaches for data analysis. However, with the dramatic increase in the volume and complexity of the collected behavioral data, this is now beginning to change. This paper aims to systematize the field of automation in canine science. We provide an examination of current automation processes and pipelines by providing a literature review of state-of-the-art studies applying automation in this field. In addition, via an empirical study with researchers in animal behavior, we explore their perceptions and attitudes toward automated approaches for better understanding barriers for a wider adoption of automation. The insights derived from this research could facilitate more effective and widespread utilization of automation within canine science, addressing current challenges and enhancing the analysis of increasingly complex and voluminous behavioral data. This could potentially revolutionize the field, allowing for more objective and quantifiable assessments of dog behavior, which would ultimately contribute to our understanding of dog-human interactions and canine welfare.

KEYWORDS

automation, canine science, artificial intelligence, animal behavior, motion tracking

1 Introduction

Dogs, *canis lupus familiaris*, are of increasing interest in many different disciplines, as can be witnessed by the increase in the scientific production on their cognitive and behavioral aspects (1). First of all, this interest can be attributed to the fact that dogs are useful clinical models for hundreds of human disorders. Indeed, they are large animal models, being physiologically and clinically more similar to human than other commonly used animal models, such as mice. Moreover, as companion animals also share the environmental conditions of their owners, similarly to humans they are affected by them. Numerous canine conditions are analogous to human diseases such as diabetes, cancers, epilepsies, eye diseases and autoimmune diseases, as well as rare diseases (2). Additional factors explaining dogs' popularity in science include fascination with the origins of dogs and their domestication, behavior, and cognition, as well as the need to better understand and regulate consequences of dog-human interactions and welfare, e.g., of working dogs and shelter dogs (1).

As a consequence of their living close to humans as pets, working or sheltered animals, dogs exhibit immense behavioral variability, stemming from their innate capacities as well as from environmental influence (3). Therefore, methods of *canine behavioral testing* are popular in research and practice. They are extensively used in cognitive science, veterinary science, working dog organizations, shelters for various purposes such as selection for breeding (4), learning abilities (5), prediction of suitability for work (6), adaptability in shelters (7, 8), animal models for human diseases (9), welfare (10).

Traditionally, data analysis in behavioral testing paradigms is done through direct and systematic human observation (11, 12) in a process where behaviors are defined in precise terms (usually they can have types of either event or state), deciding on the type of measurement, sampling method, etc. Properly trained human observers can typically provide accurate measures of almost any behavior. However, relying on human observation imposes severe limitations on behavioral data acquisition and analysis. As highlighted by Anderson and Perona (13), it is firstly a laborious and tedious task, limiting the volumes of processed data, as well as the number of analyzed behaviors or behavioral variables. But even more importantly, human analysis of behavior is prone to *subjectivity*, strongly depending on human perceptual abilities, leaving room for human error. Moreover, human understanding and interpretation of behavior is in itself subjective and sometimes inconsistent.

Advances in artificial intelligence (AI) open the door to new exciting opportunities to overcoming these limitations. Automated methods in human applied behavior research are already revolutionizing the field (14), as they can provide increased precision of measurement across smaller temporal units and for longer periods of time compared with human observers. These advancements have significant implications for understanding human behavior (15), mental health (16), and cognitive processes (17).

In the animal domain the need for promoting more objective and quantifiable assessment and measurement of behavior is also well-acknowledged [cf. (18–20)], pushing what is referred to “computational animal behavior analysis” (CABA) (13, 21, 22), also referred to as “behavioral imaging” (23). The release of deep learning frameworks such as DeepLabCut (24) has unleashed the potential of video-based motion tracking and pose recognition in many animal species. Additional tools such as EZtrack (25), LEAP (26), DeepPoseKit (27), idtracker.ai (28) provide more lightweight options, and support advanced settings such as large group tracking. An additional step in the field of CABA is taken by the paradigm shift from two-dimensional to three-dimensional data using multi-view cameras, enhancing our abilities to track every single point on the animal’s body and addressing full behavioral repertoires of various species. Large scale projects employ systems integrating multiple camera views to allow continuous 3D tracking of freely behaving animals, such as CAPTURE (29) and CaT-z (30) for rodents, and Open Monkey Studio (31) for rhesus macaques.

In canine science, however, the adoption of automation for data analysis has been quite slow. One reason for this is that the

generic deep learning platforms discussed above are not easily adaptable from controlled laboratory environments. For instance the JAABA system (32) which allows the user to annotate a small set of data to train a model specific to the study and is likely to perform poorly on dogs due to the diversity of morphologies and breeds. DeepLabCut (24) has been recently utilized for canine pose detection for emotion recognition (33) and recognition of stress related behaviors (34); both these studies noted the limitations of the system due to breed diversity.

The lack of tools tailored for the canine domain leads to the need for self-developed, domain-specific systems, which in its turn implies the necessity for multidisciplinary collaborations, leading to differences in terminology, research methods and expectations of different stakeholders. We explore in this study, this and other human-related adoption barriers for automation in the field.

The growing interest in automation in canine research, makes this a timely moment for a reflection and systematization of processes in this domain. What is usually presented in scientific papers applying automation in dog behavior analysis are just the end results, with the process of getting to them being left out of scope. Scientific papers focusing on the use of automation in analyzing dog behavior often only report the final outcomes, typically omitting the detailed process that led to these results. However, these details are essential to evaluate the insights gained and to explore future directions in this field.

This paper aims to promote automation in canine science by gaining a better understanding not only into its current usage, but also into *barriers* toward a wider adoption, and by scrutinizing not only artifacts of such analysis (i.e., study results), but also the *ways* to obtain them. To this end, we address the following research questions:

- How is automation currently used in canine science for dog behavior analysis?
- What are the challenges and barriers toward a wider adoption of automation in this context?

To answer the first question, we provide a comprehensive review of $N = 16$ studies that have applied some kind of automated analysis in the context of dog behavior. We dissect and categorize the reviewed works, identifying important dimensions, which represent the way automation is used today in this field. To address the second question, we perform an empirical study with $N = 24$ researchers who have experience with applying automation in the field, scrutinizing their perceptions and attitudes of the integration of automation and potential barriers toward their wider adoption.

The remainder of this paper is structured as follows: Section 2 encompasses a comprehensive review of studies that implement automated analysis in the realm of canine behavior. In Section 3, we present an empirical inquiry into the perceptions and attitudes of researchers in an effort to clarify the challenges impeding the broader adoption of automation in this field. The paper concludes with a discussion.

2 A mapping of automated approaches in dog behavioral data analysis

To promote systematization, as well as to provide an overview of the automated analysis methods that are relevant for animal behavior analysis outside of laboratory settings, we survey in this section studies that apply automated analysis in the form of machine learning techniques of some type.

2.1 Review protocol

2.1.1 Search strategy

We conducted the literature search using a snowballing approach (35) employing Google Scholar, as it has a broad reach covering most academic databases to avoid publisher bias, and because it includes relevant gray literature and pre-prints, which is particularly relevant to include in literature reviews of novel fields where new approaches are rapidly developed and published online.

We used the following query in Scholar to identify a first set of relevant core papers:

(animal OR dog OR canine AND (automated OR machine learning OR deep learning OR artificial intelligence OR ML OR AI) AND (behavior recognition OR behavior analysis))

The chosen keywords were selected to ensure a comprehensive and inclusive search. The terms “animal”, “dog”, and “canine” were used to focus on studies specifically related to dogs. The terms “automated”, “machine learning”, “deep learning”, “artificial intelligence”, “ML”, and “AI” were picked to capture all studies utilizing these computational methods in their research. Finally, “behavior recognition” and “behavior analysis” were included to focus on studies that apply these automated methods to behavior analysis or recognition, which is the core subject of our review.

We then identified a number of core papers, which we then conducted forward and backwards snowballing from, reading through both the literature cited and citing these works. We continued this approach by identifying relevant works using the selection criteria below, and for each identified relevant work, again snowballing by reviewing the cited and citing literature to identify further relevant literature to include.

We stopped the snowballing when we no longer identified any potentially relevant works to assess in the lists of works citing or cited by the selected papers.

2.1.2 Selection criteria

We liberally applied the following inclusion criteria while reading papers:

- **inc1** The paper applies an automated approach in dog behavior analysis;

and then refined the selection with these more concrete exclusion criteria:

- **exc1** The paper does not apply the automated approach in the context of a concrete question related to dog behavior, health,

welfare, cognition, etc. Thus, papers which solely design or propose new computational methods were excluded.

- **exc2** The approach is not applied on animal data (e.g., human behavioral data, or animal data obtained by human scoring).

2.1.3 Data extraction and dissection

We extracted the following data from the included papers (see [Supplementary Table 1](#)):

- What is the general domain/topic of the study, in the context of which the behavior is analyzed? The topics include dog-human relationships, health, welfare, etc.
- What are the concrete research questions explored in the study?
- How was the data obtained?
- How is behavior quantified, measured or computationally represented?
- What are the extracted features from the behavior representation?
- How are the extracted features used to answer the research questions?

The first author extracted the data from the included papers, which another author independently verified.

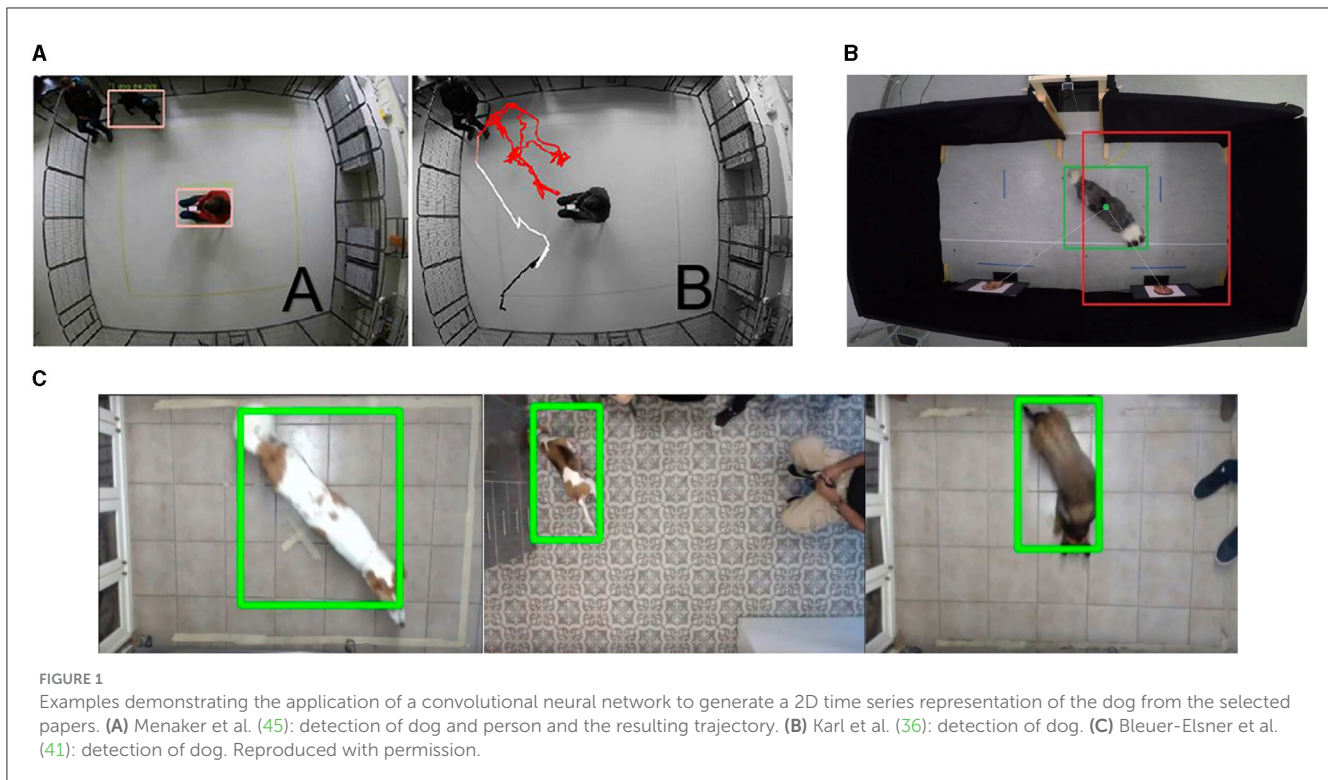
2.2 Findings

2.2.1 Selected papers

[Supplementary Table 1](#) presents a table of the 16 selected papers. Four of them address the topic of dog-human relationships (36–39), five of them focus on clinical behavioral aspects, of which three focus on ADHD-like behavior (40–42), one on separation anxiety (43), and one on ataxic gait patterns (44). Three of them focus on automation of scoring and assisting in behavioral testing (45–47), an additional three on emotion recognition (33, 48, 49), and one on shelter dog welfare (50).

2.2.2 Quantification of behavior

The idea behind automation of behavior analysis in these studies is in transforming obtained raw data into some computational representations of behavior that can then be manipulated either by machine learning or statistical methods. In most collected works, the data is obtained as visual data (videos or images) and needs to be manipulated. The most common representation of behavior used in the majority of these works is by tracking the dog's body (either in two or three dimensions). Bleuer-Elsner et al. (41), Karl et al. (36), Byosiére et al. (50), Fux et al. (40), Menaker et al. (51), Farhat et al. (46), Tsiourti et al. (47), and Watanangura et al. (42) use a convolutional neural network for object detection, producing a time series representing a trajectory of the dog from an above view, see examples in [Figure 1](#). Similarly, Völter et al. (37, 38) and Ren et al. (39) use a convolutional neural network on multiple cameras producing a 3D time series representation of multiple key points on the dogs' bodies. More subtle behavioral representations include Ferres et al. (33) which uses a convolutional neural network for



landmark detection, producing up to 24 points on the dogs' body in images, and Boneh-Shitrit et al. (48) which uses deep learning to extract both facial action units and deep learning learnt features from frontal images of dog faces. The remainder of works obtain acceleration and angular velocity measurements of movement from the accelerometer and gyroscope sensors on the dog's body. Aich et al. (49), Wang et al. (43), and Engelsman et al. (44) format the collected signals as 3D time series of the dog's body parts.

2.2.3 Extracted features

As highlighted in (52), a major focus in computer vision is on the use of *features* that can be extracted from images. Features can be hand-crafted, or manually designed, or learnt, as is the case in deep learning paradigms. Most of the studies reviewed here represent behavior in the form of time series or trajectories. Some of them extract from these trajectories some high-level meaningful features, such as average speed, residence in areas of interest, distance from and interaction with certain objects or people, studies that used time series of key points additionally extracted specific limb movement such as head angle, tail angle, velocity and amplitude. Two studies (46, 48) used deep learning to extract features automatically from the computational representation of the behavior.

2.2.4 Answering the research question: statistical testing vs. machine learning

The way in which (automatically) extracted features are used for addressing a biologically meaningful research question deserves special attention, as this point is not sufficiently studied in animal

behavior research. Traditionally in this research a hypothesis is formed, for which animal behavior is measured (either by manual coding, or in the more novel automated approaches discussed here). The measurements are then used for statistical testing of the hypothesis. However, the integration of the machine in data analysis means that we can use a different strategy. Machine learning classification is an alternative, powerful approach which is not discussed enough in the scientific community addressing animal behavior.

Li and Tong (53) discuss the differences between the two approaches of statistical testing vs. machine learning classification, which are rooted in two different cultures of inference vs. prediction. Inferential tasks aim to infer an unknown truth from observed data, and hypothesis testing is a specific framework for doing so. Prediction tasks, on the other hand, aim to predict an unobserved property of an instance based on the available (observed) features of that instance. Such prediction relies on building a prediction rule from the features to the unobservable property of interest, either based on human knowledge, or established from data.

2.2.5 The automation process: from quantification of behavior to answering research questions in behavior

We now present a conceptual model of the process of automation emerging from our analysis of the studies reviewed here. The steps involved in automated analysis in these works are presented in Figure 2. They are typically the following: (1) quantification of behavior, (2) extraction of features, and (3) using these features to answer some biological question related to

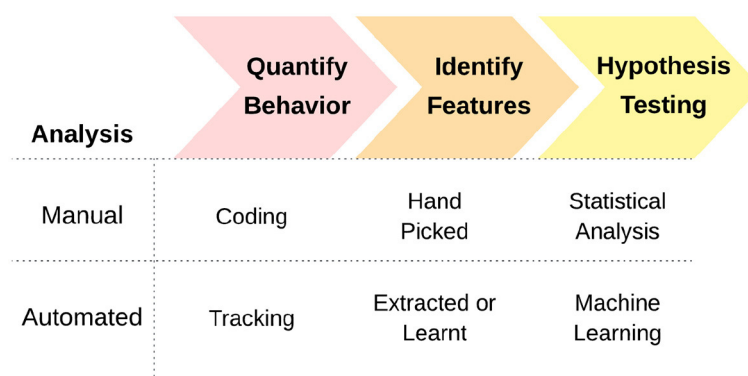


FIGURE 2
The proposed conceptual model of the process of automating behavior analysis.

behavior. Each of these steps can be done manually, or involving automation. Traditional ethological analysis keeps all of these stages manual: coding behavioral categories, quantification and then statistical analysis.

Below, we explore how the reviewed works make use of automation at each of these steps:

1. Bleuer-Elsner et al. (41) and Fux et al. (40) addressed the analysis of ADHD-like behavior in dogs. Dogs were recorded moving freely during a behavioral consultation visit in a veterinary clinic, with some diagnosed with ADHD-like behavior and others in a control group with no reported behavioral problems.

RQ: Are there differences in behavior in the consultation room between dogs diagnosed with ADHD-like behavior and control dogs?

Behavior Quantification: 2D time series (trajectory) of the dog was extracted.

Feature Extraction: The features extracted were hand-picked (traveled distance, average speed, straightness, intensity of use, etc.)

Answering the RQ: A machine learning model was developed to separate between the two types of dogs' behaviors.

2. Aich et al. (49) explored the feasibility of using wearable sensors for analyzing activity and emotional patterns of dogs. Data was collected using sensors placed on the necks and tails of the participants $N = 10$, while performing seven distinct activities and three emotional states (positive, neutral and negative). These sensors, equipped with accelerometers and gyroscopes, measured linear and rotational motions in all directions.

RQ: Can machine learning techniques recognize activity and emotional patterns of dogs from wearable sensors?

Behavior Quantification: 3D time series of the head and tail, received from the wearable sensors.

Feature Extraction: The extracted features were hand-picked (statistical and peak based features).

Answering the RQ: Two machine learning models were developed, the first detects the activity, the second detects the emotional state.

3. Karl et al. (36) investigated the engagement of an attachment-like system in dogs when they see human faces. Stimuli in the

form of videos of the caregiver, a familiar person, and a stranger were presented to the dog participants $N = 24$, showing either happy or angry facial expressions.

RQ: What are the neural, visual, and behavioral responses of dogs to human faces, and how do they differ between caregivers, strangers, and familiar persons?

Behavior Quantification: 2D time series (trajectory) of the dog was extracted.

Feature Extraction: The extracted features were hand-picked (time of residence in areas of interest, distance from screens, field of view, see Figure 3).

Answering the RQ: Statistical analysis was used to test the differences between conditions.

4. Byosiére et al. (50) examined the behavior of shelter dogs before COVID-19 pandemic restrictions and during the restrictions for a period of two weeks each. The participants $N = 34$ were recorded in their kennels for 15 seconds each hour in daytime.

RQ: Are there differences in shelter dogs activity levels before and during the COVID-19 pandemic?

Behavior quantification: 2D time series (trajectory) of the dog was extracted.

Feature extraction: The extract feature was hand-picked (step count; which is a quantification of the participants' movement).

Answering the RQ: Statistical analysis was used to test the differences between conditions and times of day.

5. Ferres et al. (33) addressed the different dog emotional states and detecting them based on posture analysis of images. Four emotional states were used (anger, fear, happiness, and relaxation)

RQ: Can machine learning techniques recognize emotional state of dogs from body language in images?

Behavior quantification: 24 key points on the dogs' body.

Feature extraction: The extracted features were hand-picked pose metrics (tail position, weight distribution, mouth condition etc.).

Answering the RQ: Machine learning models (neural network on landmarks, and a classification tree on the extracted features) were used to classify the different emotional states.

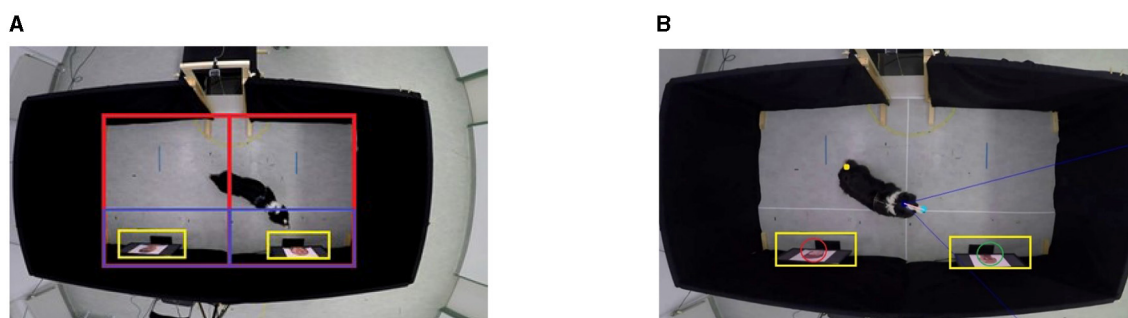


FIGURE 3

Examples of hand-picked features in Karl et al. (36). (A) Defined areas of interest that were used to calculate time of residence. (B) Example of participant field of view calculation. Reproduced with permission.

6. Völter et al. (37) aimed to investigate whether dogs are sensitive to the intentions underlying human actions using the unwilling-unable paradigm. The study involved two pre-registered experiments, the first, a within-subject design where the participants $N = 48$ observed a human actor who either intentionally (teasing) or unintentionally (clumsiness or blocked area) failed to provide a treat. In the second experiment, participants $N = 48$ observed two different human actors who performed either a clumsy or teasing demonstration of attempting to provide a treat.

RQ: Are dogs sensitive to the intentions underlying human actions?

Behavior quantification: 3D time series of four key points on the dog's body (snout, head center, base of tail, and tip of tail).

Feature extraction: The extracted features were hand-picked (visited areas, tail angle, visiting caregiver).

Answering the RQ: Statistical analysis was used to test the differences between the dog's reaction to the experimenters' intentions.

7. Menaker et al. (45) explored the automatic analysis of dogs' behavior in a behavioral test. Participating dogs $N = 30$ were recorded during a stranger test, which allowed the dog to move freely in a room with a stranger.

RQ: What is the potential use of unsupervised learning for pattern discovery dog behavior during a stranger test?

Behavior quantification: 2D time series (trajectory) of the dog and stranger were extracted.

Feature extraction: The chosen features were hand-picked (speed, covered movement area, approaching test person etc.).

Answering the RQ: Machine learning (k-means clustering) was used to identify dogs with similar behavior in the stranger test.

8. Boneh-Shitrit et al. (48) explored the recognition of dog emotional state, comparing the results of facial action units and deep learning techniques. Participating dogs $N = 29$ were recorded in a controlled experiment, inducing negative (frustration) and positive (anticipation) emotional states using treats.

RQ: Can machine learning techniques recognize emotional state of dogs from facial expression?

Behavior quantification: Facial action units

Feature extraction: The extracted features were automatically

learned from the video using deep learning.

Answering the RQ: Machine learning models (neural network on extracted features, tree based models on facial action units) were used to classify the different emotional states.

9. Ren et al. (39) investigated the effects of social cues on tail wagging during dog-human interactions. The participating dogs $N = 10$ were recorded over three consecutive days, interacting with the experimenter for 5 minutes (neutral postured experimenter, provided treats without direct contact).

RQ: How does the dog's tail wagging behavior during dog-human interactions manifest, and what are the underlying neural and behavioral mechanisms of this behavior?

Behavior quantification: 3D time series of four key points on the dogs' body (withers, back, croup, and tail tip).

Feature extraction: The extracted features were hand-picked (tail angle, amplitude, and velocity).

Answering the RQ: Statistical analysis was used to test the differences in tail wagging behavior across the test days.

10. Wang et al. (43) aimed to monitor and detect dog separation anxiety symptoms using wearable sensors. The sensors, placed on back and neck, were used to monitor home-alone cage-free $N = 8$ participants to identify behavior patterns that indicate separation anxiety symptoms.

RQ: Can machine learning detect and manage separation anxiety in dogs?

Behavior Quantification: 3D time series of the head and body, received from the wearable sensors.

Feature Extraction: The features were automatically extracted using machine learning models (head posture and body posture events).

Answering the RQ: Machine learning (Complex event processing and fuzzy logic) was used to classify the dogs' behavior pattern as normal or abnormal, the latter indicating a symptom of separation anxiety.

11. Engelsman et al. (44) investigated the use of smartphone sensors to measure ataxic gait patterns in dogs. The sensor was attached to the participating dog's back using a harness, and the dog was then walked on a leash at a steady pace 5 times. Which resulted to the capture of 770 walking sessions of $N = 55$ healthy dogs, and $N = 23$ dogs with ataxia.

RQ: What is the feasibility of using body-worn smartphone

sensors to automatically classify between dogs diagnosed with ataxia and a healthy control group?

Behavior quantification: 3D time series of the dog, using body worn sensor.

Feature extraction: The extracted features were hand-picked statistical and frequency features.

Answering the RQ: Machine learning was used to classify the gait patterns as either healthy or ataxic.

12. Völter et al. (38) examined the behavioral change of dogs N=37 exploring a room with new objects in the presence of the owner and/or stranger.

RQ: How does separation from their caregiver, and presence of a stranger, affect dogs' exploratory behavior in a novel environment?

Behavior quantification: 3D time series of eight key points on the dog's body (snout, head center, right ear, left ear, base of neck, hip region, tail base and tail tip).

Feature extraction: The extracted features were hand-picked (residence in areas of interest, traveled distance, distance from objects, field of view, and tail angle).

Answering the RQ: Statistical analysis was used to test the differences between the dog's behavior in presence and absence of owner.

13. Farhat et al. (46) investigated a computational approach to assess behavioral traits. The participated dogs N=53 were recorded in a behavioral test reacting to the presence of a strange, their coping styles were categorized into neutral, negative (reacting away from the stressor), and positive (reacting toward the stressor) reactions.

RQ: Can the machine learning techniques identify different behavioral profiles in an objective, human-free way in the stranger test?

Behavior quantification: 2D time series (trajectory) of the dog and stranger was extracted.

Feature extraction: The chosen features were automatically learned using machine learning (unsupervised clustering).

Answering the RQ: Machine learning models were used to classify dogs to the different behavior groups.

14. Tsiourti et al. (47)

RQ: Do dogs use logical reasoning or associative learning to solve an invisible displacement task?

Behavior quantification: 2D time series (trajectory) of the dog was extracted.

Feature extraction: The chosen features were hand-picked (pace, travel straightness, covered area).

Answering the RQ: Statistical analysis was used to test the differences between the dogs' travel based on placement type (visible/invisible) and object (toy/food).

15. Watanangura et al. (42) investigated the effects of fecal microbiota transplantation (FMT) on behavioral comorbidities in a dog model of epilepsy. Participants $N = 9$ underwent FMTs from a donor with controlled epilepsy three times, two weeks apart. Follow-ups were conducted at three and six months post-FMTs. Various evaluations were performed, including behavioral analysis and a range of biological tests.

RQ: Can FMT improve behavioral comorbidities and cognitive dysfunction in dogs with drug-resistant epilepsy?

Behavior quantification: 2D time series (trajectory) of the dog (and owner and/or stranger when relevant) was extracted.

Feature extraction: The extracted features were hand-picked (time spent in areas of interest, movement area, speed etc.).

Answering the RQ: Statistical analysis was used to test the FMTs effect overtime in the various follow-up tests (behavioral and biological).

2.3 Conclusions

The aim of this review was to assess the existing landscape of automation applications in canine science for dog behavior analysis. We have mapped the current situation with respect to how automation is currently utilized, analyzing 16 studies applying state-of-the-art automation in the context of dog behavior analysis. A conceptual model consisting of three steps arose from our analysis: (i) quantification of behavior, (ii) feature extraction, and (iii) hypothesis testing/answering the research question (either using statistical analysis or machine learning).

Overall, we observe a tendency to use straightforward and basic techniques at both stages (i) and (ii). In particular, a significant portion of the studies we reviewed utilized relatively simple object tracking methods for quantifying behavior [stage (i)]. However, there is a vast array of more sophisticated techniques in computer vision-like landmark localization (54), activity recognition (55), segmentation (56), and zero-shot learning (57) that hold great potential for this field. These more advanced approaches are yet to be fully explored and harnessed in the future.

Furthermore, most of the studies we reviewed tend to select a restricted range of features by hand (stage (ii)). Although these features are chosen based on expert knowledge in the field and can be quite informative, this method has its limitations. Specifically, the narrow scope of these hand-picked features might not provide the comprehensive data necessary to fully address the research questions at hand. Using alternative approaches in machine learning, such as deep learning, automated feature selection and unsupervised learning may be beneficial in this domain [see, e.g., the approaches used in (46, 48)] and should be further explored.

It is also notable that testing the hypothesis [stage (iii)] is most commonly addressed by statistical methods, following traditional approaches in behavior research (58). However, some recent studies have applied machine learning techniques.

2.4 Threats to validity

In conducting this literature review, several potential threats to validity were considered, acknowledging the need for a critical assessment of the findings.

- *Incomplete retrieval of studies:* despite thorough search strategies, there is a risk of incomplete retrieval. Cross-referencing citations and consulting multiple databases were employed to minimize this threat.

- *Inappropriate or incomplete search terms in automatic search:* the procedure of determining the search query was based on iterative refinement using identified papers as example inputs.
- *Review process:* this review is not a systematic review, and as such, the inclusion of studies may not follow a strict protocol. While efforts were made to conduct a comprehensive and thorough review, the absence of a formal systematic approach introduces a limitation in terms of standardized study selection.

3 Barriers to adoption of automation: an empirical study

To investigate the perceptions of animal researchers toward automated analysis tools, an exploratory study (59) was designed. This group consisted of researchers working on both fundamental and applied animal behavior research, excluding those in laboratory settings, and having at least some experience with dog behavior. The study adopted a mixed-method approach, incorporating both qualitative and quantitative approaches, as detailed in the “Procedure” subsection. The experimental design was reviewed and approved by the Ethical Committee of the University of Haifa. Informed consent was obtained from all participants. All relevant institutional guidelines and codes of conduct for ethical research were followed.

3.1 Participants

Twenty-four animal researchers were recruited via an invitation to participate in our study posted on an active international Facebook group on animal-centered computing, and by using the authors’ own networks. To ensure privacy, we did not collect any personal information that could be used to identify any demographic subgroups. All participants volunteered to take part in the study and did not receive any compensation. Table 1 provides an overview of the participants’ backgrounds, the species they currently investigate, and their years of experience in their respective fields.

3.2 Procedure

The data collection instrument was designed as a questionnaire incorporating both Likert-like scale and open-ended questions, designed to gather comprehensive insights into participants’ experiences and perceptions. The formulation of questions drew from the authors’ prior involvement in collaborations with animal researcher (36, 46, 50). Furthermore, the questionnaire underwent a pilot phase, involving a behavioral veterinarian who had experience with an automated tool, and an animal behavior researcher who had not previously utilized an automated analysis tool. The valuable feedback provided by both participants contributed to the refinement and improvement of the questionnaire’s content.

TABLE 1 Overview of the backgrounds of the animal researchers, the species they investigate, and their years of experience in their respective fields.

ID	Background	Species currently investigating	Years of experience
P1	Cognitive science.	Laboratory animals.	5–10 years
P2	Animal behavior.	Companion animals.	10+
P3	Veterinary science.	Companion animals.	10+
P4	Veterinary science.	Companion animals.	0–2 years
P5	Veterinary science.	Companion animals.	10+
P6	Animal computer interaction.	Various.	5–10 years
P7	Animal behavior.	Companion animals.	10+
P8	Animal behavior.	Companion animals.	10+
P9	Computer science, neuroscience.	Dogs, horses, zoo animals.	10+
P10	Animal behavior.	Companion animals.	5–10 years
P11	Computing and design.	Various.	5–10 years
P12	Founder, dog trainer.	Companion animals.	5–10 years
P13	Veterinary science.	Companion animals.	0–2 years
P14	Computer science, robotics.	Companion animals.	3–5 years
P15	Veterinary science.	Companion animals.	5–10 years
P16	Neuroscience.	Laboratory animals.	3–5 years
P17	Cognitive science.	Companion animals.	0–2 years
P18	Animal behavior.	Companion animals.	10+
P19	Animal behavior.	Companion animals.	5–10 years
P20	Veterinary science.	Companion animals.	0–2 years
P21	Veterinary science.	Companion animals.	3–5 years
P22	Animal behavior.	Dairy calves.	0–2 years
P23	Animal behavior.	Companion animals.	10+
P24	Animal behavior.	Chickens.	3–5 years

We structured the questionnaire into multiple sections to examine both current experience and future perspective on automation in the context of animal research. Concretely, we elicited the following data:

- *Section 1: Background information.* Inquires about participants’ background, the animal species they study, their years of experience in animal behavior, and their previous experience using automated tools for data analysis.
- *Section 2: Experience with automated tools.* Asks participants about their experience with automated tools for data analysis, the reasons for choosing automated tools, their satisfaction with the results achieved, challenges faced while collaborating with data scientists, and how the outcome compared to their initial expectations.

- **Section 3: Issues and challenges.** Explores participants' confidence in understanding the precision of the automated analysis tool, their trust in the outcomes produced by the tool, steps data scientists can take to help participants understand and trust the results, difficulties in communication with data scientists, preprocessing required for data preparation, benefits of automated analysis for research, and preference for non-automated analysis methods.
- **Section 4: Reflection on the future of automation in animal behavior research.** Asks participants about barriers to wider adoption of automated analysis tools, suggestions to overcome these barriers, likelihood of using automated tools in future research, and use of other technologies to assist in previous projects.

The full questionnaire can be found in [Supplementary Data Sheet 1](#).

3.3 Data analysis

We obtained 24 responses to the questionnaire, which were analyzed using thematic analysis (60). All authors individually familiarized themselves with the data. Following on from this, multiple collaborative sessions were held where all authors discussed and framed interpretations of the data to propose and refine key themes focused on the challenges of adopting automation in animal behavioral data analysis, with a particular focus on dogs.

The first author coded the qualitative data obtained in the questionnaire in two iterations, identifying repeated meaningful patterns and using this to construct an initial classification framework for the data. This framework was then applied to the data in a similar iterative fashion and discussed with all authors. In total, we constructed three core themes that encapsulated the different codes.

3.4 Findings

3.4.1 Descriptive analysis

When queried about their prior experience with automated data analysis tools, 11 respondents (46%) affirmed previous use, while the remaining 13 (54%) indicated otherwise. [Figure 4A](#) provides an overview of the satisfaction levels reported by participants who have employed these automation tools, reflecting on their achieved results. [Figure 4B](#) presents the participants' anticipated likelihood of incorporating automated data analysis tools into their future work.

3.4.2 Thematic analysis

The thematic analysis of the qualitative data yielded two prominent themes, which we explore next.

Automation is useful for overcoming human limitations

One important motivation for opting to work with automated tools for analysis of behavioral data, mentioned by some participants, was **the need to increase accuracy and objectivity of analysis** [*"My interest was to find an objective measurement tool"*

(P5); *"I was looking for objectivity"* (P8); *"It is the most accurate way I have heard of for tracking multiple animals."* (P24)].

Another reason was **the necessity to work with large volumes of data and reducing time for manual labor** [*"I needed to do a huge amount of tracking behavior in limited time"* (P7).; *"It removed hundreds of hours of work and was reapplied to nearly a dozen experiments."* (P16)].

Perhaps the most interesting reason for using automated analysis as revealed by several participants, was **the need for an analysis that could not (or are extremely hard to) be performed manually** [*"Without the use of automated analysis, we would not have been able to calculate the main behavioral parameters of interest for our experiment."* (P14)]. Specific examples of such parameters are:

- Speed of movement [*"We wanted to analyze behavioral measurements (specifically speed of movement of a dog) that were very hard to calculate 'manually' though video observation. Therefore we resorted to the use of an automated video analysis software."* (P14)].
- Length of trajectory [*"We wanted to code travel length of free-moving dogs in an open field, which would have been very difficult manually."* (P10)].
- Approximating a dog's visual field [*"We needed a sophisticated analysis, e.g. dogs visual field."* (P23)].

However, human limitations also challenge the adoption of automation

The majority of the barriers challenging the wider adoption of automation were discovered to be human-related. Among the barriers that have been mentioned by participants for a wider adoption of automation in the field of animal behavior analysis, particularly dominant were **lack of awareness to existing tools and a steep learning curve**. The **complexity of computational tools** is another issue, which is closely related to the latter [*"... tools should be simple and minimal and not try to cover every variation. They can become a lot more complicated to learn and there are more issues that could come up which we may not realize when building it."* (P16)].

Suggestions to overcome these barriers included enhancing education on computational topics; some interesting suggestions were:

- *"Courses organized for universities/post grad schools/animal behavior departments"* (P15),
- *"Animal Behavior courses should teach programming in a fun way! Make it less daunting to learn. Also easier no-code tools are now possible, esp with LLMs"* (P17).
- *"Make the technology more accessible and more outreach/demos on how they work."* (P21).

An important factor mentioned by our participants was also **communication gaps** between people from different disciplines, which can be overcome by having more discussions about the tools and the implicit assumptions made about them:

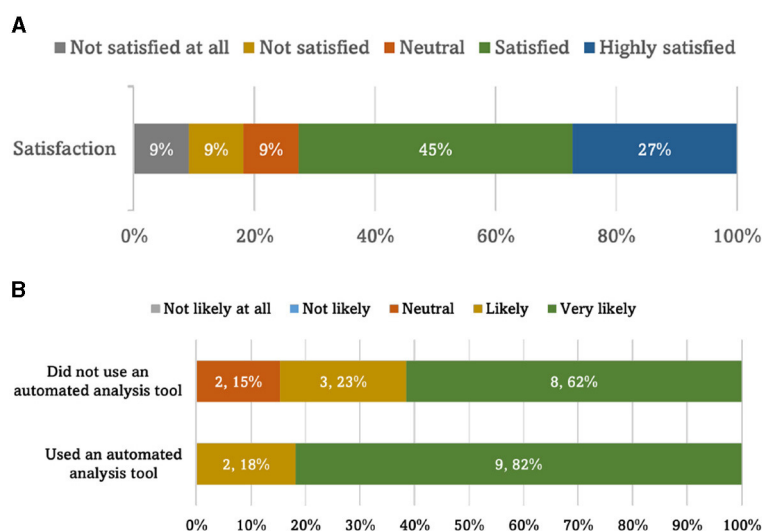


FIGURE 4

(A) Participants' satisfaction with the achieved results from an automated analysis tool. (B) The likelihood of participants using an automation data analysis tool in future work divided by their previous use of an automated analysis tool.

- “There is often a communication gap between the authors of the tool and researchers, making the tools harder to adopt and leading to potential problems where the researchers don’t know what went wrong and may think the entire tool doesn’t work or they aren’t qualified.” (P10).
- “Being aware about different hidden assumptions of the collaborators and talking about those explicitly is helpful.” (P14).

The **availability of adequate funding** is another obvious issue contributing to the limited adoption of automation in the field of animal behavior. [“... budget not obtained for this research.” (P1), “Price of software capable of this ...” (P21), “... cost or lack of university subscription.” (P22)].

3.5 Threats to validity

This section addresses potential threats to the validity of the exploratory study. The examination of these threats aims to provide a transparent acknowledgment of limitations associated with the design, execution, and findings of this section.

- **Questionnaire Formulation:** The formulation of questionnaire questions by the same author introduces a potential bias in the inquiry process. Despite piloting for understandability, the selective framing of questions may have overlooked relevant obstacles, potentially affecting the comprehensiveness of participant responses.
- **Selective Inquiry:** Deliberate crafting of questions focusing on challenges in utilizing automated analysis tools might inadvertently neglect other pertinent obstacles. This selective inquiry may introduce response bias, limiting our ability to capture a comprehensive view of participants’ experiences.

- **Participant Pool and Recruitment:** The generalization of results is constrained by the relatively small number of participants, and the recruitment from social circles and collaborative networks. A larger and more diverse sample is essential for enhancing the validity of the findings and ensuring broader generalizability.

4 Discussion

The aim of this work was to advance the application of automated techniques in the field of canine science. We have mapped the current situation with respect to how automation is currently utilized, by analyzing 16 studies applying state-of-the-art automation in the context of dog behavior analysis. A conceptual model consisting of three steps arose from our analysis: (i) quantification of behavior, (ii) feature extraction, and (iii) hypothesis testing/answering the research question (either using statistical analysis or machine learning). The outcomes of these analyses underscored the necessity of understanding the perceptions and attitudes of animal researchers, particularly in identifying barriers hindering the broader adoption of automation in the field. This realization prompted the initiation of the empirical study. Together, these parts aid in a broader understanding of the current and future use of automation in the field of canine science.

Overall, our analysis reveals a tendency to employ straightforward and basic techniques in both the behavior quantification and feature extraction stages [stages (i) and (ii)] within the proposed conceptual model. This underscores the need for more accessible introductory materials tailored for canine science researchers, providing insights into new methods for both behavior quantification and feature extraction. Additionally, our findings indicate that the testing of hypotheses [stage (iii)] is predominantly addressed through statistical

methods. This observation highlights the current reliance on traditional statistical approaches in the canine science research community. However, some recent studies have applied machine learning techniques. By making this distinction explicit, we aspire to spark productive discussions within the canine science, and the more general animal behavior community about the methodological issues related to these complementary data analysis approaches. We also hope this will encourage a greater openness to these alternative approaches among community members.

The above discussion highlights the predominance of basic and straightforward computational methods in current applications, that need to evolve further. However, feedback from the participants in our empirical study points to current challenges in mastering even these computational tools. Many participants expressed concerns over steep learning curves and a general lack of understanding regarding the operation and underlying assumptions of these tools. The relatively small number of participants, as well as using our network to reach them should be acknowledged as a limitation of this study, and future research should include a more representative sample of relevant researchers.

We believe that in today's rapidly evolving research landscape, AI proficiency is becoming increasingly vital for the field of animal behavior. Without this expertise, researchers might struggle to effectively apply AI in their research, potentially missing out on critical findings or misinterpreting AI-generated data. Moreover, a lack of proficiency could lead to a knowledge gap, hindering collaboration with other disciplines and slowing the progress of multidisciplinary research. Therefore, the development of AI education tailored to the needs of this community, as well as fostering multidisciplinary collaboration and dialogue between various communities seem to be the right way forward.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

Ethics statement

The animal studies were approved by University of Haifa has reviewed the study and waived approval. The studies were

conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

NF: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. AZ: Conceptualization, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. DV: Conceptualization, Formal analysis, Validation, Writing – original draft, Writing – review & editing. TA: Investigation, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2024.1394620/full#supplementary-material>

References

- MacLean EL, Fine A, Herzog H, Strauss E, Cobb ML. The new era of canine science: reshaping our relationships with dogs. *Front Vet Sci.* (2021) 762:675782. doi: 10.3389/fvets.2021.675782
- Hytönen MK, Lohi H. Canine models of human rare disorders. *Rare Dis.* (2016) 4:e1006037. doi: 10.1080/21675511.2016.1241362
- Diederich C, Giffroy JM. Behavioural testing in dogs: a review of methodology in search for standardisation. *Appl Anim Behav Sci.* (2006) 97:51–72. doi: 10.1016/j.applanim.2005.11.018
- Wilsson E, Sundgren PE. The use of a behaviour test for the selection of dogs for service and breeding. I: Method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex and breed differences. *Appl Anim Behav Sci.* (1997) 53:279–95. doi: 10.1016/S0168-1591(96)01174-4
- Pongrácz P, Miklósi Á, Vida V, Csányi V. The pet dogs ability for learning from a human demonstrator in a detour task is independent from the breed and age. *Appl Animal Behav Sci.* (2005) 90:309–23. doi: 10.1016/j.applanim.2004.08.004

6. Brady K, Cracknell N, Zulch H, Mills DS. A systematic review of the reliability and validity of behavioural tests used to assess behavioural characteristics important in working dogs. *Front Vet Sci.* (2018) 5:103. doi: 10.3389/fvets.2018.00103
7. Dowling-Guyer S, Marder A, D'arpino S. Behavioral traits detected in shelter dogs by a behavior evaluation. *Appl Anim Behav Sci.* (2011) 130:107–14. doi: 10.1016/j.applanim.2010.12.004
8. Ledger R, Baxter M. The development of a validated test to assess the temperament of dogs in a rescue shelter. *UFAW J.* (1997) 87–92.
9. Head E, Milgram N. Changes in spontaneous behavior in the dog following oral administration of L-deprenyl. *Pharmacol Biochem Behav.* (1992) 43:749–57. doi: 10.1016/0091-3057(92)90404-4
10. Beerda B, Schilder MB, van Hooff JA, de Vries HW. Manifestations of chronic and acute stress in dogs. *Appl Anim Behav Sci.* (1997) 52:307–19. doi: 10.1016/S0168-1591(96)01131-8
11. Bailey JS, Burch MR. *Research Methods in Applied Behavior Analysis.* New York, NY: Routledge (2017). doi: 10.4324/9781315543369
12. Martin P, Bateson P. *Measuring Behavior: An Introductory Guide.* Cambridge: Cambridge University Press. (2007).
13. Anderson DJ, Perona P. Toward a science of computational ethology. *Neuron.* (2014) 84:18–31. doi: 10.1016/j.neuron.2014.09.005
14. Crowley-Koch BJ, Van Houten R. Automated measurement in applied behavior analysis: a review. *Behav Intervent.* (2013) 28:225–40. doi: 10.1002/bin.1366
15. Michie S, Thomas J, Johnston M, Aonghusa PM, Shawe-Taylor J, Kelly MP, et al. The human behavior change project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. *Implement Sci.* (2017) 12:1–12. doi: 10.1186/s13012-017-0641-5
16. Rajkishan SS, Meitei AJ, Singh A. Role of AI/ML in the study of mental health problems of the students: a bibliometric study. *Int J Syst Assur Eng Manage.* (2023) 2023:1–23. doi: 10.1007/s13198-023-02052-6
17. Bhatt P, Sethi A, Tasgaonkar V, Shroff J, Pendharkar I, Desai A, et al. Machine learning for cognitive behavioral analysis: datasets, methods, paradigms, and research directions. *Brain Informat.* (2023) 10:18. doi: 10.1186/s40708-023-00196-6
18. Overall KL. The ethogram project. *J VetBehav.* (2014) 9:1–5. doi: 10.1016/j.jveb.2013.12.001
19. Hall C, Roshier A. Getting the measure of behavior ... is seeing believing? *Interactions.* (2016) 23:42–6. doi: 10.1145/2944164
20. Miklósi Á. *Dog Behaviour, Evolution, and Cognition.* Oxford: oUp Oxford. (2014).
21. Egnor SR, Branson K. Computational analysis of behavior. *Annu Rev Neurosci.* (2016) 39:217–36. doi: 10.1146/annurev-neuro-070815-013845
22. Zamansky A, Sinitca A, van der Linden D, Kaplun D. Automatic animal behavior analysis: opportunities for combining knowledge representation with machine learning. *Procedia Comput Sci.* (2021) 186:661–8. doi: 10.1016/j.procs.2021.04.187
23. Knaebe B, Weiss CC, Zimmermann J, Hayden BY. The promise of behavioral tracking systems for advancing primate animal welfare. *Animals.* (2022) 12:1648. doi: 10.3390/ani12131648
24. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci.* (2018) 21:1281. doi: 10.1038/s41593-018-0209-y
25. Pennington ZT, Dong Z, Feng Y, Vetere LM, Page-Harley L, Shuman T, et al. ezTrack: an open-source video analysis pipeline for the investigation of animal behavior. *Sci Rep.* (2019) 9:1–11. doi: 10.1038/s41598-019-56408-9
26. Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SSH, Murthy M, et al. Fast animal pose estimation using deep neural networks. *Nat Methods.* (2019) 16:117–25. doi: 10.1038/s41592-018-0234-5
27. Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, et al. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife.* (2019) 8:e47994. doi: 10.7554/eLife.47994.sa2
28. Romero-Ferrero F, Bergomi MG, Hinz RC, Heras FJ, de Polavieja GG. Idtracker AI: tracking all individuals in small or large collectives of unmarked animals. *Nat Meth.* (2019) 16:179–82. doi: 10.1038/s41592-018-0295-5
29. Marshall JD, Aldarondo DE, Dunn TW, Wang WL, Berman GJ, Ölveczky BP. Continuous whole-body 3D kinematic recordings across the rodent behavioral repertoire. *Neuron.* (2021) 109:420–37. doi: 10.1016/j.neuron.2020.11.016
30. Gerós A, Magalhães A, Aguiar P. Improved 3D tracking and automated classification of rodents' behavioral activity using depth-sensing cameras. *Behav Res Meth.* (2020) 52:2156–2167. doi: 10.3758/s13428-020-01381-9
31. Bala PC, Eisenreich BR, Yoo SBM, Hayden BY, Park HS, Zimmermann J. Openmonkeystudio: automated markerless pose estimation in freely moving macaques. *BioRxiv.* (2020) 11:4560. doi: 10.1101/2020.01.31.928861
32. Kabra M, Robie AA, Rivera-Alba M, Branson S, Branson KK. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat Methods.* (2013) 10:64. doi: 10.1038/nmeth.2281
33. Ferres K, Schloesser T, Gloor PA. Predicting dog emotions based on posture analysis using DeepLabCut. *Future Internet.* (2022) 14:97. doi: 10.3390/fi14040097
34. Yu R, Choi Y. OkeyDoggy3D: a mobile application for recognizing stress-related behaviors in companion dogs based on three-dimensional pose estimation through deep learning. *Appl Sci.* (2022) 12:8057. doi: 10.3390/app12168057
35. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering.* New York, NY: Association for Computing Machinery (2014). p. 1–10.
36. Karl S, Boch M, Zamansky A, van der Linden D, Wagner IC, Völter CJ, et al. Exploring the dog-human relationship by combining fMRI, eye-tracking and behavioural measures. *Scient Rep.* (2020) 10:1–15. doi: 10.1038/s41598-020-79247-5
37. Völter CJ, Lonardo L, Steinmann MG, Ramos CF, Gerwisch K, Schranz MT, et al. Unwilling or Unable? Using 3D tracking to evaluate dogs' reactions to differing human intentions. *bioRxiv.* (2022). doi: 10.1101/2022.07.09.499322
38. Völter CJ, Starić D, Huber L. Using machine learning to track dogs' exploratory behaviour in the presence and absence of their caregiver. *Animal Behav.* (2023) 197:97–111. doi: 10.1016/j.anbehav.2023.01.004
39. Ren W, Wei P, Yu S, Zhang YQ. Left-right asymmetry and attractor-like dynamics of dog's tail wagging during dog-human interactions. *Iscience.* (2022) 25:104747. doi: 10.1016/j.isci.2022.104747
40. Fux A, Zamansky A, Bleuer-Elsner S, van der Linden D, Sinitca A, Romanov S, et al. Objective video-based assessment of adhd-like canine behavior using machine learning. *Animals.* (2021) 11:2806. doi: 10.3390/ani11102806
41. Bleuer-Elsner S, Zamansky A, Fux A, Kaplun D, Romanov S, Sinitca A, et al. Computational analysis of movement patterns of dogs with ADHD-like behavior. *Animals.* (2019) 9:1140. doi: 10.3390/ani9121140
42. Watanangura A, Meller S, Farhat N, Suchodolski JS, Pilla R, Khattab MR, et al. Behavioral comorbidities treatment by fecal microbiota transplantation in canine epilepsy: a pilot study of a novel therapeutic approach. *Front Vet Sci.* (2023).
43. Wang H, Atif O, Tian J, Lee J, Park D, Chung Y. Multi-level hierarchical complex behavior monitoring system for dog psychological separation anxiety symptoms. *Sensors.* (2022) 22:1556. doi: 10.3390/s22041556
44. Engelsman D, Sherif T, Meller S, Twele F, Klein I, Zamansky A, et al. Measurement of canine ataxic gait patterns using body-worn smartphone sensor data. *Front Vet Sci.* (2022) 9:912253. doi: 10.3389/fvets.2022.912253
45. Menaker T, Zamansky A, van der Linden D, Kaplun D, Sinitca A, Karl S, et al. Towards a methodology for data-driven automatic analysis of animal behavioral patterns. In: *Proceedings of the Seventh International Conference on Animal-Computer Interaction.* New York, NY: Association for Computing Machinery (2020). p. 1–6.
46. Farhat N, Lazebnik T, Monteny J, Moons CPH, Wydooghe E, van der Linden D, et al. Digitally-enhanced dog behavioral testing. *Sci Rep.* (2023) 13:21252. doi: 10.1038/s41598-023-48423-8
47. Tsiourti C, Baynham K, Langner L, Ross M, Farhat N, Zamansky A, et al. Comparing the effects of food and toy rewards on canine problem-solving behavior: insights from the Watson task. *arXiv preprint* (2023).
48. Boneh-Shitrit T, Feighelestein M, Bremhorst A, Amir S, Distelfeld T, Dassa Y, et al. Explainable automated recognition of emotional states from canine facial expressions: the case of positive anticipation and frustration. *Sci Rep.* (2022) 12:22611. doi: 10.1038/s41598-022-27079-w
49. Aich S, Chakraborty S, Sim JS, Jang DJ, Kim HC. The design of an automated system for the analysis of the activity and emotional patterns of dogs with wearable sensors using machine learning. *Appl Sci.* (2019) 9:4938. doi: 10.3390/app9224938
50. Byosiére SE, Feighelestein M, Wilson K, Abrams J, Elad G, Farhat N, et al. Evaluation of shelter dog activity levels before and during COVID-19 using automated analysis. *Appl Anim Behav Sci.* (2022) 250:105614. doi: 10.1016/j.applanim.2022.105614
51. Menaker T, Monteny J, de Beeck LO, Zamansky A. Clustering for automated exploratory pattern discovery in animal behavioral data. *Front Vet Sci.* (2022) 9:884437. doi: 10.3389/fvets.2022.884437
52. Broome S, Feighelestein M, Zamansky A, Carreira Lencioni G, Haubro Andersen P, Pessanha F, et al. Going deeper than tracking: a survey of computer-vision based recognition of animal pain and emotions. *Int J Comput Vis.* (2023) 131:572–90. doi: 10.1007/s11263-022-01716-3
53. Li JJ, Tong X. Statistical hypothesis testing versus machine learning binary classification: Distinctions and guidelines. *Patterns.* (2020) 1:100115. doi: 10.1016/j.patter.2020.100115
54. Rathod D, Vinay A, Shylaja S, Natarajan S. Facial landmark localization-a literature survey. *Int J Current Eng Technol.* (2014) 4:1901–7. doi: 10.48550/arXiv.1805.05563

55. Ramasamy Ramamurthy S, Roy N. Recent trends in machine learning for human activity recognition a survey. *Wiley Interdiscip Rev Data Min Knowl Discov.* (2018) 8:e1254. doi: 10.1002/widm.1254
56. Cheng Y, Li L, Xu Y, Li X, Yang Z, Wang W, et al. Segment and track anything. *arXiv preprint arXiv:2305.06558* (2023).
57. Cao W, Wu Y, Sun Y, Zhang H, Ren J, Gu D, et al. A review on multimodal zero-shot learning. *Wiley Interdiscip Rev Data Min Knowl Discov.* (2023) 13:e1488. doi: 10.1002/widm.1488
58. Bateson M, Martin P. *Measuring Behaviour: An Introductory Guide.* Cambridge: Cambridge University Press. (2021). doi: 10.1017/9781108776462
59. Runeson P, Höst M. Guidelines for conducting and reporting case study research in software engineering. *Empirical Softw Eng.* (2009) 14:131–64. doi: 10.1007/s10664-008-9102-8
60. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol.* (2006) 3:77–101. doi: 10.1191/1478088706qp063oa



OPEN ACCESS

EDITED BY

Jasmin Nicole Nessler,
University of Veterinary Medicine Hannover,
Germany

REVIEWED BY

Stefanie Winzer,
DOS Software-Systeme, Germany
Anna Knebel,
University Hannover, Germany

*CORRESPONDENCE

Thomas Flegel
✉ flegel@kleintierklinik.uni-leipzig.de

[†]These authors have contributed equally to
this work and share first authorship

RECEIVED 24 March 2024

ACCEPTED 05 July 2024

PUBLISHED 22 July 2024

CITATION

Flegel T, Neumann A, Holst A-L,
Kretzschmann O, Loderstedt S, Tästensen C,
Gutmann S, Dietzel J, Becker LF, Kalliwooda T,
Weiß V, Kowarik M, Böttcher IC and
Martin C (2024) Machine learning algorithms
predict canine structural epilepsy with high
accuracy.

Front. Vet. Sci. 11:1406107.

doi: 10.3389/fvets.2024.1406107

COPYRIGHT

© 2024 Flegel, Neumann, Holst,
Kretzschmann, Loderstedt, Tästensen,
Gutmann, Dietzel, Becker, Kalliwooda, Weiß,
Kowarik, Böttcher and Martin. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Machine learning algorithms predict canine structural epilepsy with high accuracy

Thomas Flegel^{*†}, Anja Neumann^{2†}, Anna-Lena Holst¹,
Olivia Kretzschmann¹, Shenja Loderstedt¹, Carina Tästensen¹,
Sarah Gutmann¹, Josephine Dietzel¹, Lisa Franziska Becker¹,
Theresa Kalliwooda¹, Vivian Weiß¹, Madlene Kowarik¹,
Irene Christine Böttcher¹ and Christian Martin²

¹Department for Small Animals, Veterinary Faculty, Leipzig University, Leipzig, Germany, ²Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Leipzig University, Leipzig, Germany

Introduction: Clinical reasoning in veterinary medicine is often based on clinicians' personal experience in combination with information derived from publications describing cohorts of patients. Studies on the use of scientific methods for patient individual decision making are largely lacking. This applies to the prediction of the individual underlying pathology in seizing dogs as well. The aim of this study was to apply machine learning to the prediction of the risk of structural epilepsy in dogs with seizures.

Materials and methods: Dogs with a history of seizures were retrospectively as well as prospectively included. Data about clinical history, neurological examination, diagnostic tests performed as well as the final diagnosis were collected. For data analysis, the Bayesian Network and Random Forest algorithms were used. A total of 33 features for Random Forest and 17 for Bayesian Network were available for analysis. The following four feature selection methods were applied to select features for further analysis: Permutation Importance, Forward Selection, Random Selection and Expert Opinion. The two algorithms Bayesian Network and Random Forest were trained to predict structural epilepsy using the selected features.

Results: A total of 328 dogs of 119 different breeds were identified retrospectively between January 2017 and June 2021, of which 33.2% were diagnosed with structural epilepsy. An overall of 89,848 models were trained. The Bayesian Network in combination with the Random feature selection performed best. It was able to predict structural epilepsy with an accuracy of 0.969 (sensitivity: 0.857, specificity: 1.000) among all dogs with seizures using the following features: age at first seizure, cluster seizures, seizure in last 24 h, seizure in last 6 month, and seizure in last year.

Conclusion: Machine learning algorithms such as Bayesian Networks and Random Forests identify dogs with structural epilepsy with a high sensitivity and specificity. This information could provide some guidance to clinicians and pet owners in their clinical decision-making process.

KEYWORDS

dog, seizures, artificial intelligence, Random Forest, Bayesian Network, feature selection

1 Introduction

Medical decisions on diagnostic interventions, interpretation of results of diagnostic tests and treatment protocols in an individual case are currently heavily influenced by the attending clinician's level of veterinary training, participation in continuing education courses, level of knowledge of pertinent scientific literature and personal experiences gained from prior cases. That may result in completely different decisions in cases that are presented with identical clinical signs.

There are first attempts to add more objective information to the medical decision-making process using artificial intelligence (AI) (1–8). Several of those are centered around forecasting seizures. A support vector machine algorithm was investigated to predict seizures in 5 dogs based on intracranial electroencephalography (EEG) tracings (4). Authors highlight the importance of correct feature selection. Another study describes the training of a subject-specific deep-learning convolutional neural network model to predict seizures in 4 dogs based on ambulatory intracranial EEG recordings (8). This model forecasted seizures with a mean sensitivity of 0.79. Similarly, a precision-recall genetic algorithm in line with a probabilistic support vector machine classifier was used for seizure forecasting in canine epilepsy, again based on intracranial EEG recordings in 3 dogs (2).

Another focus is the prediction of epilepsy types with the regard to the underlying etiology. Epilepsy can be classified as idiopathic or structural. The latter is defined as epileptic seizures that are caused by a structural intracranial pathology, whereas the underlying cause is either genetic or not known yet in idiopathic epilepsy (9). Abani et al. (1) investigated the possibility to use ChatGPT to determine the underlying pathology causing seizures in dogs based on age, clinical signs, seizure characteristics and results of the neurological examination (2023). ChatGPT correctly identified dogs with idiopathic epilepsy based on the clinical history in 2 out of 5 cases. Adding results of the clinical examination improved the correct rate to 4 out of 5 dogs. Cases with structural brain abnormalities were correctly diagnosed in 1 out of 5 cases just based on clinical history, which could again be improved to 2 out of 5 cases by adding the results of the clinical examination. Cases with paroxysmal dyskinesia, however, were not identified by ChatGPT.

In this study we evaluated two machine learning algorithms to detect structural epilepsy in dogs.

2 Materials and methods

2.1 Data acquisition

This study is based on data that has been collected retrospectively and prospectively. In the first phase, the hospital database of the Department for Small Animals at Leipzig University was screened for dogs using the following search terms: astrocytoma, adenoma, adenocarcinoma, encephalitis, epilepsy, esthesioneuroblastoma, glioma, hydrocephalus, intoxication, lymphoma, meningioma, neoplasia, neuroblastoma, postictal, seizures, shunt, status, and cluster. The following information was extracted for dogs being identified:

age, sex, breed, body weight, age at first seizure, first symptoms observed by the owner, type of seizure, number of seizures before first presentation, observed cluster of seizures, observed status epilepticus, results of neurological examination, results of diagnostic investigation, survival time, and final diagnosis. In a second phase, the features identified to be most relevant in the first phase based on the validation of different feature selection methods were collected prospectively for patients being presented to the hospital from this time point on.

2.2 Data preprocessing

Dogs were included retrospectively as well as prospectively. Data sets of both groups of dogs were not identical. In order to integrate the data from both groups of dogs, the following parameters were chosen for further analysis: age at presentation, body weight, sex including castration status, age at first seizure, duration of seizure history, seizure in last 24 h, seizures in last week (excluding those in the last 24 h), seizures in last month (excluding those in the last week), seizures in last 6 months (excluding those in the last month), seizures in last year (excluding those in the last 6 months), neurological deficits at presentation, history of cluster seizures, history of status epilepticus, lateralized neurological deficit at presentation, first clinical signs observed by the owner, and seizure type (generalized tonic-clonic, partial, both). Entries with missing values were removed. Table keys were adapted and unified, the values were cleaned. First clinical signs were grouped to paroxysmal events, abnormal behavior, vocalization, abnormal coordination, abnormal motor movements, gastrointestinal signs, recumbency, salivation and head tremor. All classes with less than 5 members were labeled with other clinical signs. The resulting table was extended by three further columns “age in days”, “structural brain disease” (yes or no) “and “weight groups” to represent breeds. The final resulting table was preprocessed in two different ways in order to allow further processing using two different algorithms, the Bayesian Network algorithm and the Random Forest algorithm. The Bayesian Network algorithm does not allow analysis of numerical data. Therefore, numerical parameters were grouped based on clinical relevance as shown in Table 1. The grouping is done separately for each numerical parameter.

For the Random Forest model, one hot encoding was used for the features “first clinical signs observed by the owner” and “type of seizure” in order to adapt categorical data. The column “sex” was subdivided in two boolean columns “sex” and “castrated”.

2.3 Feature selection

After data preprocessing, a total of 33 features for the Random Forest algorithm and 17 for the Bayesian Network algorithm were available for analysis. In general, a smaller number of features is advantageous to achieve better results, since this reduces overfitting and generates more robust and more explainable models. Non-informative features can distract the model and may cause poorer results. Therefore, we focused on identifying features that are most relevant for the prediction. The following four feature selection methods were used: Permutation Importance, Forward Selection, Random Selection and Expert Opinion.

Permutation Importance measures the decrease in model score when a single feature value is randomly shuffled. During Forward

Abbreviations: AI, artificial intelligence; EEG, electroencephalography; MDI, mean decrease in impurity.

TABLE 1 Grouping of numerical data based on clinical relevance in order to allow analyzes using the Bayesian Network algorithm.

Body weight [kg]	0–7; > 7–15; > 15–30; > 30
Grouped age / grouped age at presentation / grouped age at first seizure [years]	<p>≤ 7 kg: a (0–0.5); b (> 0.5–9); c (> 9–12); d (> 12)</p> <p>> 7–30 kg: a (0–0.5); b (> 0.5–6); c (> 6–10); d (> 10)</p> <p>> 30 kg: a (0–0.5); b (> 0.5–5); c (> 5–9); d (> 9)</p> <p>Boxer and French Bulldog: (0–0.5); b (> 0.5–4); c (> 4–6); d (> 6)</p>
Number of seizures before first presentation	<p>within the last 24 h: 0; 1; 2–3; 4–10; > 10</p> <p>with the last week: 0; 1; 2–3; 4–10; > 10</p> <p>within the last month: 0; 1; 2–3; 4–10; > 10</p> <p>within the last 6 months: 0; 1–4; 5–10; > 10</p> <p>within the last year: 0; 1–4; 5–10; > 10</p>
Duration of seizure history before first presentation [months]	0–1; > 1–3; > 3–12; > 12

Selection, a model is trained for every single feature. The feature that performed best and the corresponding accuracy score is stored. For the remaining features each is combined with the selected feature and a model is trained. The two features performing best are selected and the model score is stored. For the remaining features the process is continued until no feature is left. That way, the features performing best are added step wise. Finally, the scores of the model series are compared and the feature set of the model with the best score is chosen. In Random Selection, a subset of the features is randomly selected to train a model. The feature subset and the accuracy score are stored. The process is repeated 1,000 times and the feature subset with the best score is chosen. In Expert Opinion, the feature subset is selected by a clinician with 20 years of experience in this field.

For the Random Forest model, additional mean decrease in impurity (MDI), a feature importance score for tree models, was used for feature selection. MDI counts the times a feature is used to split a node, weighted by the number of samples it splits (10). Feature Selection based on MDI was utilized similar to Forward Selection, but the MDI score, previously calculated through training a model with all features, determined the order of the sequentially added features. For model selection, the accuracy score was used too.

2.4 Data analysis

For data analysis, we applied two machine learning algorithms, the Bayesian Network and the Random Forest. The Bayesian Network is a probabilistic graphical model that allows to compute probabilities between symptoms and diseases (11). Given clinical data (including symptoms), it allows to compute the probability of structural epilepsy. The Random Forest is a classifier which is based on multiple decision trees and is a widely used approach for this type of classification task (12).

We used the RandomForestClassifier from the package scikit-learn in Python with its default parameters (100 trees, Gini criterion to measure the quality of a split, no maximal tree depth, at least two samples for a split, at least one sample in a leaf, initialization with the same random state 0).

For model validation, we performed a 10-fold cross validation. A k-fold-cross validation is a resampling method for model validation. For k = 10, the data is split into 10 parts, whereas 9 parts of the data is used for training and one part for testing. This is repeated 10 times,

resulting in 10 different collections of disjoint training and testing data sets. Thus, it is guaranteed that the trained models are never tested on the same data. The accuracy metric was used for model comparison and the best model was chosen.

3 Results

A total of 444 dogs of 119 different breeds were identified retrospectively between January 2017 and June 2021. Some dogs had to be removed because of incomplete data, whereas 279 dogs have been used for further analysis. An additional 49 dogs were included prospectively resulting in a total number of 328 dogs included. The most frequent breeds that were represented by at least 5 individuals were: mixed breed dog ($n = 90$), French bulldog ($n = 39$), Labrador retriever ($n = 20$), Chihuahua ($n = 11$), Bolonka zwetna ($n = 9$), Yorkshire terrier ($n = 8$), Beagle ($n = 8$), Pug dog ($n = 7$), Australian shepherd dog ($n = 6$), Jack Russell terrier ($n = 6$), Golden retriever ($n = 5$), and Great Swiss mountain dog ($n = 5$). The type and number of observed features in those 328 dogs are summarized in Table 2 for categorical features and Table 3 for numeric features. The statistics of the grouped features are in Table 1 of the supplementary.

In total, 89,848 models were trained. The following five features performed best for the Bayesian Network algorithm based on Random Selection: cluster seizures, grouped age of first seizure, seizure in last 24 h, seizure in last 6 months, and seizure in last year. The mutual correlations of these features as well as the target variable structural epilepsy are shown in Figure 1. For the Random Forest based on MDI feature selection, the following five features performed best: age at first seizure, age at presentation, age, duration of seizure history, and body weight. The mutual correlations of these features as well as the target variable structural epilepsy are shown in Figure 2.

The performance of the applied feature selection methods to predict structural epilepsy in dogs is shown in Table 4. The best results were obtained for the Bayesian Network combined with Random Selection as well as the Random Forest combined with MDI. Both approaches reached an accuracy of 0.969. The Bayesian Network reached slightly better values for sensitivity (0.857) and AUC (0.971). The selected features for all feature selection methods can be found in Table 2.

By varying the thresholds inside the two best performing models, the receiver operating characteristic (ROC) curves of the models can

TABLE 2 Features selected for training both models (Bayesian Network and Random Forest).

		All	Structural epilepsy	No structural epilepsy
n		328	102	226
Sex	Female intact	90	29	61
	Female spayed	45	18	27
	Male intact	135	38	97
	Male neutered	58	17	41
Neurological deficits on initial presentation		213	89	124
Lateralized neurological deficits on initial presentation		78	42	36
Cluster seizures		150	58	92
Status epilepticus		53	16	37
Clinical signs observed by the owner	Paroxysmal events	150	43	107
	Recumbency	47	16	31
	Abnormal behavior	38	14	24
	Salivation	24	13	11
	Gastrointestinal signs	15	4	11
	Vocalization	9	2	7
	Abnormal coordination	9	1	8
	Abnormal motor movements	6	2	4
	Head tremor	5	1	4
	Other clinical signs	25	6	19
Seizure types	Generalized seizures	279	86	193
	Focal seizures	27	6	21
	Generalized and focal seizures	22	10	12
Body weight	≤ 7 kg	70	22	48
	> 7 - ≤ 15 kg	88	33	55
	15 - ≤ 30 kg	107	26	81
	> 30 kg	63	21	42

be obtained (Figure 3). The false positive rate ($1 - \text{specificity}$) on the x-axis is plotted against the true positive rate (sensitivity) on the y-axis. The ROC curves are both far away from the diagonal (dashed line), corresponding to a random classifier, and the Areas Under the Curve (AUC) are close to 1. By walking along the ROC curve (which is done by varying the threshold inside the model), a sensitivity of 1.000 can be obtained at the cost of a lower specificity (0.800 for Bayesian network and 0.667 for Random Forest).

The Random Forest computes a feature importance based on the Gini index for each available feature as a side result. These feature importance are shown in Figure 4 (for the Random Forest combined with the MDI score). The five most important features (MDI value > 0.75) were selected by the feature selection algorithm. After that, the feature importance drops considerably. The performance of the model would deteriorate if more than the five most important features were used by the model.

4 Discussion

By applying the machine learning algorithms Bayesian Network, a sensitivity of 0.857 and a specificity of 1.000 resulting in a total

accuracy of 0.969 could be obtained when classifying dogs with and without structural epilepsy. Thus, dogs with structural epilepsy and those without were identified correctly in about 97% of cases. The Random Forest model with MDI feature selection performed similar to the Bayesian Network. It reached also an accuracy of about 97% but with slightly smaller values for sensitivity and AUC than the Bayesian Network.

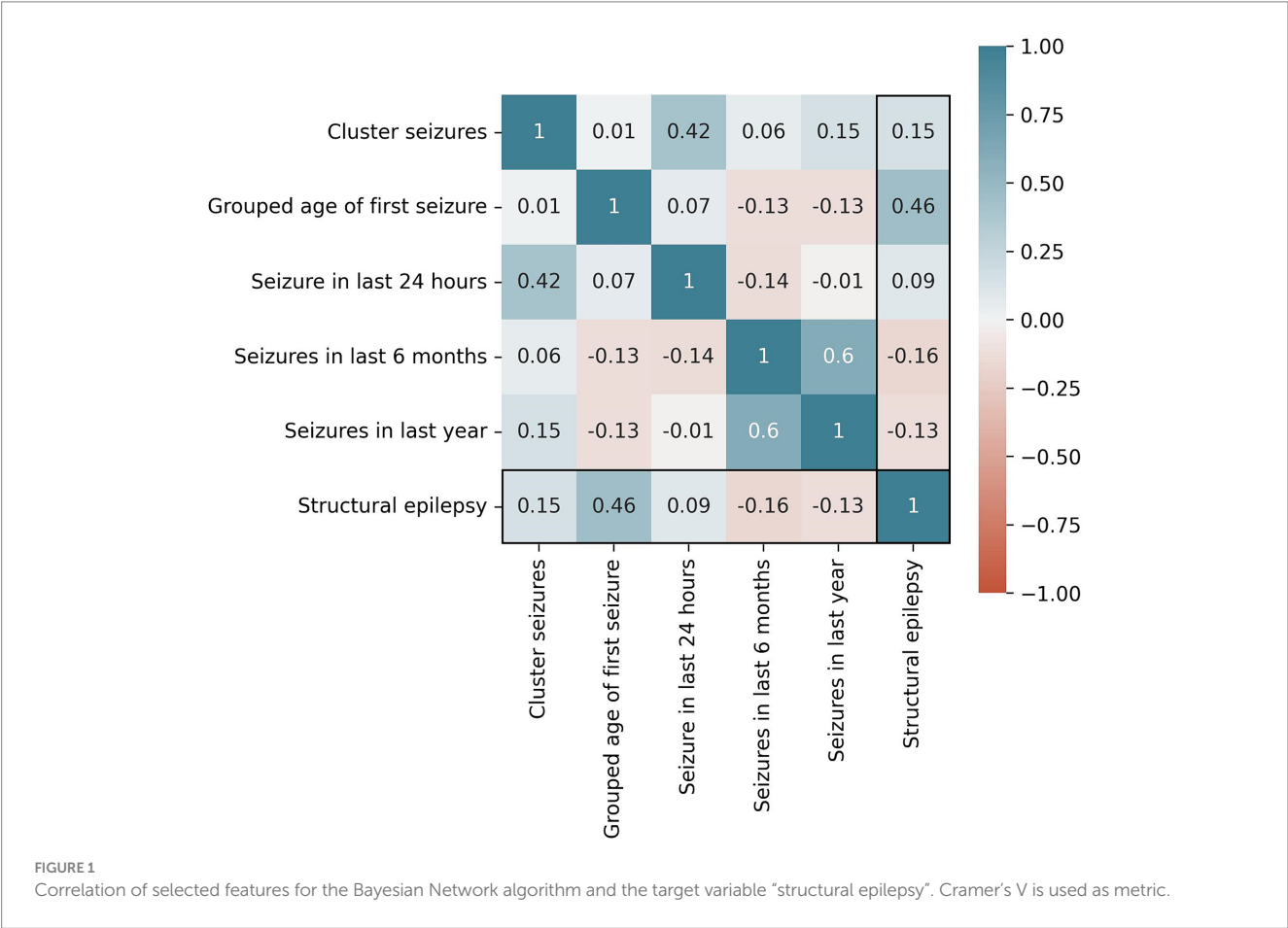
There are hardly any studies for comparison in veterinary medicine. However, these accuracies are higher than previously reported results, where ChatGPT did accurately identified structural epilepsy in 2 out of 5 dogs (1). These differences may be caused by the different approach applied in this study. ChatGPT that was used for prediction is a language processing model, that largely depends on random input made by internet users. It was applied to data without previous training. In contrast, in the study presented here, we have trained two different algorithms for prediction and the results were validated. In addition, the differences in accuracy to identify dogs with structural epilepsy may result from the large number of 238 dogs included here, whereas the study using ChatGPT looked at the small number (5 dogs) with structural epilepsy only (1).

Using artificial intelligence is an important step to support clinical reasoning. So far, most clinical decisions were based on clinicians'

TABLE 3 Comparison of numerical features between dogs with structural epilepsy and no structural epilepsy ($n = 328$) with median, range, and IQR (Q1–Q3).

	All dogs	Dogs with structural epilepsy	Dogs without structural epilepsy	p -value
Body weight [kg]	$m = 15.55$ [0.46–90] IQR: 8.3–27.5	$m = 13.73$ [1.77–49] IQR: 8.39–26.6	$m = 16.55$ [0.46–90] IQR: 8.21–27.6	0.6224
Age [days]	$m = 3,656$ [459–8,438] IQR: 2502–4,832	$m = 4,781$ [1275–7,969] IQR: 3781–5,692	$m = 3,210$ [459–8,438] IQR: 2317–4,170	1.65×10^{-12} (***)
Age at presentation [days]	$m = 2008$ [50–6,197] IQR: 939–3,287	$m = 3,273$ [79–5,866] IQR: 2309–4,253	$m = 1,586$ [50–6,197] IQR: 744–2,447	5.00×10^{-16} (***)
Age at first seizure [days]	$m = 1,646$ [48–6,197] IQR: 751–3,115	$m = 3,214$ [48–5,865] IQR: 2176–4,074	$m = 1,214$ [50–6,197] IQR: 521–2037	0.76×10^{-18} (***)
Seizures in last 24 h	$m = 1$ [0–60] IQR: 0–2	$m = 1$ [0–60] IQR: 0–3	$m = 1$ [0–12] IQR: 0–2	0.1468
Seizures in last week	$m = 0$ [0–60] IQR: 0–2	$m = 1$ [0–60] IQR: 0–2	$m = 0$ [0–12] IQR: 0–1	0.0003 (***)
Seizures in last month	$m = 0$ [0–60] IQR: 0–2	$m = 0$ [0–60] IQR: 0–1	$m = 0$ [0–30] IQR: 0–2	0.1402
Seizures in last 6 month	$m = 0$ [0–60] IQR: 0–2	$m = 0$ [0–60] IQR: 0–1	$m = 1$ [0–60] IQR: 0–2	0.0007 (***)
Seizures in last year	$m = 0$ [0–60] IQR: 0–2	$m = 0$ [0–60] IQR: 0–0	$m = 0$ [0–50] IQR: 0–1	0.0018 (**)
Duration of seizure history [days]	$m = 29$ [–271–3,249] IQR: 1–189.25	$m = 7.5$ [–223–3,249] IQR: 1–52	$m = 69$ [–271–2,557] IQR: 1–309	0.0002 (***)

Comparison of medians between both groups was performed using the Mann–Whitney–U-Test. (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.



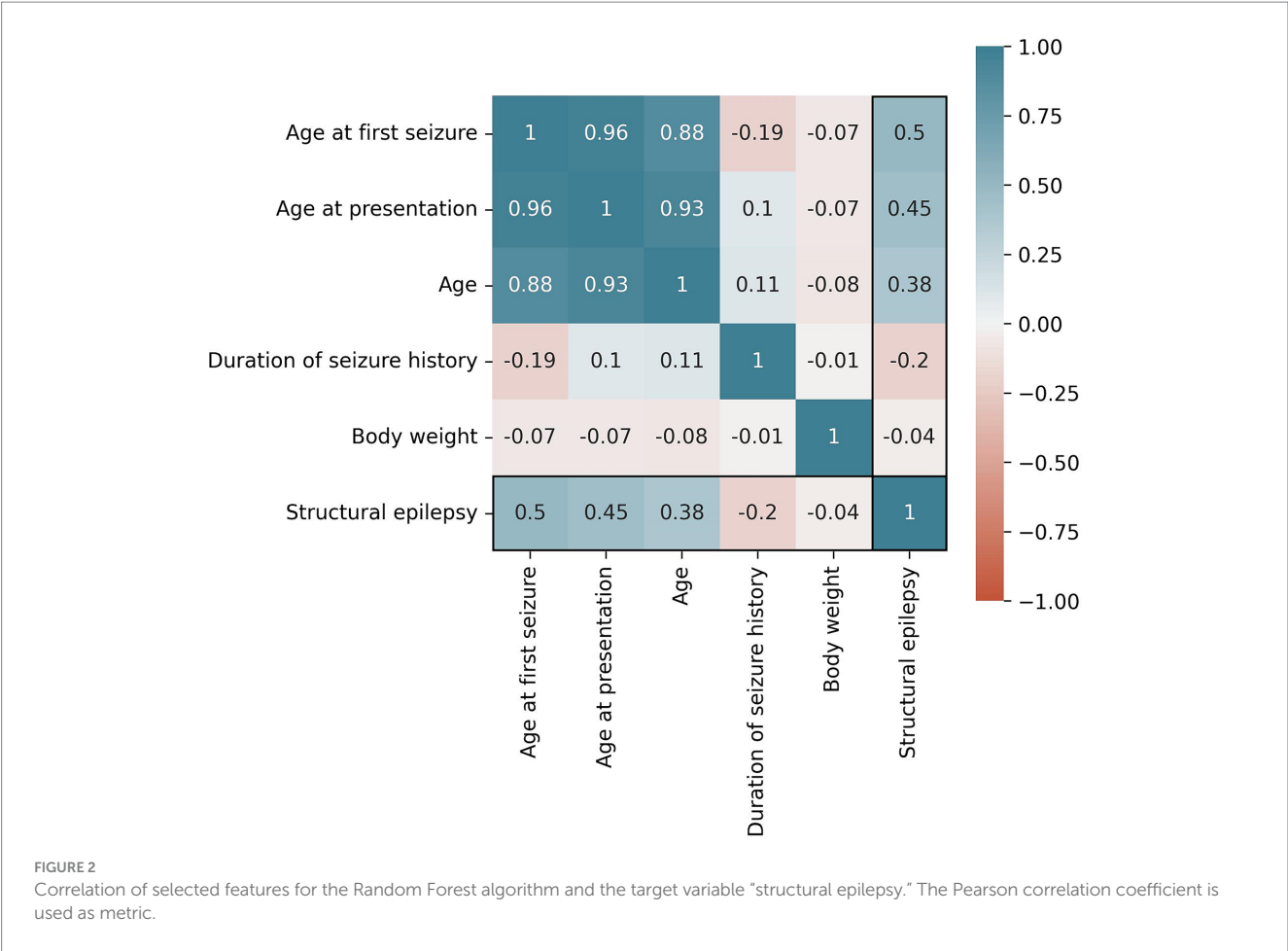


TABLE 4 Performance of the Bayesian Network and the Random Forest combined with the different feature selection methods to predict structural epilepsy (AUC: area under the curve; MDI: mean decrease in impurity).

		Permutation importance	Forward selection	MDI	Random selection	Expert opinion
Bayesian Network	Sensitivity	0.857	0.857	----	0.857	0.800
	Specificity	0.920	0.880	----	1.000	1.000
	Accuracy	0.906	0.875	----	0.969	0.938
	AUC	0.954	0.886	----	0.971	0.916
Random Forest	Sensitivity	0.714	0.714	0.800	0.714	0.857
	Specificity	0.960	1.000	1.000	0.960	0.960
	Accuracy	0.906	0.938	0.969	0.906	0.938
	AUC	0.923	0.909	0.937	0.957	0.911

expertise and personal experiences combined with knowledge derived from studies describing cohorts of patients. Attempts to predict the risk of certain intracranial pathologies in dogs with seizures were already made in the past. It was found that dogs having just one single seizure were less likely to suffer from a lateralized structural brain disease, whereas those dogs with abnormal findings on the neurological examination had a 16.5 times higher risk for such a lateralized lesion and a 12.5 times higher risk for a symmetrical structural lesion (13). However, that is a completely different approach than in our study, which was aiming for the prediction of the individual risk of dogs to suffer from structural epilepsy.

In the study presented here, two algorithms, the Bayesian Network and the Random Forest had been intentionally selected in order to make usage of the advantages of different algorithms. A Bayesian Network is a probabilistic graphical model. It treats uncertainty explicitly, and is thus suitable for small and incomplete data sets and inference with incomplete data. This algorithm provides a conditional probability distribution for every combination of variable values. However, the Bayesian Network requires discretization of continuous variables, hence it is poor in finding linear relationships, but it performs well if relationships between the variables are non-linear and complex.

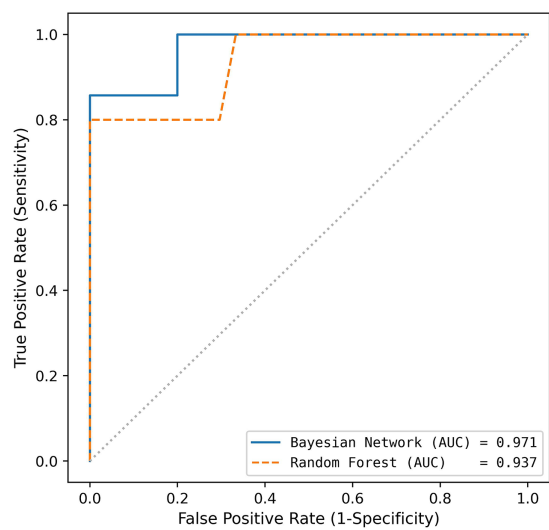


FIGURE 3
Receiver operating characteristics for both algorithms used (blue: Bayesian Network using Random Selection; orange: Random Forest using MDI Selection).

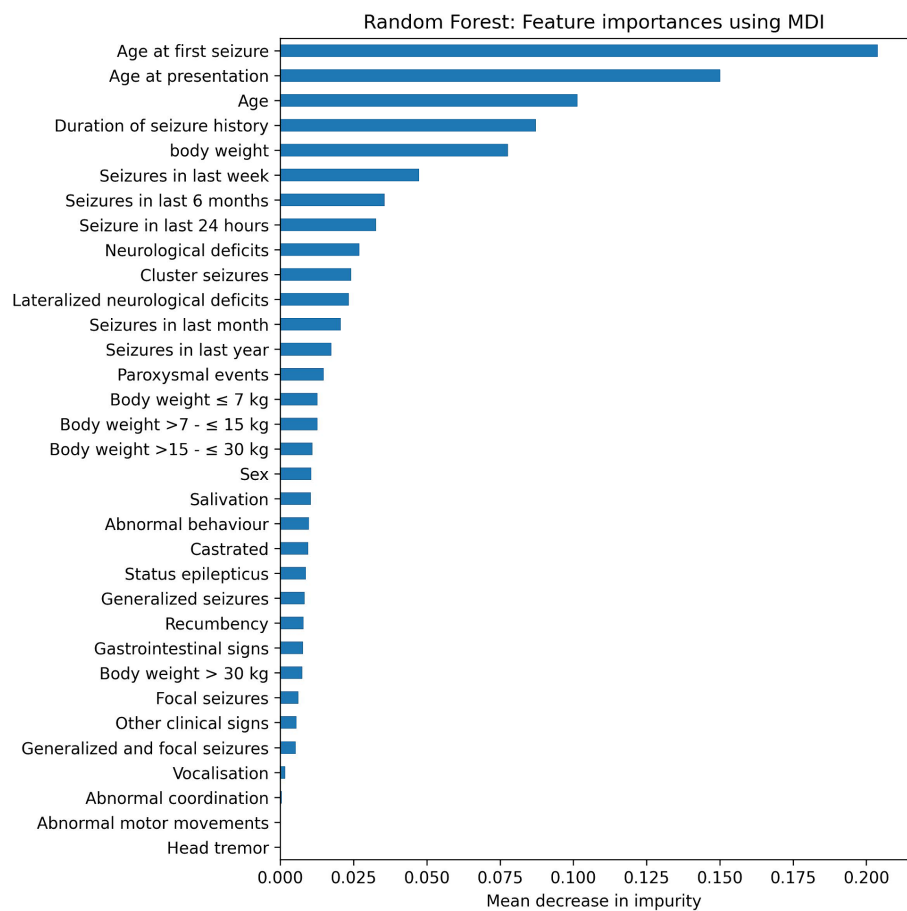


FIGURE 4
Feature importance using the Random Forest algorithms with Forward Selection in combination with the mean decrease in impurity (MDI).

The Random Forest model is an ensemble method using decision trees. It is robust to outliers and overfitting, can deal with linear and non-linear relationships and offers feature importance as a byproduct. It can handle numeric and categorical variables. It is not necessary to construct artificial categories when transforming numeric variables into categorical ones. To use categorical variables with equal order, these variables can be transformed with one hot encoding to boolean variables. Strongly correlated features are often problematic, however, Random Forest can handle them much better than linear regression models. If there are two strongly correlated features, the Random Forest algorithm randomly picks only one of them in a split. Thus, the correlated features are less likely to show up together at the same position in a tree.

For the reasons explained, we considered those two algorithms to be among the best for analyzing the data obtained in this study. However, it cannot be excluded that other algorithms could have resulted in an even higher diagnostic accuracy. We used the standard parameters when training the models. A hyperparameter tuning might be applied to further improve the classification results.

For any of the algorithms available, the decision on which features to be included is crucial (4). Feature selection allows to identify and rank the clinical parameters according to their importance with respect to the classification. Therefore, different features selection methods have been tested here, all of which have advantages and disadvantages. The Random Selection method can identify feature subsets consisting of low ranked single features but it does allow double selection. Forward Selection, however, provides good performing feature sets, where multiple solo ranked features may perform potentially better, but it can be a relatively slow model, whereas Permutation Importance is fast, easy to use and it is readily available since it is included in scikit-learn, a freeware online library for machine learning tools.

The Random Selection methods provided best results here if applied to the Bayesian Network or Random Forest algorithm. However, it cannot be excluded that other features that had been excluded during the initial phase of feature selection or features, that were not even collected could perform better. In addition, the number of dogs included, even though it appears relatively large for a veterinary study, was still not large enough to investigate certain features. Specifically, including the feature “breed” would have been very interesting, since clinical observation indicates that certain canine breeds are predisposed for specific structural intracranial diseases. For instance, Pug dogs and Yorkshire terriers are predisposed for necrotizing encephalitis, whereas Boxers and French bulldogs are more frequently affected by intracranial neoplasias (14–18). However, only 12 of the 88 different breeds were represented by at least 5 individuals and only 5 breeds were represented by at least 10 individuals. Therefore, a meaningful analysis of the feature “breed,” as interesting as it would have been, was not feasible.

Interestingly, the feature selection method Expert Opinion did not perform as good as the Random or MDI feature selection methods in identifying features to determine if dogs are affected by structural epilepsy, although an accuracy of about 94% was reached. It performed better than Forward Selection and Permutation Importance. This method is the closest of all methods to the current clinical decision-making process, where intracranial neoplasias are commonly associated with neurological deficits (specifically lateralized neurological signs) and therefore those had been included as features into the Expert Opinion model (13, 19). This discrepancy between

clinical experience and a slightly lower accuracy of the Expert Opinion model may be at least partially explained by the fact that neoplasias in certain brain areas such as the olfactory and frontal lobes rarely cause interictal neurological deficits (18). In addition, pituitary neoplasia may be associated with rather unspecific signs such as disorientation or obtundation, that do not present any lateralization (20). Therefore, the importance of lateralized neurological deficits for prediction of structural epilepsy may have been overrated by the expert in the Expert Opinion feature selection method. That is supported by an only mid-range MDI of lateralized neurological deficits using the Random Forest algorithms with Forward Selection (Figure 4).

Of the 328 dogs, only 102 were diagnosed with structural epilepsy, therefore the target variable is slightly imbalanced. In order to also capture the performance of the model on the minority class, we used the metrics sensitivity, specificity, accuracy, did a ROC analysis, and computed AUC values. In some scenarios, a specificity of 1.000 was reached, but this often comes at the cost of lower sensitivity.

To avoid overfitting, feature selection and cross-validation were used. Feature selection acts as regularization, as only the most important features are selected for the final model. Cross-validation ensures that the trained model is always tested on test data that has not been used for training the model.

A weakness of this study is that some of the data was based on the observations made by the owners during collecting the clinical history of the pet. Different owners may have described the same clinical signs in a different way resulting in seemingly different observations. However, the data could only be extracted retrospectively from the patients' files for most dogs based on the study designs. We have grouped some of this information during data processing in order to reduce the tremendous amount of different observations made by the owners for a meaningful analysis. Erroneous classification may have occurred in some cases during this grouping process that was done in order to transform individual verbal owner descriptions into repetitive clinical findings. That obstacle could be overcome in the future by providing pet owners and clinicians with a set of specific questions that have to be answered by selecting options from a predetermined menu. Another potential weakness of this study might be, that some data was based on the subjective assessment of the dogs' neurological status by clinicians. Neurological examination performed by other clinicians may have resulted in slightly different findings. Therefore, the results presented here may not be safely applied to every other clinical setting. For further studies, it is desirable to identify features or feature expressions that can be repeatably and reliably obtained by different clinicians.

In conclusion, it can be said that structural epilepsy can be predicted with high sensitivity and specificity in dogs with seizures using machine learning algorithms. This information is not meant to replace further diagnostic tests in affected dogs, but it may rather facilitate client communication. The knowledge of the likelihood of structural epilepsy could be used as guidance to decide about appropriate diagnostic steps in dogs presented for seizures.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The animal studies were approved by Ethics Committee of the Veterinary Faculty of Leipzig University EK 2/2024. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was not obtained from the owners for the participation of their animals in this study because this study did retrospectively evaluate patient data available in the hospital data base.

Author contributions

TF: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing. AN: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. AH: Investigation, Writing – original draft, Writing – review & editing. OK: Investigation, Validation, Writing – original draft, Writing – review & editing. SL: Investigation, Writing – original draft, Writing – review & editing. CT: Investigation, Writing – original draft, Writing – review & editing. SG: Investigation, Writing – original draft, Writing – review & editing. JD: Investigation, Writing – original draft, Writing – review & editing. LB: Investigation, Writing – original draft, Writing – review & editing. TK: Investigation, Writing – original draft, Writing – review & editing. VW: Investigation, Writing – original draft, Writing – review & editing. MK: Investigation, Writing – original draft, Writing – review & editing. IB: Investigation, Writing – original draft, Writing – review & editing. CM: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing.

References

- Abani S, De Decker S, Tipold A, Nessler JN, Volk HA. Can ChatGPT diagnose my collapsing dog? *Front Vet Sci.* (2023) 10:1245168. doi: 10.3389/fvets.2023.1245168
- Assi EB, Nguyen DK, Rihana S, Sawan M. A functional-genetic scheme for seizure forecasting in canine epilepsy. (2018) *IEEE trans. Biomed Eng.* (2018) 65:1339–48. doi: 10.1109/TBME.2017.2752081
- Barge P, Oevermann A, Maiolini A, Durand A. Machine learning predicts histologic type and grade of canine gliomas based on MRI texture analysis. *Vet Radiol Ultrasound.* (2023) 64:724–32. doi: 10.1111/vru.13242
- Brinkmann H, Patterson EE, Vite C, Vasoli VM, Crepeau D, Stead M, et al. Forecasting seizures using intracranial EEG measures and SVM in naturally occurring canine epilepsy. *PLoS One.* (2015) 10:e0133900. doi: 10.1371/journal.pone.0133900
- Feighelstein M, Ehrlich Y, Naftaly L, Alpin M, Nadir S, Shimshoni I, et al. Deep learning for video-based automated pain recognition in rabbits. *Sci Rep.* (2023) 13:14679. doi: 10.1038/s41598-023-41774-2
- Nejedly P, Kremen V, Sladky V, Nasser M, Guragain H, Klimes P, et al. Deep-learning for seizure forecasting in canines with epilepsy. *J Neural Eng.* (2019) 16:036031. doi: 10.1088/1741-2552/ab172d
- Spiteri M, Knowler SP, Rusbridge C, Wells K. Using machine learning to understand neuromorphological change and image-based biomarker identification in cavalier king Charles spaniels with Chiari-like malformation-associated pain and syringomyelia. *J Vet Intern Med.* (2019) 33:2665–74. doi: 10.1111/jvim.15621
- Varatharajah Y, Berry BM, Worrell GA, Brinkmann BH. Seizure forecasting and the Preictal state in canine epilepsy. *Int J Neural Syst.* (2017) 27:1650046. doi: 10.1142/S0129065716500465
- Berendt M, Farquhar RG, Mandigers PJJ, Pakozdy A, Bhatti SFM, De Risio L, et al. International veterinary epilepsy task force consensus report on epilepsy definition, classification and terminology in companion animals. *BMC Vet Res.* (2015) 11:182. doi: 10.1186/s12917-015-0461-2
- Lee C (2017) Feature importance measures for tree models part I. Available at: <https://medium.com/the-artificial-impostor/feature-importance-measures-for-tree-models-part-i-47f187c1a2c3>
- Niedermayer D. An introduction to Bayesian networks and their contemporary applications. In: *Innovations in Bayesian networks: Theory and applications.* eds. DE Holmes and LC Jain Berlin: Springer (2008). 117–30.
- Ho TK. (1995). Random decision forests. *Proceeding of 3rd International Conference on Document Analysis and Recognition* 278–282.
- Armaşu M, Packer RMA, Cook S, Solcan G, Volk HA. An exploratory study using a statistical approach as a platform for clinical reasoning in canine epilepsy. *Vet J.* (2014) 202:292–6. doi: 10.1016/j.tvjl.2014.08.008
- Elbert JA, Yau W, Rissi DR. Neuroinflammatory diseases of the central nervous system of dogs: a retrospective study of 207 cases (2008–2019). *Can Vet J.* (2022) 63:176–83.
- Higginbotham MJ, Kent M, Glass EN. Noninfectious inflammatory central nervous system diseases in dogs. *Comp Contin Educ.* (2007) 29:488–97.
- Higgins RJ, Bollen AW, Dickson PJ, Sisó-Llonch S. Tumors of the nervous system In: DJ Meuten, editor. *Tumors in domestic animals. 5th ed.* Ames, IA: Iowa State Press, Ames (2017). 834–91.
- Kishimoto TE, Uchida K, Chambers JK, Kok MK, Son NV, Shiga T, et al. A retrospective survey on canine intracranial tumors between 2007 and 2017. *J Vet Med Sci.* (2020) 82:77–83. doi: 10.1292/jvms.19-0486
- Miller AD, Miller CR, Rossmeisl JH. Canine primary intracranial Cancer: a Clinicopathologic and comparative review of glioma, meningioma, and choroid plexus tumors. *Front Oncol.* (2019) 9:1151. doi: 10.3389/fonc.2019.01151
- Snyder JM, Shofer FS, Van Winkle TJ, Masicotte C. Canine intracranial primary neoplasia: 173 cases (1986–2003). *J Vet Med Int Med.* (2006) 20:669–75. doi: 10.1111/j.1939-1676.2006.tb02913.x
- Menchetti M, De Risio L, Galli G, Cherubini GB, Corlazzoli D, Baroni M, et al. Neurological abnormalities in 97 dogs with detectable pituitary masses. *Vet Q.* (2019) 39:57–64. doi: 10.1080/01652176.2019.1622819

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The APC was supported by the Open Access Publishing Fund of Leipzig University as well as by Center for Scalable Data Analytics and Artificial Intelligence of Leipzig University, Germany.

Acknowledgments

The authors AN and CM acknowledge the financial support by the German Federal Ministry of Education and Research (BMBF) and the Saxony State Ministry SMWK in the program Center of Excellence for AI-research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



OPEN ACCESS

EDITED BY

Sebastian Meller,
University of Veterinary Medicine Hanover
Foundation, Germany

REVIEWED BY

Amber Adams Progar,
Washington State University, United States
Muhammad Amirul Abdullah,
Universiti Malaysia Pahang, Malaysia

*CORRESPONDENCE

Jiajia Wang
✉ johnnyjiajia@163.com

RECEIVED 25 April 2024

ACCEPTED 23 October 2024

PUBLISHED 20 November 2024

CITATION

Zhang F, Sun H, Wang J, Wang X, Qiu Y, Cui X
and Ali S (2024) Research on motion
characterization of goose neck in narrow
space. *Front. Vet. Sci.* 11:1423453.
doi: 10.3389/fvets.2024.1423453

COPYRIGHT

© 2024 Zhang, Sun, Wang, Wang, Qiu, Cui
and Ali. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Research on motion characterization of goose neck in narrow space

Fu Zhang^{1,2}, Haoxuan Sun^{1,2}, Jiajia Wang^{2,3*}, Xinyue Wang^{1,2},
Yubo Qiu¹, Xiahua Cui⁴ and Shaukat Ali⁵

¹College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang, China, ²Longmen Laboratory, Luoyang, China, ³College of Engineering and Technology, Jilin Agricultural University, Changchun, China, ⁴College of Biological and Agricultural Engineering, Jilin University, Changchun, China, ⁵Wah Engineering College, University of Wah, Wah Cantt, Pakistan

Introduction: Inspired by the obstacle avoidance mechanism of goose neck, a theoretical design method of bionic robotic arm was proposed to solve the contradiction between high flexibility and strong bearing capacity in narrow space.

Methods: Taking the goose neck as the test object, a narrow space test environment with a width of 10 cm was built, and a 6 × 4 obstacle matrix was set up, to analyze the maximum value of joint angle, motion rate and trajectory in different target areas.

Results: The test results showed that the goose neck movement has continuity and transmissibility. The overall posture of the goose neck was adjusted through the synergistic movement of the anterior, middle and posterior segments to move toward the target position. The regulating effect of the anterior segment was significantly stronger than that of the middle and posterior segments. Specifically, the anterior segment of goose neck exhibited mostly transverse movement, with significant horizontal regulation; the middle segment of the goose neck was coupled with longitudinal movement, with similar movement ability in all directions, the posterior segment of the goose neck has mostly longitudinal movement, with significant height regulation.

Conclusion: In addition, the YOLOv7-pose recognition network was used to recognize goose neck motion pose, which provides a new method for animal behavior research.

KEYWORDS

bionics, goose neck, narrow space, motion trajectory, joint angle

1 Introduction

Robotic arms have been used in a wide range of applications in fields such as healthcare, automotive manufacturing, aerospace, food production, and agricultural production (1–3). With the increase of industrialization level and equipment integration, a large number of jobs need to be operated in narrow space, which has high requirements for the flexibility, stability and load capacity of robotic arms (4–6). Traditional robotic arms, while having high structural rigidity and strong load capacity, lack the dexterity required for operations in narrow space. Bionic flexible robotic arms offer high flexibility and the ability to continuously deform, exhibiting greater movement and operational capabilities in narrow spaces compared with traditional articulated robotic arms. However, it is accompanied by the problem of lower load capacity. Avian neck has ultra-high motion stability with the cervical vertebra serving as the most flexible part of its spine (7). Typically, the number of vertebrae ranges from 11 to 25 (8), each capable of bending in two directions, providing flexible and stable control of the head, which can weigh several times more than the neck, to perform various activities (9). Research into the motion mechanism of the avian neck in

narrow spaces can provide a theoretical basis for resolving the contradiction between the flexibility and load capacity of robotic arms.

Existing research on avian neck kinematics primarily focuses on domestic chickens, owls, ostriches, geese, etc. Movement forms include natural walking, feeding, and flying. van der Leeuw et al. (10) analyzed the characteristics and patterns of feeding and drinking movements in domestic chickens and geese, revealing that the cervical vertebrae of domestic chickens adhere to the geometric principle of maximizing angular efficiency. Furet et al. (11, 12) used CT scanning to obtain the three-dimensional surface characteristics and static maximum range of flexion of an owl's cervical vertebrae, and established a motion model imitating the bird neck linkage based on surface contact characteristics. Krings et al. (13, 14) employed X-ray technology to capture the natural neck posture when the heads of live and cadaveric owls were rotated, and utilized CT scanning to study the shape of a single vertebra. Their study indicated that rotational motion can be described as a combination of movement on the yawing axis (yawing) and the cross-rolling axis (rolling). Panyutina et al. (15) utilized CT scanning and a joint coordinate system (JCS) to investigate extreme head turns in owls, discovering that during full head turns, the maximal joint angles alternate along the neck in three planes, with maximal axial rotation (to the side of the head turn) followed by maximal sagittal bending (in the ventral direction) and then by maximal lateral bending (to the side of the head turn). Kambic et al. (16) categorized the range of motion of the avian neck into three areas: the cranial joints, predominantly in ventral flexion with high axial rotation and lateral flexion activity; the caudal joints, predominantly in dorsiflexion with low axial rotation activity and high lateral flexion activity; and the intermediate joints, variable in axial rotation activity and exhibiting low lateral flexion. Wang et al. (17, 18) studied the structural characteristics of the bird neck skeleton and static motion characteristics using CT scanning, while employing biplane X-ray techniques to study the passive stabilizing motion of the goose neck. The results indicated that the goose neck has the best passive motion stability in the sagittal plane for two directions of motion. Abourachid et al. (19, 20) combined the contact characteristics of the vertebrae to establish an intervertebral motion model suitable for birds, demonstrating that the saddle-shaped articular processes of avian vertebrae limit joint mobility, with the position and orientation of the articular processes determining vertebral mobility. The orientation of the articular eminence surfaces dictates the range of motion in dorsiflexion and lateral flexion, while the axial angle of the articular zygapophyses surfaces effectively restricts longitudinal rotation. Dzemski and Christian (21) studied the ostrich's neck and showed that it could be divided into three segments with varying flexibility, the upper portion being more flexible in the dorsal-ventral and lateral directions, the middle portion having the highest dorsal-ventral flexibility, and the bottom portion having the highest lateral flexibility. Gunji et al. (22, 23) analyzed typical ostrich neck behavior by establishing an ostrich neck dynamics model and found that ostrich neck movement involved both lever and rolling actions. He et al. (24) observed that during walking, a chicken's neck extends and contracts alternately to provide intermittent fixed stability for the head. They designed a bionic vertebrae unit combining springs and universal joints to

simulate chicken cervical vertebrae, investigated the connection and motion characteristics, and found that the unit could achieve an S-shaped bionic bending configuration and successfully wind and lift objects of interest, proving that the proposed robot has excellent flexibility and application potential, and that the design method is effective.

Research has been conducted on the structural characteristics of bird neck, range of motion, and the maintenance of head stability, among other aspects. Most of the studies have used CT scanning equipment to obtain the structural characteristics of bones. Anatomical methods were used to obtain the structural characteristics of muscles. *In vivo* motion studies have used X-ray video or biplane X-ray motion analysis system to test the motion of bird neck vertebrae. The research on the structure and motion characteristics of live bird neck have mainly focused on free motion in open space, and mostly on vertebral kinematic analyses. However, the motion characteristics of the bird neck in narrow space based on the overall motion posture and muscle distribution characteristics of the bird neck are still unclear, and the research to analyze its obstacle avoidance mechanism in depth has not been reported. In this research, the goose neck was selected as the subject of investigation, muscle distribution characteristics were analyzed by using goose neck MRI data. The high-speed camera system was used to collect the movement data of the goose neck in the narrow space, and the goose neck motion pose was extracted by YOLOv7-pose recognition network. Then, combining the characteristics of goose neck muscle distribution and movement posture to analyze the angular changes and movement trajectories of goose neck joints. Finally, the movement mechanism of the goose neck was explored in the narrow space, which provide the theoretical basis for the research on goose neck kinematics and design optimization of the bionic multi-joint robotic arms.

2 Materials and methods

2.1 Experimental equipment and software

The following equipment and software were used in this research:

- 1) MRI scanning was accomplished by Siemens Spectra 3.0T MRI scanning system (Siemens, Munich, Germany).
- 2) Goose neck motion data acquisition and analysis was by Phantom Miro series M110 high-speed camera system (Vision Research Inc, New Jersey, USA), and the PCC software (Vision Research Inc, New Jersey, USA).
- 3) Goose neck key point dataset was produced Labelme software (Massachusetts Institute of Technology, Boston, USA).
- 4) The Pytorch framework was used to build the YOLOv7-pose pose recognition network, and the model was trained on a DELL Precision 7820 workstation. The hardware configuration of the workstation was as follows: CPU is Intel® Xeon® Silver 4210R with a main frequency of 2.4 GHz, RAM is 32 GB, GPU is NVIDIA Quadro RTX 5000, Video Memory is 16 GB, and the operating system is Windows 10 Professional. For the software platform,

TABLE 1 MRI scanning parameters.

Plane	TR/ms	TE/ms	Slice thickness/mm	SNR
Sagittal	1,100.00	36.00	0.80	1.00
Coronal	2,950.00	98.00	4.00	1.00
Transverse	8,550.00	95.00	3.50	1.00

PyCharm 2023.2.1 was used as the IDE, CUDA 12.1 as the GPU-accelerated computing platform, Python version 3.10, and PyTorch version 2.1.0.

2.2 Goose neck MRI experiment

In order to investigate the effect of goose neck muscles on its movement, Magnetic Resonance Imaging (MRI) data was collected from the neck of goose to obtain the distribution of goose neck muscles. The body of the goose was cleaned before the test to ensure that there is no dirt or foreign matter on the surface of the goose's body in order to minimize the noise points of the MRI data. To prevent the goose from moving and affecting the scanning effect, the goose was injected with an appropriate amount of Zoletil 50 anesthesia drug before the MRI scanning to ensure that was in sleeping state during the test and could return to the normal state after the test. The anesthesia process was completed by a professional veterinarian.

The goose neck MRI scanning test was performed at Dongfang Hospital, Luoyang City, Henan Province, P. R. China, with Siemens Spectra 3.0T MRI scanning system. The feeding conditions of test animals are in accordance with the standard (GB14925) and comply with the requirements of animal welfare (GB/T 42011-2022). Adult gray goose with growth period of 2 years and weight of about 4 kg was selected as test subject, which was able to move neck freely and had no history of neck disease. The sleeping goose was positioned in a recumbent posture on the MRI scanning platform to ensure stable fixation on the platform to minimize movement and distortion. The neck of the goose was adjusted according to the requirements of the experiment ensuring the location of the area of interest within the scanning range.

Appropriate scanning parameters were set according to the characteristics of the goose neck as shown in Table 1. The MRI scanner was started and goose neck was scanned according to the preset parameters. The stable and accurate scanning process were ensured, and the scanning time and sequence were recorded.

Based on the MRI scanning images, the sagittal plane, coronal plane, and transverse plane muscle distribution of the anterior, middle, and posterior segments of the goose neck were comparatively analyzed. Observing from the sagittal plane of the goose neck, it can be seen that the goose neck muscle was a long muscle that extends from the anterior to the posterior segment, as shown in Figure 1. This long muscle controlled the movement of the goose neck by means of a longitudinal connection. A similar muscular connection can be seen from the coronal plane of the goose neck in Figure 1. The muscle still showed a longitudinal distribution, connecting the anterior and posterior ends of the

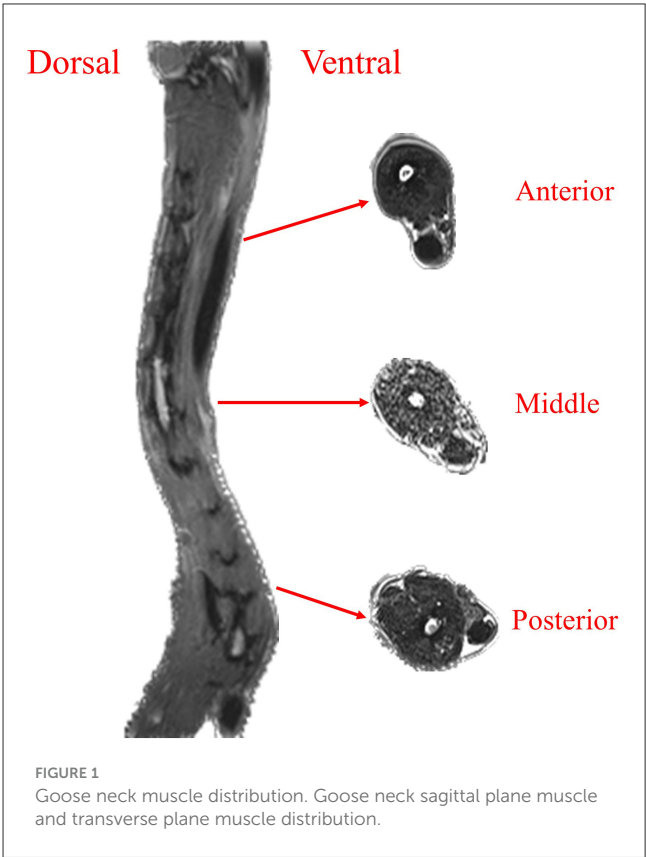


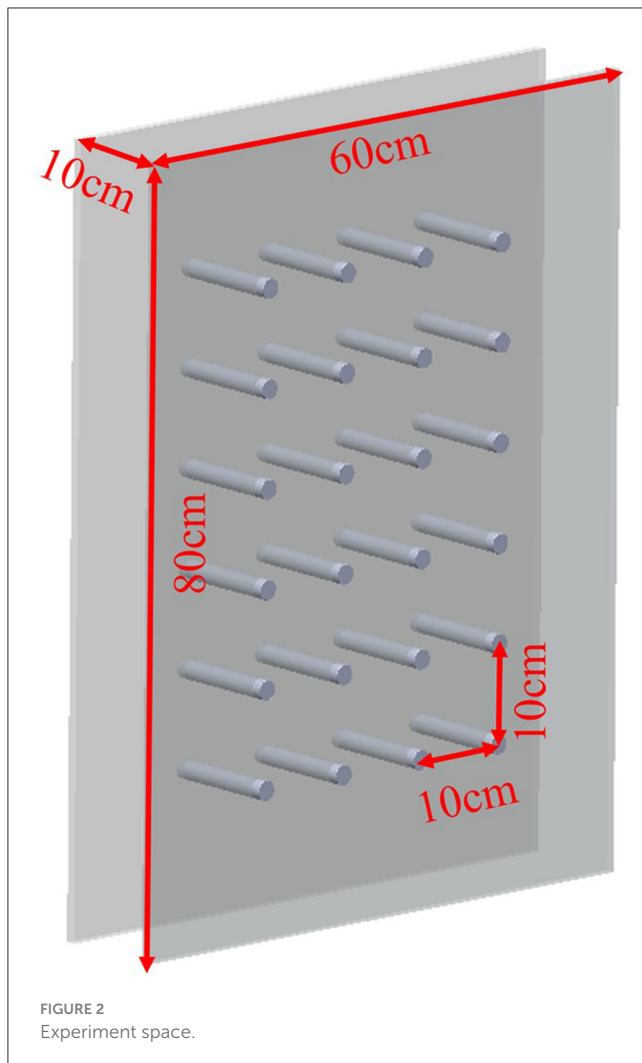
FIGURE 1
Goose neck muscle distribution. Goose neck sagittal plane muscle and transverse plane muscle distribution.

goose neck and controlling the lateral bending and rotational movements of the goose neck.

From the transverse plane of the goose neck, the area of the ventral muscles in the anterior segment of the goose neck was significantly larger than that of the dorsal muscles. The distribution of the muscles in the middle segment of the goose neck was relatively uniform, and there was no significant difference in the area of the ventral and dorsal muscles. The area of the ventral muscles in the posterior segment of the goose neck was significantly smaller than that of the dorsal muscles. It indicated that the ventral muscles in the anterior segment of the goose neck required greater muscle strength to support the ventral flexion movement of the anterior segment of the goose neck, the dorsal muscles in the posterior segment of the goose neck required greater muscle strength to support the dorsiflexion movement of the posterior segment of the goose neck.

2.3 Narrow space goose neck motion test

To investigate the movement of a goose neck in a narrow space, it was necessary to establish a testing area containing obstacles. When constructing the test space, it was important to ensure that the space was safe and stable to ensure a reliable and accurate test. Two transparent acrylic panels were used, and arranged in parallel at 10 cm intervals to form the narrow space. Using transparent acrylic rods with a diameter of 4.5 mm, a matrix of obstacles with



six rows and 4 columns was set up at 10 cm intervals in the narrow space as shown in Figure 2.

According to the distribution of goose neck muscles, the goose neck was divided into three parts, named as anterior, middle and posterior. Marking points were made at the goose cervical vertebrae C2, C5, C8, C10, and C12 with a red marker pen, labeled as L1, L2, L3, L4, and L5, respectively. The change of angle L1–L2–L3 was selected to indicate the motion of the anterior middle segment of the cervical vertebrae. The change of angle L2–L3–L4 indicated the motion of the anterior middle segment of the goose cervical vertebrae relative to the middle segment of the cervical vertebrae. While, the change of angle L3–L4–L5 indicated the motion of the middle segment of the goose cervical vertebrae relative to the posterior segment of the cervical vertebrae. They were labeled as θ_1 , θ_2 , and θ_3 , respectively, to facilitate the subsequent analysis.

The experiments were conducted using Phantom Miro series M110 high-speed camera system for the acquisition of goose neck motion data in a confined space, and PCC software was used for the data acquisition and processing.

The motion of the goose neck exhibits uncertainty. Hence, for the ideal joint angle, motion trajectory, motion posture and other

such parameters, the specific test preparation and process were carried out as follows:

- 1) A narrow space test environment with a width of 10 cm was built before the test.
- 2) In order to obtain a complete image sequence, the lens shooting direction was kept perpendicular to the direction of goose neck movement.
- 3) To set the high-speed camera parameters, Resolution was $1,280 \times 720$, Sample Rate was 200 fps, Exposure Time was $4,999.54 \mu\text{s}$, Exposure Index was 6,400.
- 4) In the six rows and four columns obstacle matrix, the target position was adjusted in turn, every test was repeated and saved.

The experiment as shown in Figure 3.

2.4 Goose neck motion data processing

2.4.1 Goose neck motion joint angle treatment

In order to obtain the joint angle of the goose neck movement in a narrow space, the goose neck movement data was processed by PCC software. The measurement unit parameters and the scale were set and calibrated. The distance unit was set to meters (m), the velocity unit was set to meters per second (m/s), the acceleration unit was set to meters per square second (m/s^2), the angle unit was set to radians (rad), and the angular velocity unit was set to radians per second (rad/s). The 3-point method was used to measure the angle data of the goose neck joint angles θ_1 , θ_2 , and θ_3 , and the measurements were carried out by extracting 1 frame at every interval of 10 frames.

Goose neck kinematic joint angles θ_1 , θ_2 , and θ_3 were processed by Matlab software. The smoothing of joint angle data was performed by six data smoothing methods, namely, moving, lowess, loess, sgolay, rlowess and rloess. The Root Mean Square Error (RMSE), Signal-to-Noise Ratio (SNR) and decidability coefficient (R^2) were used as the evaluation indexes to select the optimal data smoothing method. Among these metrics, the RMSE value indicates the deviation between the denoised data and the original data, with smaller values indicating less deviation. The SNR value reflects the credibility of the data, with larger values indicating higher credibility. Lastly, the R^2 value signifies the fit between the denoised data and the original data, with larger values indicating a higher degree of fit. It can be found from Table 2, that the sgolay method has the smallest RMSE value and the largest SNR and R^2 values, so it was finally chosen for data smoothing.

According to the different distances and heights of the target points, the experimental areas were divided into six distinct areas labeled as Area 1, Area 2, Area 3, Area 4, Area 5, and Area 6. The distribution of these areas is depicted in Table 3 and illustrated in Figure 3.

2.4.2 Goose neck motion trajectory processing

In order to obtain the trajectory of the goose neck in a narrow space, the data collected by the high-speed camera needed to

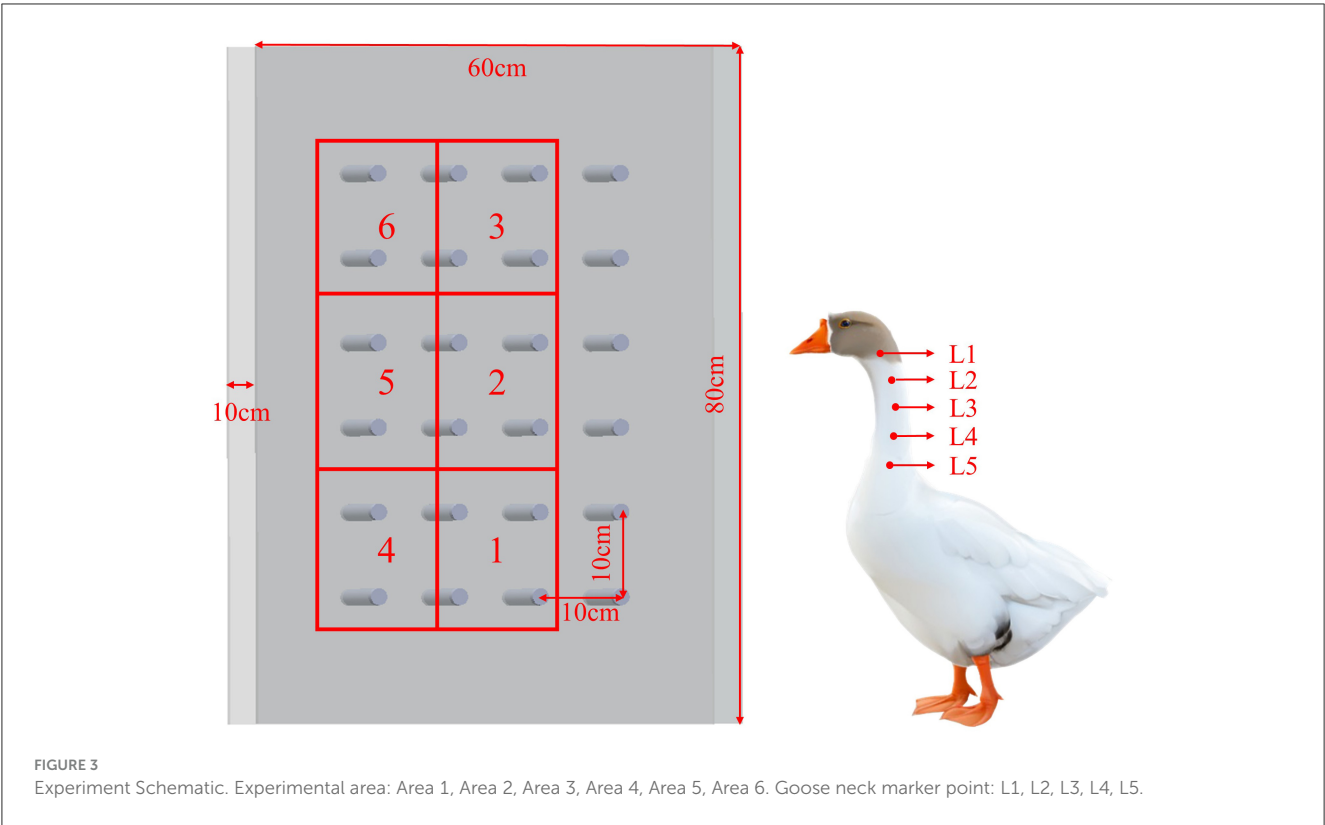


TABLE 2 Data smoothing results.

Norm	Angel	Moving	Lowess	Loess	Sgolay	Rlowess	Rloess
RMSE	θ_1	2.1159	2.5306	1.6042	1.1952	2.4558	1.4159
	θ_2	1.7605	2.0548	0.9801	0.9534	2.0431	1.0528
	θ_3	2.1113	2.6469	1.2942	0.8510	2.7277	1.1098
SNR	θ_1	36.6728	35.1180	39.0771	41.6336	35.3788	40.1615
	θ_2	37.7778	36.4351	42.8647	43.1049	36.4846	42.2434
	θ_3	36.5137	34.5502	40.7647	44.4061	34.2889	42.1000
R^2	θ_1	0.9057	0.8651	0.9458	0.9699	0.8730	0.9577
	θ_2	0.8829	0.8406	0.9637	0.9656	0.8424	0.9581
	θ_3	0.6438	0.4402	0.8661	0.9421	0.4055	0.9016

be processed to obtain the coordinates of each marking point. First, the high-speed camera system was calibrated and a reference coordinate system was established. A two-dimensional reference coordinate system was established according to the fore-and-aft and pitching motion directions of the goose neck. The lower right corner of the image area was taken as the origin of the coordinate system, and the forward and backward motion direction of the goose neck was taken as the X-axis, and the pitching motion direction of the goose neck was taken as the Y-axis. Among them, the forward movement direction of the goose neck was defined as the positive direction of X-axis, and the upward movement direction of the goose neck was defined as the positive direction of Y-axis, and the reference coordinate system as shown in Figure 4.

YOLOv7-pose recognition algorithm was used to extract the coordinate information of the marker points of the goose neck.

TABLE 3 Experimental area distribution.

Experimental area	Heights/cm	Lengths/cm
Area 1	15–30	20–30
Area 2	30–45	20–30
Area 3	45–60	20–30
Area 4	15–30	30–40
Area 5	30–45	30–40
Area 6	45–60	30–40

Firstly, the *.cine format video data collected through the high-speed camera system was transcoded and converted all into *.mp4



FIGURE 4
Reference coordinate system establishment.

format video data. Second, the *.mp4 format video data were processed by using video processing algorithms to extract all the frame images. Then, in order to construct a complete dataset, frame images under different postures and lighting conditions were selected to ensure that the dataset could cover various postures and lighting conditions. Finally, the labeling tool Labelme software was used to label the selected frame images, including the whole goose neck and goose neck key point information, and generated *.json type labeling files. The YOLOv7-pose network structure is shown in Figure 5.

The 2:1 ratio was used to divide the training set (3,685 images) and test set (1,806 images). Among them, the training set was used for model training and parameter optimization, and the test set was used to evaluate the model performance. In order to augment the dataset, increase the difficulty of model training and improve the model recognition accuracy, physical transformation operations were performed on the divided dataset using the following four augmentation methods; random rotation, Gaussian noise, horizontal flip, and luminance change.

The YOLOv7-pose recognition network was used, and the software platform was Pycharm 2023.2.1+CUDA 12.1+Python 3.10+torch 2.1.0. In terms of the parameter settings for the YOLOv7-pose pose recognition network, the size of the training image was set to 640×640 pixel, the training batch size was set to 16 and the epoch was set to 300. The optimal weights would be automatically saved after each test.

The YOLOv7-pose network was used for 300 epoch, and the AP reached 99.6% after training, with excellent training results.

Based on the YOLOv7-pose pose recognition model, goose neck pose recognition was performed on all frames extracted from the high-speed camera video, as shown in Figure 6 and Table 4. The recognition results showed that the goose neck as a whole as well as the five marking points of the neck could be accurately recognized under different poses and different lighting conditions, which can be used for the analysis of the goose neck movement trajectory.

3 Results

3.1 Goose neck motion joint angle analysis

The joint angles of the goose neck moving at different heights and distances in a narrow space were analyzed to obtain the maximum, minimum and range of the rotation angles of each joint of the goose neck, as shown in Table 5.

From the Table 5, the maximum values and rotation ranges of each joint angle of the goose neck were analyzed. The maximum rotation angle of the anterior segment θ_1 was 179.89° , the minimum rotation angle was 88.38° , and the maximum rotation range was 91.51° . The maximum rotation angle of the middle segment θ_2 was 178.51° , the minimum rotation angle was 96.15° , and the maximum rotation range was 82.36° . The maximum rotation angle of the posterior θ_3 was 179.66° , the minimum rotation angle was 100.95° , and the maximum rotation range was 78.71° . The maximum rotation angle of the end segment θ_3

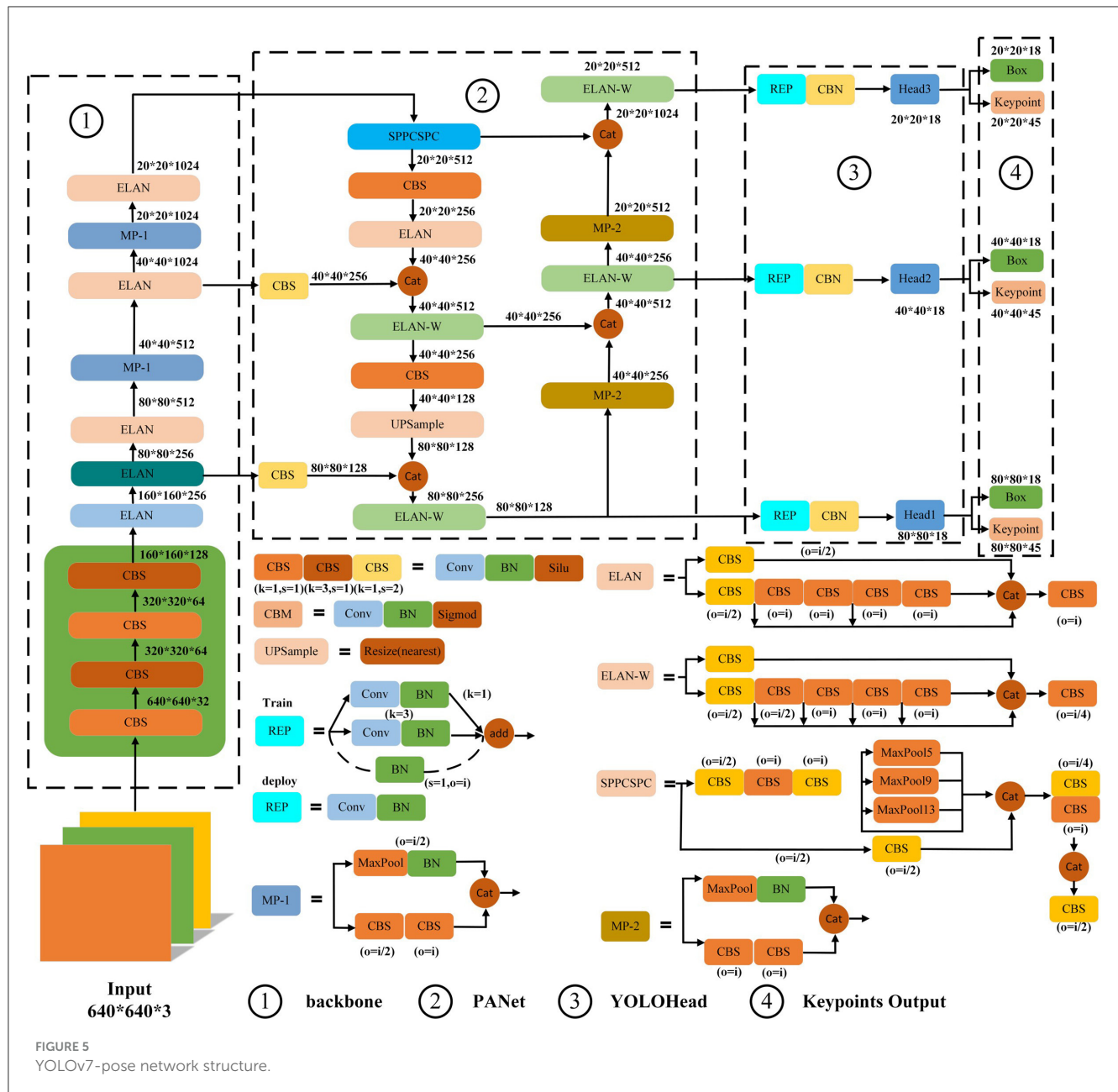


FIGURE 5
YOLOv7-pose network structure.

was 179.66°, the minimum rotation angle was 100.95°, and the maximum rotation range was 78.71°.

Origin software was used to analyze and process the test data of the joint angles. The rotation angles between different joints of the goose neck in each area of the narrow space are plotted in Figure 7.

As shown in Figure 7a, when the goose neck moved toward the target Area 1, the anterior segment joint angle θ_1 firstly decreased about 10°, then gradually increased to about 180°, and finally decreased 15°; the middle segment joint angle θ_2 firstly decreased about 20°, and then gradually increased to about 180°; and the posterior segment joint angle θ_3 firstly increased about 10°, and then decreased about 40°, and finally increased to about 170°.

As shown in Figure 7b, when the goose neck moved toward the target Area 2, the anterior segment joint angle θ_1 firstly decreased by about 10° and then increased to about 170°; the middle segment

joint angle θ_2 of firstly decreased by about 30° and then increased to about 175°; and the posterior segment joint angle θ_3 of firstly decreased by about 20° and then increased to about 165°.

As shown in Figure 7c, when the goose neck moved toward the target Area 3, the anterior segment joint angle θ_1 firstly decreased by about 10° and then increased to about 170°; the middle segment joint angle θ_2 firstly decreased by about 10° and then increased to about 174°; and the posterior segment joint angle θ_3 firstly decreased by about 5° and then increased to about 178°.

As shown in Figure 7d, when the goose neck moved toward the target Area 4, the anterior segment joint angle θ_1 firstly decreased by about 40° and then increased to about 180°; the middle segment joint angle θ_2 firstly decreased by about 40° and then increased to about 170°; and the posterior segment joint angle θ_3 firstly decreased by about 20° and then increased to about 160°.

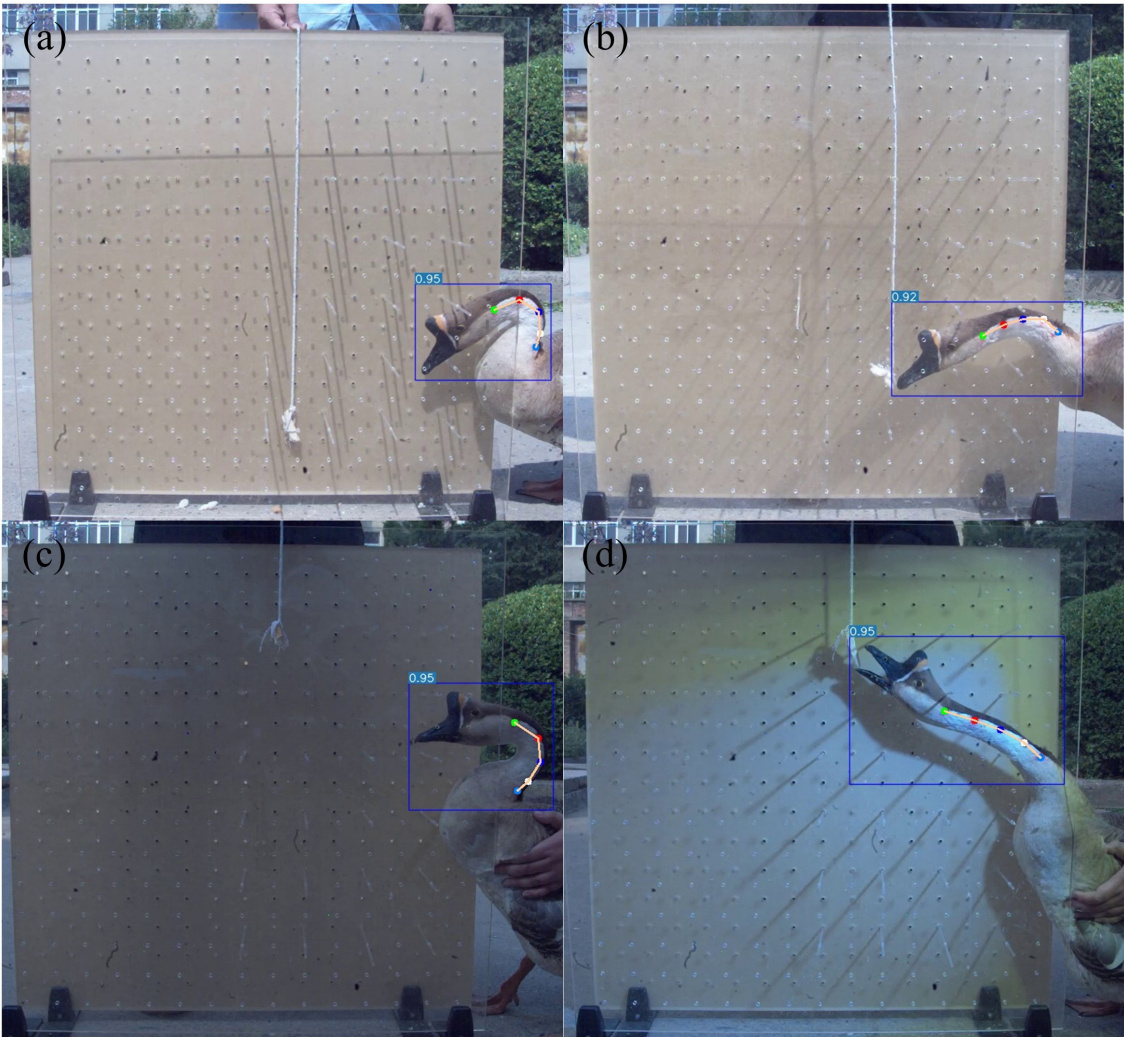


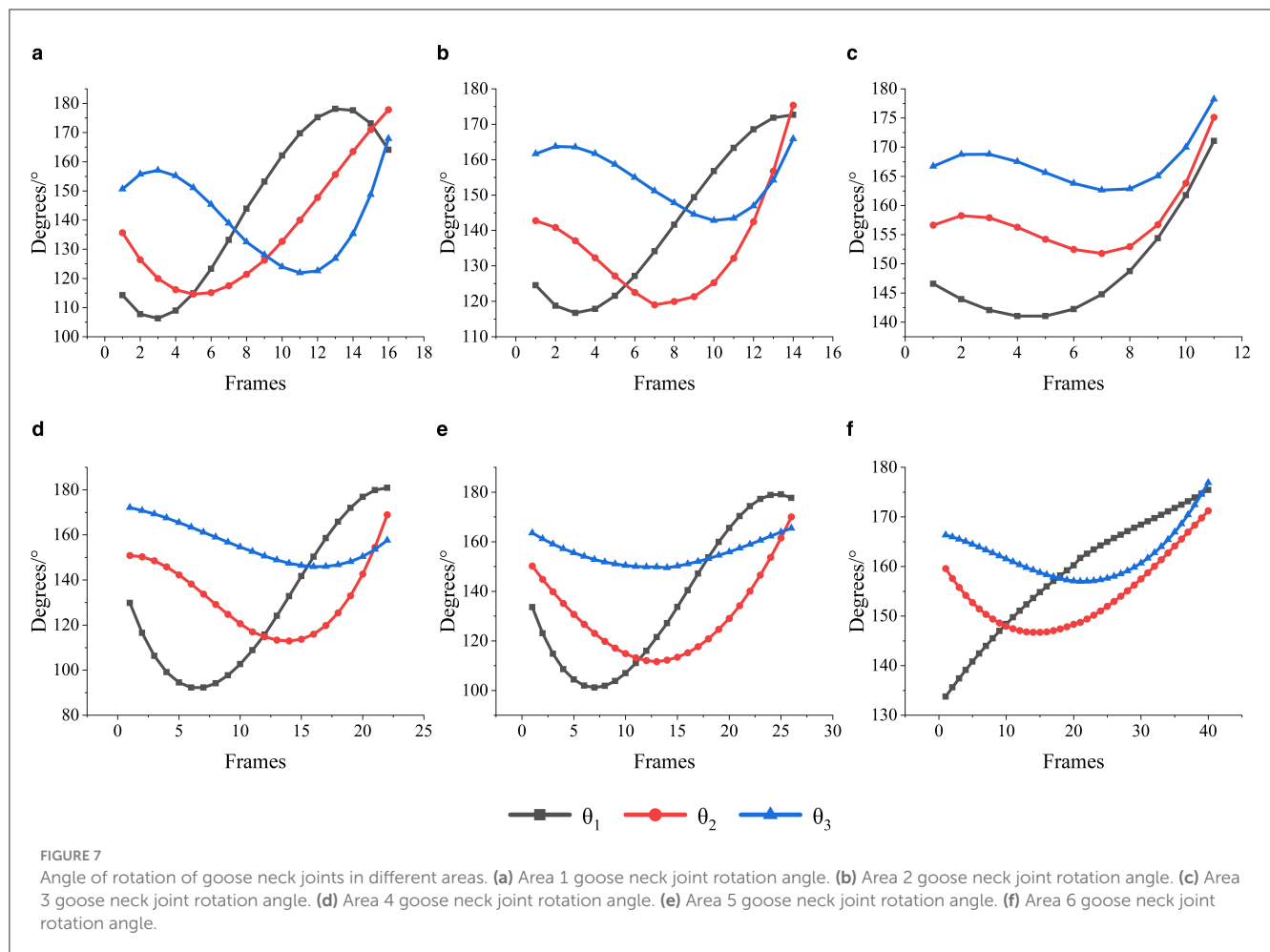
FIGURE 6 YOLOv7-pose recognition results. (a) Recognition in goose neck bent pose. (b) Recognition in goose neck extended pose. (c) Recognition in low light condition. (d) Recognition in bright light condition.

TABLE 4 Network inspection performance results.

Precision/%	Recall/%	F1 score/%	Average precision/%	Detection speed/FPS	Weight size/MB
99.4	95.1	98	99.6	36.7	153

TABLE 5 Goose neck motion joint angles.

	L1–L2–L3 (θ_1)			L2–L3–L4 (θ_2)			L3–L4–L5 (θ_3)		
	MAX/°	MIN/°	VAR/°	MAX/°	MIN/°	VAR/°	MAX/°	MIN/°	VAR/°
1	179.82	89.13	90.69	178.51	96.15	82.36	168.55	100.95	67.60
2	178.99	104.76	74.23	165.88	105.25	60.63	165.96	113.75	52.21
3	179.39	115.79	63.60	178.15	121.87	56.28	179.66	136.65	43.01
4	178.85	88.56	90.29	176.14	102.78	73.36	175.22	113.20	62.02
5	179.89	88.38	91.51	170.62	98.63	71.99	173.33	114.78	58.55
6	179.86	109.39	70.47	176.74	98.43	78.31	173.50	128.99	44.51



As shown in Figure 7e, when the goose neck moved toward the target Area 5, the anterior segment joint angle θ_1 firstly decreased by about 55° and then increased to about 180° ; the middle segment joint angle θ_2 firstly decreased by about 40° and then increased to about 170° ; and the posterior segment joint angle θ_3 firstly decreased by about 10° and then increased to about 168° .

As shown in Figure 7f, when the goose neck moved toward the target Area 6, the anterior segment joint angle θ_1 continued to increase to about 174° ; the middle segment joint angle θ_2 firstly decreased by about 15° , and then increased to about 170° ; and the posterior segment joint angle θ_3 firstly decreased by about 8° , and then increased to about 176° .

3.2 Goose neck motion trajectory analysis

To further analyze the motion of the goose neck in the narrow space, the YOLOv7-pose pose recognition algorithm was used to extract the position information of the marker points L1, L2, L3, L4, and L5, and the goose neck motion trajectory is plotted in Figure 8.

As shown in Figure 8a, when the target position was in Area 1, the overall goose neck was firstly moving downward and then forward. During the forward movement of the goose neck, the marker points L1, L2, and L3 of the anterior-middle segment moved further downward, and the movement heights

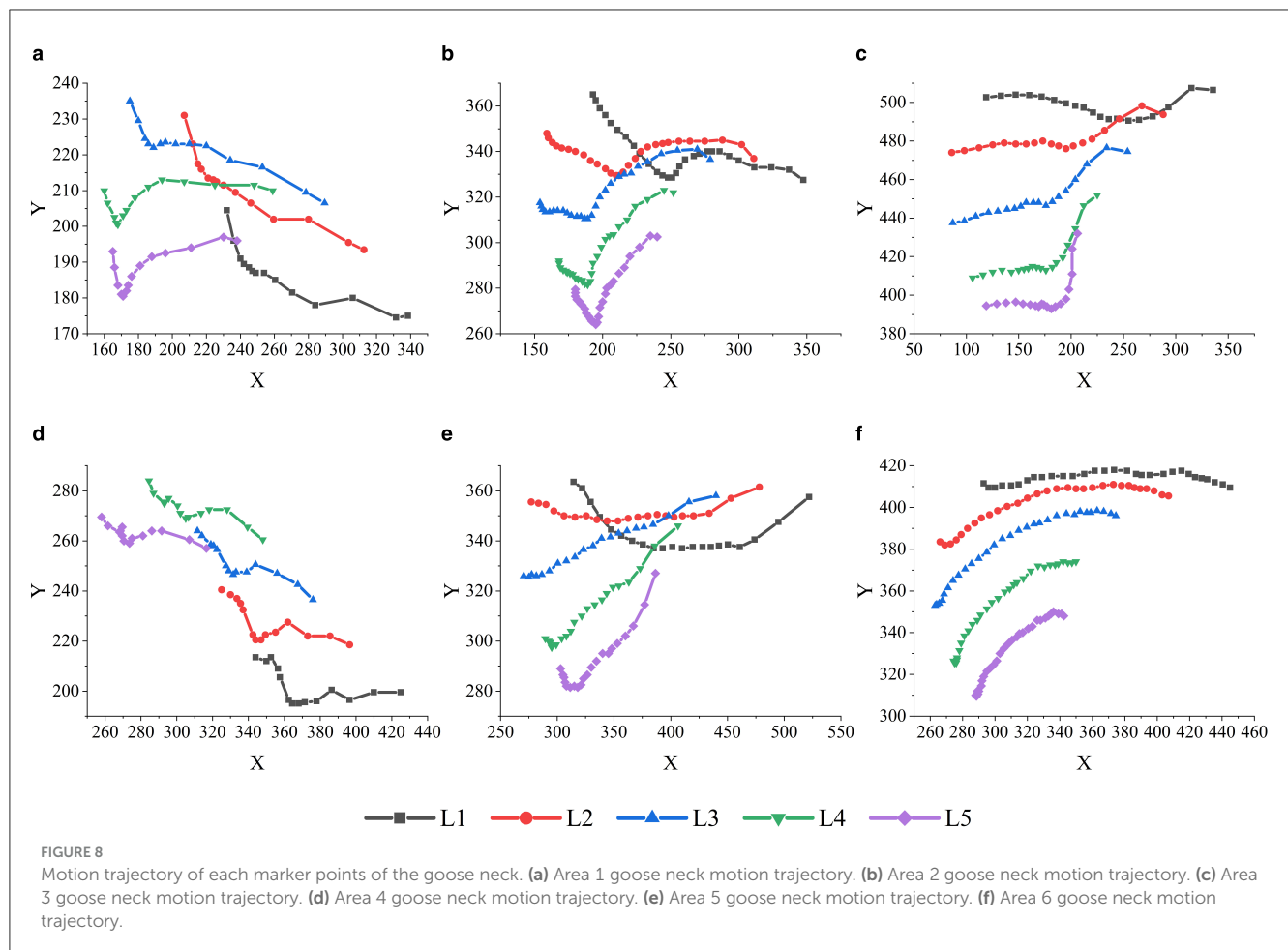
of the marker points L4 and L5 of the posterior segment were basically unchanged.

As shown in Figure 8b, when the target position was in Area 2, the overall goose neck was firstly moved downward, then adjusted upward, and finally moved forward. During the forward movement of the goose neck, the movement heights of the marker points L1, L2, and L3 in the anterior-middle segment were basically unchanged, and the marker points L4 and L5 in the posterior segment moved further upward.

As shown in Figure 8c, when the target position was in Area 3, the goose neck as a whole moved forward first, followed by a compound motion in the forward and upward directions. During the forward motion of the goose neck, the motion height of each marker point was basically unchanged; when the goose neck carried out the composite motion in the forward and upward directions, the forward motion magnitudes of each marker point decreased in turn, and the upward motion magnitudes increased in turn.

As shown in Figure 8d, when the target position was in Area 4, the overall goose neck moved downward first, and then forward. During the forward movement of the goose neck, the movement heights of the marker points L1 and L2 in the anterior-middle segment were basically unchanged, and the marker points L3, L4, and L5 in the middle-posterior segment moved further downward.

As shown in Figure 8e, when the target position was in Area 5, the overall goose neck firstly moved downward, then moved



upward previously, and finally carried out a compound motion in the forward and upward directions. During the forward movement of the goose neck, the movement height of the marker points L1 and L2 of the anterior segment was basically unchanged, and the marker points L3, L4, and L5 of the middle-posterior segment further moved upward; when the goose neck carried out the composite movement in the forward and upward directions, the amplitude of the forward movement of each marker point decreased in turn, and the amplitude of the upward movement increased in turn.

As shown in Figure 8f, when the target position was in Area 6, the overall movement of the goose neck was upward first, followed by forward movement. During the forward movement of the goose neck, anterior segment marker point L1 movement height was basically unchanged, middle-posterior segment marker points L2, L3, L4, L5 forward movement amplitude decreased in turn, and upward movement amplitude increased in turn.

4 Discussion

From the most values of the angles of the goose neck joint angles θ_1 , θ_2 , and θ_3 , it was found that the maximum rotation angles of the goose neck anterior segment joint angle θ_1 , the middle segment joint angle θ_2 and the posterior segment joint angle θ_3 are similar. Both the maximum and the minimum rotation angles decreased in order. It indicated that when the goose neck moved

in a narrow space, the anterior segment exhibited the greatest adjustment capability, followed by the middle segment, with the posterior segment demonstrating relatively weaker adjustment ability. It could be seen that the adjustment effect of the anterior segment was stronger than that of the middle and posterior segments of the goose neck, and its flexibility was also better than that of the middle and posterior segments.

From the goose neck motion joint angle analysis, the overall trend of change for each joint angle was quite similar. Primarily, the angles of goose neck anterior joint θ_1 , middle joint θ_2 , and posterior joint θ_3 reach their respective troughs sequentially before reaching to peak values. Following the trough of angle θ_1 , the rate of change of angle θ_2 begins to decline. Similarly, after the trough of angle θ_2 , the rate of change of angle θ_3 starts to decrease. Subsequently, after the trough of angle θ_3 , the rate of change of angle θ_1 begins to decrease, while the rates of change for angles θ_2 and θ_3 start to rise.

From the goose neck motion trajectory analysis, the motion trajectories of the marker points were similar, the motion trend was basically the same, and the motion of the joints exhibited obvious transmissibility. In particular, the transverse motion amplitude of the goose neck key points L1, L2, L3, L4, and L5 decreased in turn, and the longitudinal motion amplitude increased in turn. When the target position was in Area 1 and Area 4, the overall goose neck was firstly moving downward and then forward. During the forward movement of the goose neck, the marker points L1, L2, and L3 of the anterior-middle segment moved further

downward, and the movement heights of the marker points L4 and L5 of the posterior segment were basically unchanged. When the target position is in Area 2 and Area 5, the overall goose neck was firstly moved downward, then adjusted upward, and finally moved forward. During the forward movement of the goose neck, the movement heights of the marker points L1, L2, and L3 in the anterior-middle segment were basically unchanged, and the marker points L4 and L5 in the posterior segment moved further upward. When the target position was in Area 3 and Area 6, anterior segment marker point L1 movement height was basically unchanged, middle-posterior segment marker points L2, L3, L4, L5 forward movement amplitude decreased in turn, and upward movement amplitude increased in turn.

To summarize, the goose neck movement had continuity and transmissibility, and the overall posture of the goose neck was adjusted through the synergistic movements of the anterior, middle, and posterior segments to complete the movement toward the target positions. Among them, the anterior segment mostly carried out transverse movement, with the significant horizontal adjustment; the middle segment combined transverse and longitudinal movement, with similar movement ability in all directions; and the posterior segment mostly carried out longitudinal movement, with a significant height adjustment.

This research investigated the muscle distribution characteristics in sagittal plane, coronal plane and transverse plane, researched the joint rotation characteristics and partition motion law, and revealed the mechanism of goose neck motion in narrow space. The next step will construct obstacle avoidance posture curve and analysis based on goose neck motion posture in narrow space to establish the bionic obstacle avoidance trajectory of robotic arms in narrow space and experimentally verify the reasonableness and superiority of the trajectory design. At present, the mechanism of the synergistic motion of the anterior, middle, and posterior segments of the goose neck based on the overall motion postures of the goose neck was analyzed only combined the high-speed camera motion data of the goose neck. The goose neck is a rigid-flexible coupled complex structure as a whole, and the bone-muscle synergy of the goose neck motion in the narrow space is still needed to be explored, which will provide theoretical support for the combination of robotic arms control strategies and structural design and further improve the flexibility and load capacity of the bionic robotic arms.

5 Conclusions

In this article, the goose neck was used as the test object to research the movement mechanism in different target areas in a narrow space. Firstly, goose neck MRI image data was collected. It was found that there are longitudinal muscles in both sagittal plane and coronal plane connecting the anterior and terminal parts of the goose neck, which controlled the lateral bending and rotational movements of the goose neck. The muscle areas in the transverse plane of the goose neck were different, the anterior segment was dominated by ventral flexion movement, so the ventral muscle area was larger; there was no significant difference in the area of the ventral and dorsal muscles in the middle segment; and the posterior segment was dominated by dorsiflexion movement, so the dorsal muscle area was larger.

Then, the motion data of the goose neck in the narrow space was collected by a high-speed camera system to analyze the goose neck motion:

- 1) The optimum value and the range of variation of the motion joint angle of the goose neck were analyzed, it was found that the anterior segment exhibited the greatest adjustment capability, followed by the middle segment, with the posterior segment demonstrating relatively weaker adjustment ability. It indicated that the regulation of the anterior segment of the goose neck was stronger than that of the middle and posterior segments, and its flexibility was better than that of the middle and posterior segments.
- 2) The joint angles of goose neck movement were analyzed. It was found that the overall trend of change for each joint angle was similar, mostly decreased first and then increased. In terms of the rate of change of the angles, mostly when the anterior segment joint angle θ_1 reached the trough, the rate of change of the middle segment joint angle θ_2 and the posterior segment joint angle θ_3 decreased in turn. When the posterior segment joint angle θ_3 reached the trough, the rate of change of the anterior segment joint angle θ_1 began to decrease, and the rate of change of the middle segment joint angle θ_2 and the posterior segment joint angle θ_3 began to increase.
- 3) The motion trajectory of the goose neck was analyzed. It was found that the motion trajectories of each marker point were similar, the motion trend was basically the same. The motion travels of the marker points L1, L2, L3, L4, and L5 decreased in turn, and the motion of the joints exhibited obvious transmissibility. From the motion process, the overall goose neck was first height-adjusted, and when the height of the head was nearly the same as that of the target position, the horizontal movement was performed, and finally reached the target position.

The research results showed that the goose neck has excellent motion flexibility in narrow space and excellent obstacle avoidance effect. Its muscle distribution characteristics, joint rotation characteristics and partition motion law provided the important theoretical basis for designing the high flexibility and adaptability bionic robotic arm. In addition, the feasibility of the YOLOv7-pose recognition network for goose neck motion pose recognition was verified to provide a new method for animal behavior research.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The animal study was approved by Henan University of Science and Technology. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

FZ: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. HS: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft. JW: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. XW: Investigation, Validation, Writing – original draft. YQ: Investigation, Validation, Writing – original draft. XC: Formal analysis, Writing – original draft. SA: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China (Grant Nos. 51905155 and 52075149), Frontier Exploration Projects of Longmen Laboratory (Grant No. LMQYTSKT032),

Colleges and Universities of Henan Province Youth Backbone Teacher Training Program (Grant No. 2017GGJS062), and Postgraduate Education Reform Project of Henan Province (2023SJGLX180Y).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Luciani B, Braghin F, Pedrocchi ALG, Gandolla M. Technology acceptance model for exoskeletons for rehabilitation of the upper limbs from therapists' perspectives. *Sensors*. (2023) 23:1721. doi: 10.3390/s23031721
- Noronha B, Ng CY, Little K, Xiloyannis M, Keong Kuah CW, Wee SK, et al. Soft, lightweight wearable robots to support the upper limb in activities of daily living: a feasibility study on chronic stroke patients. *IEEE Trans Neural Syst Rehabil Eng*. (2022) 30:1401–11. doi: 10.1109/TNSRE.2022.3175224
- Pi J, Liu J, Zhou K, Qian M. An octopus-inspired bionic flexible gripper for apple grasping. *Agriculture*. (2021) 11:1014. doi: 10.3390/agriculture11101014
- Gupta U, Qin L, Wang Y, Godaba H, Zhu J. Soft robots based on dielectric elastomer actuators: a review. *Smart Mater Struct*. (2019) 28:103002. doi: 10.1088/1361-665X/ab3a77
- Ghalati N, Hadi M, Akbari A, Ghafarirad H, Zareinejad M. Behavior analysis of biomimetic soft bending actuators in free motion and contact. *J Bionic Eng*. (2023) 20:967–81. doi: 10.1007/s42235-022-00322-w
- Tebyani M, Spaeth A, Cramer N, Teodorescu M. A geometric kinematic model for flexible voxel-based robots. *Soft Robot*. (2023) 10:517–26. doi: 10.1089/soro.2021.0139
- Zweers GA, Angtlique HJ, Leeuw VD, Bout RG. Evolutionary morphology of the neck system in ratites, fowl and waterfowl. *Neth J Zool*. (2001) 51:243–62. doi: 10.1163/156854201X00297
- Narita Y, Kuratani S. Evolution of the vertebral formulae in mammals: a perspective on developmental constraints. *J Exp Zool B: Mol Dev Evol*. (2005) 304:91–106. doi: 10.1002/jez.b.21029
- Zweers G, Bout R, Heidweiller J. Motor organization of the avian head-neck system. In: MNO Davies, PR Green, editors. *Perception and Motor Control in Birds: An Ecological Approach*. Cham: Springer (1994), p. 201–21. doi: 10.1007/978-3-642-75869-0_12
- van der Leeuw AHJ, Bout RG, Zweers GA. Control of the cranio-cervical system during feeding in birds. *Am Zool*. (2001) 41:1352–63. doi: 10.1093/icb/41.6.1352
- Furet M. *Analyse cinéto-statique de mécanismes de ténacité: Application à la modélisation de cous d'oiseaux et de manipulateurs bio-inspirés*. Nantes: Diss École Centrale de Nantes (2020).
- Furet M, Abourachid A, Böhmer C, Chummun V, Chevallereau C, Cornette R, et al. Estimating motion between avian vertebrae by contact modeling of joint surfaces. *Comput Methods Biomech Biomed Engin*. (2022) 25:123–31. doi: 10.1080/10255842.2021.1934676
- Krings M, Nyakatura JA, Fischer MS, Wagner H. The cervical spine of the American barn owl (*Tyto furcata pratincola*): I. Anatomy of the vertebrae and regionalization in their S-shaped arrangement. *PLoS ONE*. (2014) 3:e91653. doi: 10.1371/journal.pone.0091653
- Krings M, Nyakatura JA, Boumans MLLM, Fischer MS, Wagner H. Barn owls maximize head rotations by a combination of yawing and rolling in functionally diverse regions of the neck. *J Anat*. (2017) 231:12–22. doi: 10.1111/joa.12616
- Panyutina AA, Kuznetsov AN. Are owls technically capable of making a full head turn? *J Morphol*. (2024) 285:e21669. doi: 10.1002/jmor.21669
- Kambic RE, Biewener AA, Pierce SE. Experimental determination of three-dimensional cervical joint mobility in the avian neck. *Front Zool*. (2017) 14:1–15. doi: 10.1186/s12983-017-0223-z
- Wang J, Jia W, Zhang F, Ma X, Qiu Z, Qian Z, et al. Study on the structural characteristics of bird necks and their static motion features in the sagittal plane. *Coatings*. (2021) 11:1228. doi: 10.3390/coatings11101228
- Wang J, Sun H, Jia W, Zhang F, Qian Z, Cui X, et al. *In vivo* analysis of the dynamic motion stability characteristics of geese's neck. *Biomimetics*. (2022) 7:160. doi: 10.3390/biomimetics7040160
- Böhmer C, Prevotau J, Duriez O, Abourachid A. Gulper, ripper and scrapper: anatomy of the neck in three species of vultures. *J Anat*. (2020) 236:701–23. doi: 10.1111/joa.13129
- Abourachid A, Gagnier B, Furet M, Cornette R, Delapre A, Hackert R, et al. Modeling intervertebral articulation: the rotule à doigt mechanical joint (RAD) in birds and mammals. *J Anat*. (2021) 239:1287–99. doi: 10.1111/joa.13517
- Dzemeski G, Christian A. Flexibility along the neck of the ostrich (*Struthio camelus*) and consequences for the reconstruction of dinosaurs with extreme neck length. *J Morphol*. (2007) 268:701–14. doi: 10.1002/jmor.10542
- Mochiyama H, Gunji M, Niiyama R. Ostrich-inspired soft robotics: a flexible bipedal manipulator for aggressive physical interaction. *J Robot Mechatron*. (2022) 34:212–8. doi: 10.20965/jrm.2022.p0212
- Nakano K, Gunji M, Ikeda M, Or K, Ando M, Inoue K, et al. "RobOstrich" manipulator: a novel mechanical design and control based on the anatomy and behavior of an ostrich neck. *IEEE Robot Autom Lett*. (2023). doi: 10.36227/techrxiv.21358200.v2
- He J, Wen G, Liu J. A class of bionic hyper-redundant robots mimicking the bird's neck. *Acta Mech Sin*. (2023) 39:522351. doi: 10.1007/s10409-022-22351-x



OPEN ACCESS

EDITED BY

Sebastian Meller,
University of Veterinary Medicine Hanover
Foundation, Germany

REVIEWED BY

James Edward Brereton,
Sparsholt College, United Kingdom
Anindya Sen,
Heritage Institute of Technology, India

*CORRESPONDENCE

Fu Zhang
✉ zhangfu30@126.com
Sanling Fu
✉ fusanling@126.com

RECEIVED 21 March 2024

ACCEPTED 10 February 2025

PUBLISHED 13 March 2025

CITATION

Zhang F, Zhao X, Wang S, Qiu Y, Fu S and
Zhang Y (2025) Research on herd sheep facial
recognition based on multi-dimensional
feature information fusion technology in
complex environment.
Front. Vet. Sci. 12:1404564.
doi: 10.3389/fvets.2025.1404564

COPYRIGHT

© 2025 Zhang, Zhao, Wang, Qiu, Fu and
Zhang. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Research on herd sheep facial recognition based on multi-dimensional feature information fusion technology in complex environment

Fu Zhang^{1,2*}, Xiaopeng Zhao¹, Shunqing Wang¹, Yubo Qiu¹,
Sanling Fu^{3*} and Yakun Zhang¹

¹College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang, China, ²Collaborative Innovation Center of Machinery Equipment Advanced Manufacturing of Henan Province, Luoyang, China, ³College of Physical Engineering, Henan University of Science and Technology, Luoyang, China

Intelligent management of large-scale farms necessitates efficient monitoring of individual livestock. To address this need, a three-phase intelligent monitoring system based on deep learning was designed, integrating a multi-part detection network for flock inventory counting, a facial classification model for facial identity recognition, and a facial expression analysis network for health assessment. For multi-part detection network, The YOLOv5s path aggregation network was modified by incorporating a multi-link convolution fusion block (MCFB) to enhance fine-grained feature extraction across objects of different sizes. To improve the detection of dense small targets, a Re-Parameterizable Convolution (RepConv) structure was introduced into the YOLOv5s head. For facial identity recognition, the sixth-stage structure in GhostNet was replaced with a four-layer spatially separable self-attention mechanism (SSSA) to strengthen key feature extraction. Additionally, model compression techniques were applied to optimize the facial expression analysis network for improved efficiency. A transfer learning strategy was employed for weight pre-training, and performance was evaluated using FPS, model weight, mean average precision (mAP), and test set accuracy. Experimental results demonstrated that the enhanced multi-part identification network effectively extracted features from different regions of the sheep flock, achieving an average detection accuracy of 95.84%, with a 2.55% improvement in mAP compared to YOLOv5s. The improved facial classification network achieved a test set accuracy of 98.9%, surpassing GhostNet by 3.1%. Additionally, the facial expression analysis network attained a test set accuracy of 99.2%, representing a 3.6% increase compared to EfficientNet. The proposed system significantly enhances the accuracy and efficiency of sheep flock monitoring by integrating advanced feature extraction and model optimization techniques. The improvements in facial classification and expression analysis further enable real-time health monitoring, contributing to intelligent livestock management.

KEYWORDS

YOLOv5, MCFB, SSSA, compression of model, transfer learning

1 Introduction

Sheep inventory counting, facial recognition, and health analysis are crucial components of daily management on large-scale farms. Accurate counting of sheep in a flock allows for the development of effective breeding plans, which align with animal welfare standards while also reducing farm costs (1). Facial recognition enables precise identification of individual sheep, supporting better tracking and management. The information can be used to tailor feeding plans for each sheep, promoting precise breeding and scientific farm management (2). Additionally, facial expression analysis helps assess the health status of sheep, enabling timely treatment of sick or injured animals and minimizing the risk of disease spread (3). As such, sheep inventory counting, facial recognition, and health analysis hold significant potential for application in precision sheep farming, making them essential tasks for improving farm efficiency and animal welfare (4–6).

Livestock identification methods are generally classified into contact and non-contact types. Contact methods are traditional approaches, including ear markings, ear tags, and radio frequency identification (RFID) (7, 8). Non-contact recognition, on the other hand, typically involves identifying livestock based on physiological characteristics, such as iris and retinal blood vessels (9, 10). Traditional contact methods are limited by distance, and improper installation can cause stress in animals or harm to personnel. Additionally, ear tags and similar methods require manual registration, which is time-consuming and prone to errors. Non-contact recognition, currently reliant on combining retinal and iris data with traditional machine learning, is primarily used for identifying individual animals. However, it does not meet the requirements for identifying sheep or livestock in complex environments (11, 12). Moreover, non-contact methods often involve complicated data collection processes, which are not easily cooperative with livestock, limiting their practical applicability.

In recent years, with the advancement of convolutional neural networks (CNNs), face recognition technology has become increasingly mature. However, studies on sheep recognition remain limited (13, 14). Drawing inspiration from face recognition, several studies have explored the use of CNNs to identify livestock using various biometric features (15–18). Facial recognition technology offers advantages such as being natural, intuitive, and non-contact, eliminating the need for livestock to cooperate with fixed gestures. Additionally, face recognition systems are known for their strong anti-interference capabilities and good scalability. As a result, contactless recognition using visual biometric features has become a promising trend for individual livestock identification. For instance, Xie et al. (19) developed an improved DenseNet-CBAM model for pig face recognition, integrating the Convolutional Block Attention Module (CBAM), which enhanced recognition performance. Wang et al. (20) proposed a multi-scale convolutional neural network-based model for contactless individual pig detection in complex environments, achieving an accuracy of 92%. Li et al. (21) introduced a CNN-based method for detecting pig face feature points, addressing the challenge of accurate feature point detection in livestock recognition. Wang et al. (22) implemented a lightweight pig face recognition model based on a deep convolutional neural network, which demonstrated a high recognition rate in complex environments. Yang et al. (23) applied

a YOLOv4-based target detection network that incorporated coordinate information to accurately identify individual cows, with an average recognition accuracy of 93.4%. Traditional cattle identification methods require external tools, which pose safety risks for breeders and may cause physical harm to the cattle. To overcome this, Zhu et al. (24) proposed a bovine face biometric feature extraction method based on image analysis, achieving an accuracy of 95.1%.

The studies mentioned above primarily focus on facial recognition research for pigs and cattle. In contrast, research on sheep facial recognition and expression analysis has made significant progress in the following studies. Xu et al. (25) fine-tuned seven different pre-trained classification network models through transfer learning to assess the effectiveness of existing target classification networks for sheep face recognition, achieving an average detection accuracy of 99.8%. Zhang et al. (26) proposed an improved MobileFaceNet network for sheep face recognition, which achieved 97.91% accuracy in recognizing sheep with small differences, at long distances, and under conditions of low recognition accuracy. Song et al. (27) applied the YOLOv3 recognition network for sheep face recognition, achieving fast and accurate results. Billah et al. (28) introduced a deep learning-based goat facial recognition method, which demonstrated good accuracy, with a recognition rate of 96.4%. For facial expression detection, Han et al. (29) proposed an improved STVGNet-based algorithm for detecting sheep pain expressions, which achieved an accuracy of 96.06%, addressing the challenges of high experience requirements, low recognition accuracy, high costs, and delays in disease treatment in manual recognition of sheep pain. Noor et al. (30) used transfer learning to fine-tune an existing classification network, achieving effective detection of painful expressions in sheep with accuracies of 96.69% in the validation set and 100% in the test set.

Existing livestock recognition technologies do not integrate individual counting or facial expression analysis functions, which are essential for accurate feeding and real-time health monitoring of individual sheep. To address these limitations, a three-stage system for group sheep recognition and individual multi-part classification is proposed. The system consists of three main steps: multi-part recognition of flock sheep, facial classification, and facial expression analysis. A multi-part recognition network is used to achieve flock counting by targeting the back of the sheep, facial classification is employed for individual sheep identification, and a facial expression analysis network is utilized to assess the health status of the sheep.

2 Methods

2.1 Production of datasets

2.1.1 Experimental data sources

The video footage of sheep faces used was captured in the standardized indoor sheep farm environment of Luoyang Xiangshun Agricultural and Animal Husbandry Technology Company, located in Luoyang City, Henan Province, China. The test subjects consisted of 30 adult sheep in four different states: nulliparous, fertilized, pregnant, and postpartum. A Canon camera was used for semi-monthly tracking of the sheep in a herd environment, with each video session lasting no less than 30 min. The frame rate was set to 30 fps, and the shooting scenes are shown in Figure 1.

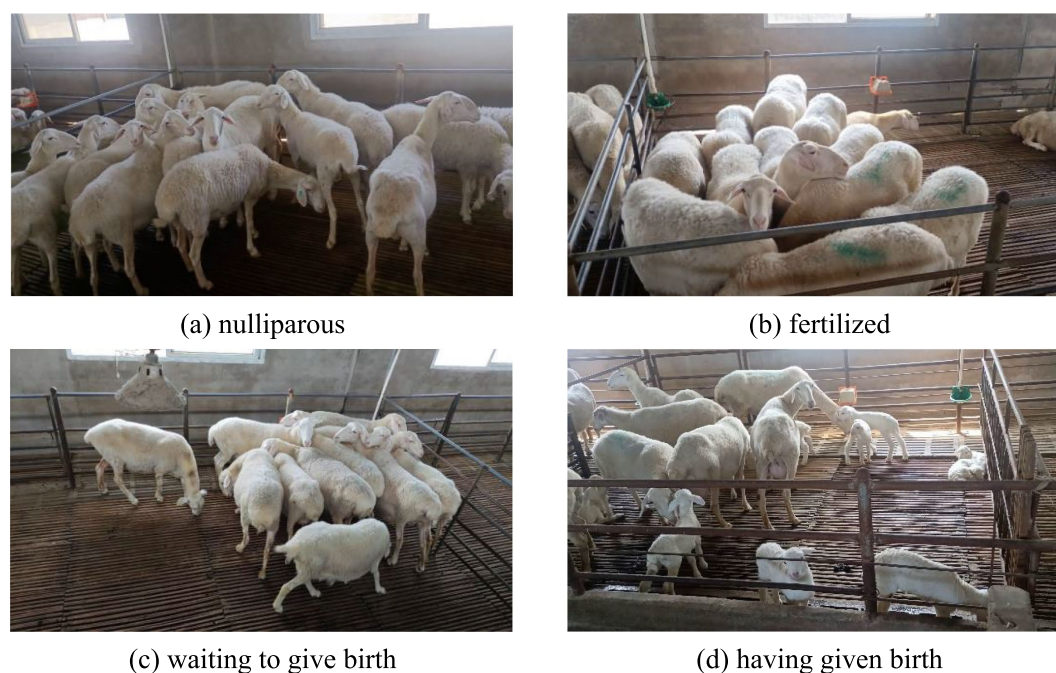


FIGURE 1

Schematic diagram of the shooting scene. (A) Nulliparous. (B) Fertilized. (C) Waiting to give birth. (D) Having given birth.

2.1.2 Data similarity processing

The videos were captured at a rate of 25 frames per second, and clear, effective images were selected as sample data. To further filter out similar images and avoid overfitting of the model due to similarity, the perceptual hash algorithm (d-Hash) was used to eliminate the similar images. Both images were resized to $8 \times 9 \times 3$ pixels and then converted to grayscale. The pixel values of each row were compared sequentially: if the former value was larger than the latter, a difference value of 1 was assigned; otherwise, 0 was assigned. This process generated two binary 8×8 difference matrices. The four characters before and after each row of the difference matrix were converted to hexadecimal, forming two 16-character hash strings. The Hamming distance between the images was then calculated using the XOR method. Only images with a Hamming distance greater than or equal to 12 were retained to eliminate interference from similar images. As a result, a total of 3,078 effective and clear images were selected.

2.1.3 Data amplification and annotation

The sheep face recognition network consisted of three parts: a multi-site recognition network for individual flock sheep, a facial classification network, and a facial expression classification network. Three separate datasets—Dataset A, Dataset B, and Dataset K—were used to train the networks, with the overall processing flow shown in Figure 2.

Dataset A contained 3,078 pre-screened images. The annotation tool LabelImage was used to annotate the images according to the Pascal VOC format, generating an annotation file in .xml format. Dataset A was then split in a 2:1 ratio. Dataset B consisted of 30 sheep, each with 90 face images, totaling 2,700 images. This dataset was divided in a 2:1 ratio between the training and test sets. Dataset K included two categories of images: healthy and sick sheep, with specific differences between the two categories shown in Figure 3. The

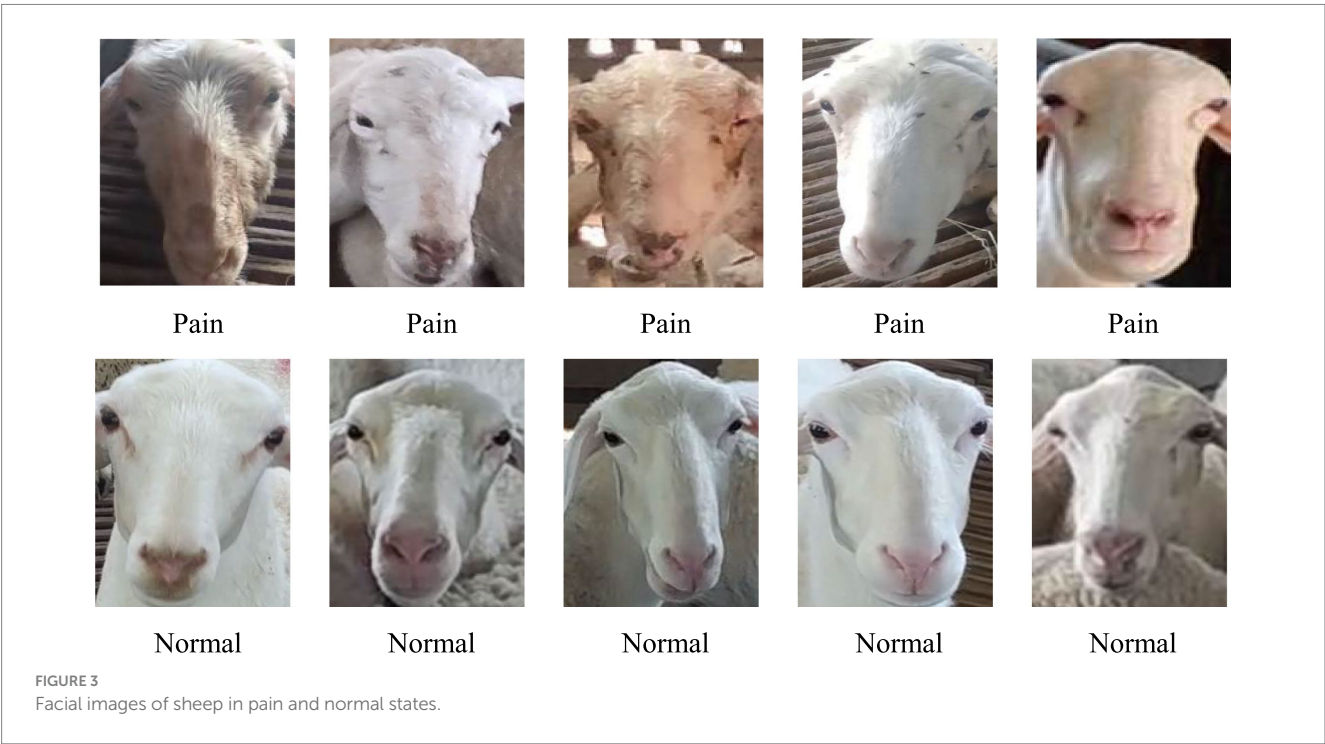
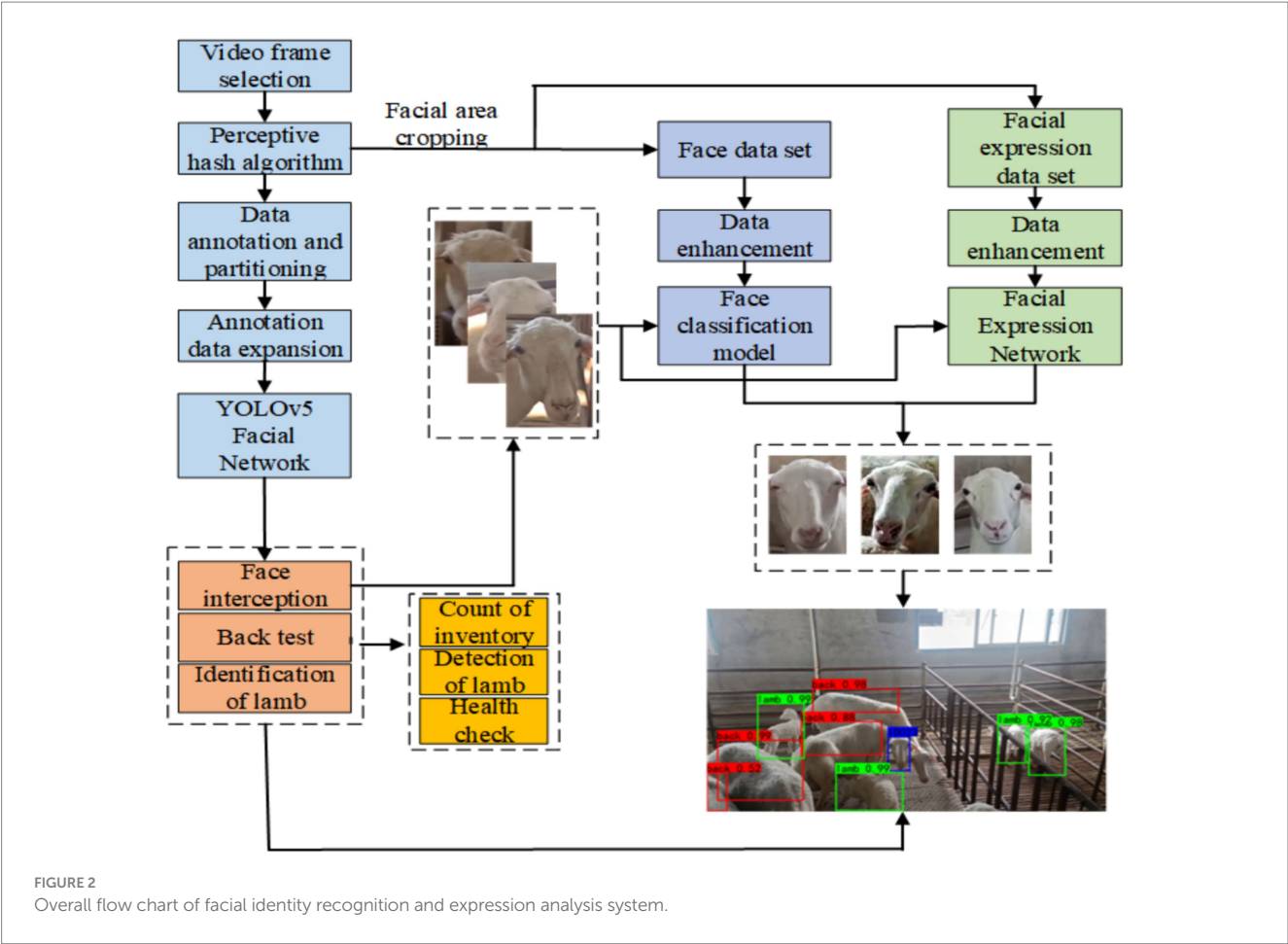
healthy images were taken from 15 adult sheep in Dataset B, while the sick images were sourced from an additional five sheep identified as diseased by a local veterinarian.

To enhance the generalization performance of the facial identity and expression analysis system, Datasets A, B, and K were augmented after splitting. Dataset A was augmented using four methods: random rotation (-15° to 15°), horizontal flip, MixUp, and brightness change. The corresponding annotation files were transformed accordingly, generating Training Set C (10,260 images) and Test Set D (5,130 images). For Dataset B, augmentation was performed using random rotation and brightness change, resulting in Training Set E (5,400 frames) and Test Set F (2,700 frames). Dataset K was augmented in the same way as Dataset B, creating Training Set M (2,400 frames) and Test Set N (1,200 frames).

2.2 Multi-part identification network for individual flock sheep

2.2.1 Improved face detection model for YOLOv5 sheep

Face and back recognition were used to identify individual sheep and lambs, enabling face interception and sheep counting functions. YOLOv5s was chosen for target detection due to its small model size, fast detection speed, and ease of deployment on mobile devices. YOLOv5s was applied to detect faces, backs, and lamb targets. However, the large number of sheep often led to issues such as stacking and partial occlusion in the large-scale breeding environment, posing significant challenges for the YOLOv5s sheep individual multi-part recognition network. Additionally, the sheep's body texture features were relatively uniform, with most being pure white, which made it difficult for the YOLOv5s network to distinguish between different



parts of the sheep. To address these challenges, improvements were made to the neck and head networks of YOLOv5s to enhance detection performance in obstructed environments and improve the accuracy of sheep individual counting.

A depth factor of $d = 0.33$ was applied to the YOLOv5s backbone network to scale down the main architecture of CSPDarkNet53, reducing the number of model parameters and creating a lighter network. A width factor of $w = 0.5$ was used to decrease the input image size, reducing computational effort and improving detection speed for the multi-part recognition network of individual flock sheep. However, this adjustment also affected the detection performance for small, medium, and large targets, making it less suitable for multi-target detection in complex group sheep environments. Due to variations in sheep body shapes, occlusion, and the different sizes of individuals, especially in the farrowing room where lambs are often obscured by ewes, the model faced significant challenges in detecting sheep in such conditions.

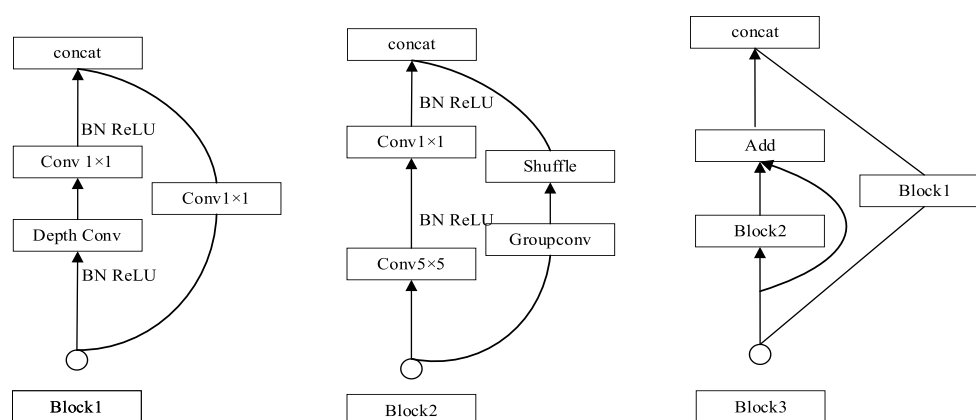
To address these issues in complex breeding environments, it was essential to enhance the target detection algorithm's ability to screen fine-grained features across multiple scales, ensuring effective detection of both large and small objects. A multi-link convolution feature fusion structure was proposed, shown in Figure 4, capable of refining features for small, medium, and large targets. The input feature map was divided into multiple sub-blocks, improving the detection of fine features. Convolution kernels of sizes 1, 3, and 5, along with various convolution types (separable, normal, and group convolution), were applied in specific blocks to enhance feature

fusion. The introduction of multi-link convolution reduced the feature fusion path in YOLOv5s, improving the model's ability to extract higher-order semantic features of sheep while preserving shallow coordinate information. This change allowed for better feature transfer and enhanced recognition of herd sheep. Additionally, multi-link convolution reduced the network model's parameters by dividing the feature map into smaller blocks, thereby improving detection accuracy and speed for each sheep and lamb part.

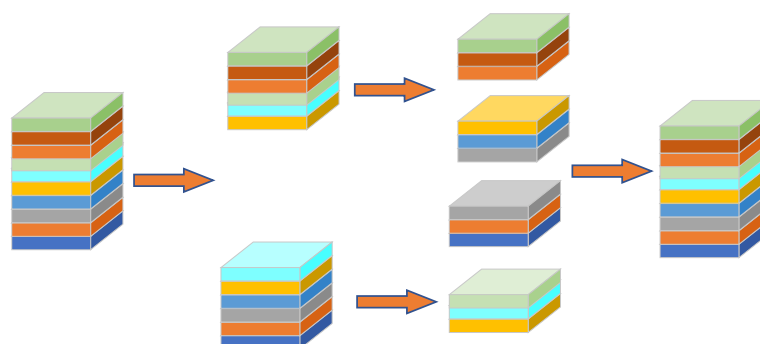
The Repconv convolution structure was introduced before the head network to improve the detection effect of different sizes of the group sheep's face, back and young under complex environment. Repconv contained two different convolution structures of 1×1 and 3×3 . 1×1 convolution can further enhance the attention to small target features, and 3×3 convolution kernel can enhance the information fusion of high and low order features. Compared with a single common convolution structure, Repconv had a more diversified screening effect on target details to improve the detection effect on targets of different sizes.

The multi-part recognition network of the improved YOLOv5s sheep is shown in Figure 5. It was easy to see that the network was divided into three parts: ① the CSPDarkNet53 trunk network structure after the same scale; ② the improved path aggregation network (PANet) and ③ the Head network structure.

Note: CSPn_X stands for Cross stage partial structure, Conv stands for convolutional, BN stands for Batch Norm, CBL stands for Conv+Batch BN + SiLU activation function synthesis module, ResUnit stands for the residual connection module, Concat stands for the feature concatenation



(a) Schematic diagram of multi-link convolution structure



(b) Spatial diagram of multi-link convolution structure

FIGURE 4

Multi-link convolution blocks. (A) Schematic diagram of multi-link convolution structure. (B) Spatial diagram of multi-link convolution structure.



minimum closure region between the prediction frame and the true frame; α is used to measure the consistency parameter between the predicted frame and the real frame, and v represents a trade-off parameter as shown in Equation 3.

The anchor frame sizes in the YOLOv5s target detection network are based on clustering from the COCO dataset, which includes 80 categories. Since the anchor frame sizes vary across categories, they cannot be directly applied to sheep face detection. The multi-part recognition network for sheep in this work primarily identifies the face, back, and lamb, each with different shapes and sizes. To improve detection across large, medium, and small targets, the k-means clustering algorithm was used to re-cluster the sizes in the annotation files. This process resulted in 9 anchor frame sizes with different length-to-width ratios: (29, 68), (67, 38), (50, 77), (46, 117), (78, 73), (104, 56), (74, 114), (79, 187), and (142, 115).

YOLOV5s target loss function consists of four parts: positive sample coordinate loss, positive sample confidence loss, negative sample confidence loss and positive sample classification loss. The loss function is calculated as shown in Equation 1. Where λ_{coord} and λ_{noobj} respectively represent positive sample weight coefficients and

$$\begin{aligned}
Loss1 = & \lambda_{coord} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} (2 - w_i \times h_i) (1 - CIOU) \\
& - \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} \left[\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - C_i) \right] \\
& - \lambda_{noobj} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{noobj} \\
& \left[\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - C_i) \right] - \\
& \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} \sum_{c \in classes} \left[\hat{p}_i \log(p_i(c)) + (1 - p_i(c)) \right] p
\end{aligned} \quad (1)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (3)$$

Aiming at the phenomenon of feature redundancy in feature extraction networks, GhostNet proposed a lightweight network

model, which used linear operations instead of partial convolution to generate a large number of redundant feature graphs to reduce the amount of network computation and improve the speed of the model. GhostNet network was composed of a series of Ghost stream modules stacked. Ghost vector was mainly composed of two structures, separable convolution structure (DWConv) and Ghost module, to achieve the effect of reducing the number of model parameters and improving the detection rate of the model.

2.3.1 Improved GhostNet facial distribution network

Although GhostNet feature network generated efficient feature graphs through simple linear feature mapping, its separable convolution structure reduced the information interaction between different feature graphs and may neglect the extraction of key features. In view of the similar facial texture of sheep, small differences in intra-class features, and difficult to distinguish fine grained features, the four-layer spatially separable self-attention mechanism (SSSA) was adopted to effectively replace the sixth stage network structure in GhostNet, so as to improve the extraction of important features in the facial region. The SSSA consisted of the local self-attention mechanism (LSA) and the global subsampled attention mechanism (GSA), the structure is shown in Figure 6. The LSA module captures fine-grained features and short-range information with a window size of 7×7 , while the GSA processes long-distance and global information, enabling fusion of local features and information exchange between different regions. The fine features of sheep face can be screened and the detection effect of facial classification network can be improved using the alternating link of LSA and GSA.

2.3.2 Facial classification network loss function

CrossEntropy Loss function was used for regression training of sheep face classification network in the improved face classification network, as shown in Equation 4, y_j represents the unique thermal coding form corresponding to the real category, and o_j represents the probability that the network predicts a certain category.

$$Loss2 = \sum_{j=1}^q y_j \log \sum_{j=1}^q \exp(o_j) - \sum_{j=1}^q y_j o_j \quad (4)$$

2.4 Expression analysis network of sheep face

The sheep facial expression analysis network is based on the EfficientNet detection framework, which consists primarily of MBConv convolution blocks. These blocks come in 3×3 and 5×5 modes, each incorporating three components: standard convolution, separable convolution, and a channel management attention mechanism. After the facial images of sheep are captured, issues such as low resolution and blurred facial expression features often arise, making it challenging for deep networks to extract and train important features. Additionally, the deep layers of the EfficientNet model and high input resolution lead to unnecessary computation, affecting the model's detection speed. To address these issues, the EfficientNet model is scaled using depth scaling factor d and resolution scaling factor r . The model was compressed with scaling factors of 1.0, 0.5, and 0.25, reducing the number of parameters, improving detection

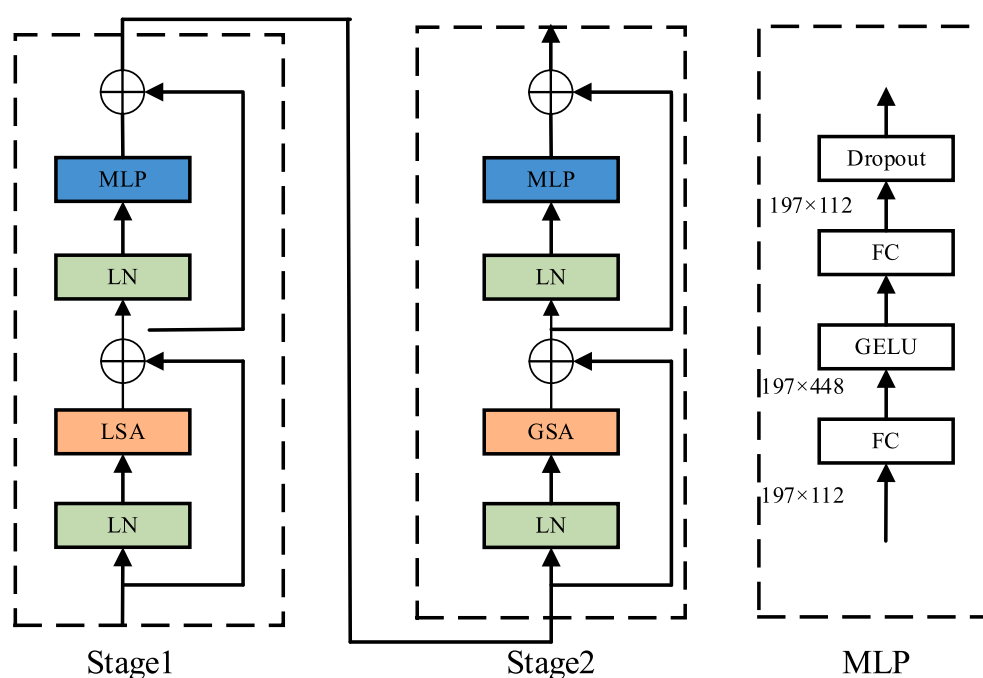


FIGURE 6
Structure diagram of separable self-attention mechanism.

accuracy, and enhancing the speed of the sheep recognition and classification system.

2.4.1 Facial classification network loss function

Focal Loss function was used to classify sheep faces and regression training sheep facial expressions in the facial expression analysis network as shown in Equation 5. P_i represents a probability factor as shown in Equation 6. When the prediction category is consistent with the real category, P_i is equal to the probability value of correct prediction, and conversely, the prediction probability value is inverse; γ represents the adjustable focusing parameter.

$$Loss3 = -(1 - P_i)^\gamma \log(P_i) \quad (5)$$

$$P_i = \begin{cases} p & \text{if } y=1 \\ 1-p & \text{otherwise} \end{cases} \quad (6)$$

2.5 Model training

2.5.1 Model training and parameter configuration

The facial identity recognition and expression classification system was developed using the PyTorch framework. The system consists of three networks, each trained separately before being integrated for sheep facial identity recognition, individual counting, and facial expression analysis. Transfer learning was applied to enhance the generalization ability and accelerate convergence of the system.

For the improved YOLOv5s multi-part recognition network, pre-training was conducted using the COCO dataset, with the network weights serving as the initialization. This process improves detection accuracy for sheep faces, backs, and lambs in a complex environment. The training settings include an image size of 640×640 , a batch size of 16, and 300 epochs. After each epoch, the system automatically saves the best-performing weights. The backbone network is frozen for the first 50 epochs, with an initial learning rate (LR) set to 0.001, which is reduced to 0.0001 for the next 250 epochs. The LR is smoothed using the cosine annealing algorithm to enhance feature extraction in the facial recognition network.

Regarding the sheep facial classification and expression analysis networks, the Mini-ImageNet dataset was used for transfer learning. The input image size was set to 224×224 , and the LambdaLR scheduling strategy was employed to adjust the LR periodically. The initial LR was set to 0.01 for both sub-networks, with the SGD optimizer used to optimize the model parameters.

2.5.2 Model evaluation index

To comprehensively evaluate the network detection performance of the facial identity recognition and expression classification system, several evaluation metrics are employed, including average precision (AP), mean average precision (mAP), accuracy, precision, recall, frames per second (FPS), model weight size (Weight), model parameter count (Params), model computation (FLOPs), and memory usage during model inference (Memory). FPS refers to the number of images processed per second. AP is calculated by plotting

the precision-recall (P-R) curve, with recall on the horizontal axis and precision on the vertical axis, and integrating to find the area under the curve. mAP is the mean of the AP values across all categories. The calculation formulas for these metrics are as follows:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$AP = \int_0^1 Precision \cdot Recall \quad (10)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (11)$$

In Equations 7–11, TP represents the number of positive samples predicted by the model that are consistent with the real label; FP represents the number of samples predicted by the model that are inconsistent with the actual positive samples; FN (False Negative) represents the number of samples predicted by the model that are inconsistent with the actual negative samples; TN represents the number of samples predicted by the model that are consistent with the actual negative samples.

3 Results and analysis

3.1 Comparison of the results of multi-part recognition model of sheep individual

3.1.1 Comparison of the results of different types of recognition algorithms

The same dataset C was used to train common target detection algorithms, such as the YOLO series, SSD, Faster-RCNN, CenterNet, and EfficientDet, to comprehensively evaluate the performance of the improved individual multi-part identification network. These detection algorithms were tested using test set D from the multi-part identification data of sheep individuals. The network models were assessed across six aspects: FPS, weight, mAP, Params, FLOPs, and memory usage. The test results are presented in Table 1. The mAP values in Table 1 were all obtained with an IOU threshold of 0.5.

As shown in Table 1, the highest FPS of YOLOv4-tiny is 129, making it the fastest in terms of detection speed; however, its mAP is only 77.12%, indicating low detection accuracy. YOLOv5-x achieves the highest detection accuracy with an mAP of 95.98%, but its weight, parameters, and FLOPs are too large, making it less suitable for lightweight model applications. YOLOv5-s has an FPS of 64, mAP of 93.29%, weight of 27.1, parameters of 711,785, and FLOPs of 8.27,

TABLE 1 Comparison of the results of different models.

Model	Fps/s	mAP/%	Weight/M	Params	Flops/G	Memory/M
SSD	83	81.01%	91.6	26,151,824	31.39	206.92
CenterNet	74	92.60%	124.0	32,665,432	22.15	610.50
Faster-RCNN	18	91.88%	521.0	28,469,983	461.65	532.34
EfficientDet	19	85.85%	15.0	3,839,060	2.12	306.24
YOLOv4-Tiny	129	77.12%	22.4	5,918,006	3.43	72.94
YOLOv4	34	92.74%	244	64,040,001	29.95	606.54
YOLOX-Nano	50	88.09%	3.7	900,459	1.24	230.00
YOLOX-Tiny	64	92.91%	19.4	5,038,395	7.59	260.22
YOLOX-s	61	93.40%	34.3	8,945,035	13.34	346.70
YOLOX-m	46	94.59%	96.8	25,291,755	36.78	646.21
YOLOX-l	29	94.32%	207.0	54,162,635	77.70	1030.10
YOLOX-x	18	94.01%	378.0	99,013,675	140.8	1498.36
YOLOv5-s	64	93.29%	27.1	711,785	8.27	286.78
YOLOv5-m	46	94.17%	80.6	21,133,185	25.33	555.53
YOLOv5-l	29	95.83%	178.0	46,733,665	57.28	908.65
YOLOv5-x	18	95.98%	378.0	87,257,832	108.68	1344.52

TABLE 2 Comparison of the test results of the improved multi-part identification models of sheep individual.

Model	Fps/s	mAP/%	Weight/M	Params	Flops/G	Memory/M
YOLOv5s	64	93.29	27.1	711,785	8.27	286.78
YOLOv5s-MCFB	57	95.77	59.7	14,522,120	15.96	485.93
YOLOv5s-MCFB+	53	95.84	73.9	18,580,104	18.12	425.77

offering a good balance between detection accuracy and speed. Therefore, YOLOv5-s is chosen as the baseline model for the multi-part recognition network.

3.1.2 Comparison of the results of improved multi-part recognition models of sheep individual

Improvements were made to YOLOv5s and experiments were conducted, as shown in Table 2. YOLOv5s-MCFB refers to the multi-part recognition network for sheep individuals that incorporates multi-link convolution feature fusion blocks into the path aggregation network, while YOLOv5s-MCFB+ further integrates the Repconv convolution structure on top of the multi-link convolution feature fusion blocks.

According to Table 2, the mAP for YOLOv5s-MCFB and YOLOv5s-MCFB+ reaches 95.77 and 95.84%, respectively. The FPS of these models is slightly lower than that of YOLOv5s, at 57 FPS and 53 FPS, respectively. Compared to YOLOX Nano, YOLOX Tiny, YOLOX-s, YOLO-m, YOLOX-l, YOLOX-x, YOLOv5s, YOLOv5m, and YOLOv5l, the mAP of YOLOv5s-MCFB+ increased by 7.75, 2.93, 2.44, 1.25, 1.52, 1.83, 2.55, 1.67, and 0.01%, respectively. This demonstrates the effectiveness of the improvements proposed in this paper.

To evaluate the detection performance of the improved multi-part recognition network for individual sheep counting, the backs, faces, and lambs in 120 images were counted. As shown in Table 3, a total of 187 faces, 936 backs, and 185 lambs were identified, with 35 individuals experiencing severe occlusion, bringing the total to 1,173

sheep. YOLOv5s-MCFB+ achieved the highest detection accuracy in counting sheep individuals, detecting 1,161 individuals with only 12 missed detections, resulting in a detection accuracy of 98.97%. When compared with YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x models, the results showed increases of 5.2, 3.32, 2.13, and 0.68%, respectively.

Since discussed in Figure 7, the prediction outcomes of the YOLOv5s-MCFB+ multi-part recognition model were visualized using class activation heat maps in order to further confirm the efficacy of the enhanced multi-part recognition network. The areas that the model concentrated on throughout the identification phase are indicated by these heat maps. It is clear that the model focuses more on the back and lamb aspects of the sheep and less on the face, which explains why the model is more successful at identifying the back and lamb than the face.

3.2 Comparison of the results of the improved facial classification network model

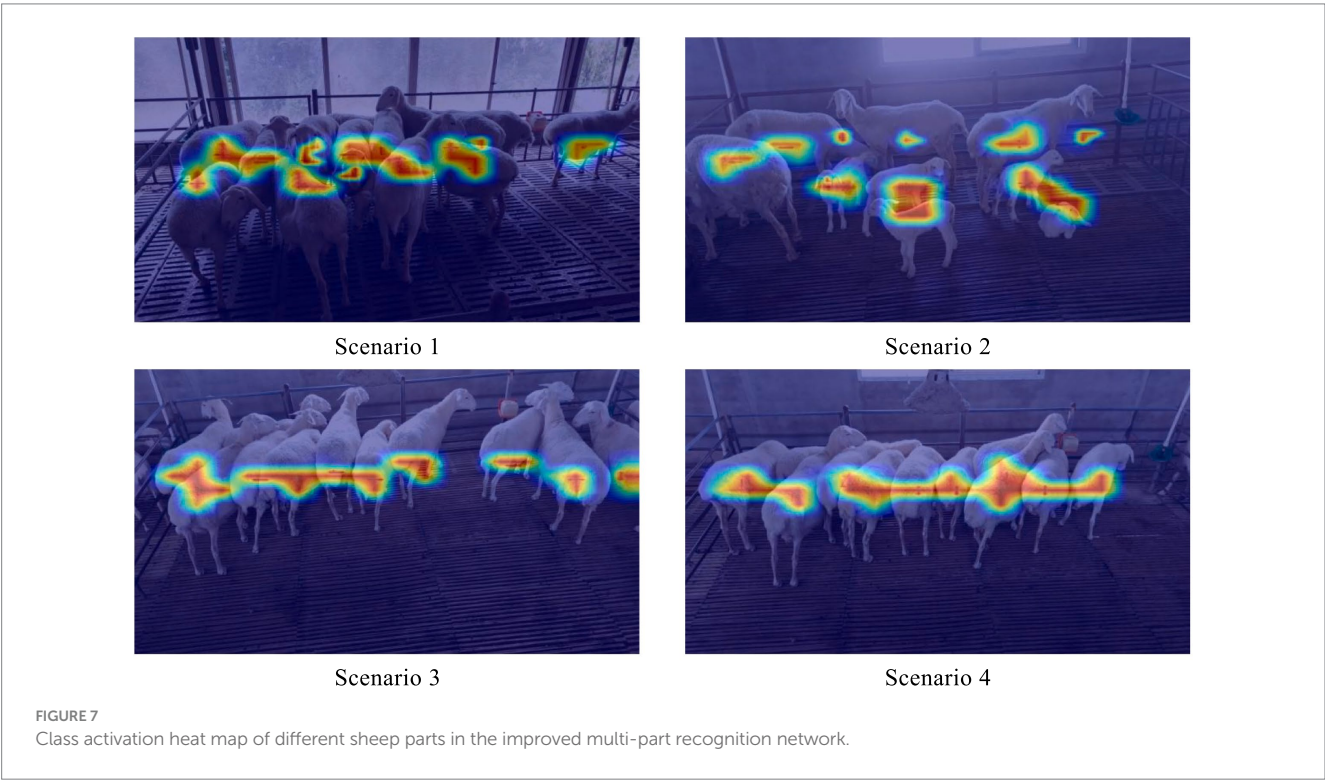
ResNet50, MobileNetV2, MobileNetV3, MobileVit, GhostNet and the improved facial classification network model were used for comparative training of sheep face classification dataset E. The performance of these networks was tested using test set F, and the results are summarized in Table 4. MobileNetVit, MobileNetVitxs, and MobileNetVitxxs represent different versions of the same classification

TABLE 3 Comparison of individual counting between the improved multi-part recognition model and the YOLOv5 series models.

Model	Face	Back	Lamb	Total	Shelter	Predict	Omission	Accuracy/%
YOLOv5s	187	936	185	1,138	35	1,100	73	93.77
YOLOv5m	187	936	185	1,138	35	1,122	51	95.65
YOLOv5l	187	936	185	1,138	35	1,136	37	96.84
YOLOv5x	187	936	185	1,138	35	1,142	31	97.35
YOLOv5s-MCCB	187	936	185	1,138	35	1,153	20	98.29
YOLOv5s-MCCB+	187	936	185	1,138	35	1,161	12	98.97

TABLE 4 Results comparison between the improved sheep face classification model and other models.

Model	Precision	Recall	Fps/s	Train_acc	Test_acc	Weight/M
ResNet50	99.4%	99.4%	47	96.5%	98.3%	81.3
MobileNetV2	96.2%	93.8%	76	94.3%	94.9%	8.86
MobileNetV3	96.5%	96.2%	66	94.7%	95.2%	16.3
MobileNetVit	92.8%	92.7%	52	90.7%	91.3%	19.2
MobileNetVitxs	95.8%	95.3%	55	93.4%	94.3%	7.81
MobileNetVitxxs	96.1%	96.3%	66	94.5%	94.7%	4.11
GhostNet	97.4%	96.3%	71	95.3%	95.8%	15.2
GhostNetVits	88.3%	84.1%	66	90.1%	89.5%	12.7
GhostNetVitm2	99.3%	99.4%	58	96.5%	98.1%	19.1
GhostNetVitm	99.4%	99.6%	58	96.7%	98.9%	19.1
GhostNetVitm8	97.4%	96.0%	58	95.4%	96.2%	19.1
GhostNetVitl	99.3%	99.2%	47	95.8%	98.2%	21.0



network. GhostNetVits, GhostNetVitm, and GhostNetViti correspond to networks in which varying amounts (1, 2, 3) of LSA-GSA three-stage face classification networks were introduced into the sixth-stage structure of the GhostNet classification network. The Multi-Head Attention mechanism, which integrates different network modules for deeper feature extraction, was also explored. Specifically, GhostNetVitm2, GhostNetVitm, and GhostNetVitm8 represent the variations of the network structure with 2, 4, and 8 Multi-Head Attention layers, respectively.

As shown in Table 4, the Precision, Recall, and Accuracy of the GhostNetVitm on the training and testing datasets are 99.4, 99.6, 96.7, and 98.7%, respectively, outperforming GhostNet, GhostNetVits, and GhostNetViti network. This indicates that adding the two-layer LSA-GSA structure to the GhostNet network yields the best results. Comparing different numbers of multi-head attention mechanisms, the results show that adding the four-layer multi-head attention mechanism yields the best performance. Therefore, the GhostNetVitm network is chosen for facial identity recognition of sheep.

To verify the effectiveness of the improved face classification network, the loss change curves for GhostNet, GhostNetVits, GhostNetVitm2, GhostNetVitm, GhostNetVitm8, and GhostNetViti on test set F are plotted, as shown in Figure 7.

As observed in Figure 8, the loss curve for the GhostNet facial recognition model exhibits significant fluctuations and a slow convergence rate. In contrast, the loss curve for GhostNetVits fluctuates more gently, although the overall loss value is slightly higher than that of the other networks. The loss curves for GhostNetVitm2, GhostNetVitm, GhostNetVitm8, and GhostNetViti show minimal differences. Among these, the improved GhostNetVitm face classification network achieves the lowest loss value and the fastest convergence, further demonstrating the effectiveness of the improved sheep face classification network.

Additionally, to demonstrate that the improved face classification network model focuses on the features of different facial regions of sheep, 12 facial images from ID001 to ID012 were selected. The class activation heat map was applied to visualize the last layer of the convolutional feature map. The specific results are shown in Figure 9.

The darker regions in Figure 8, which correspond to higher activation areas, indicate that the facial classification network model

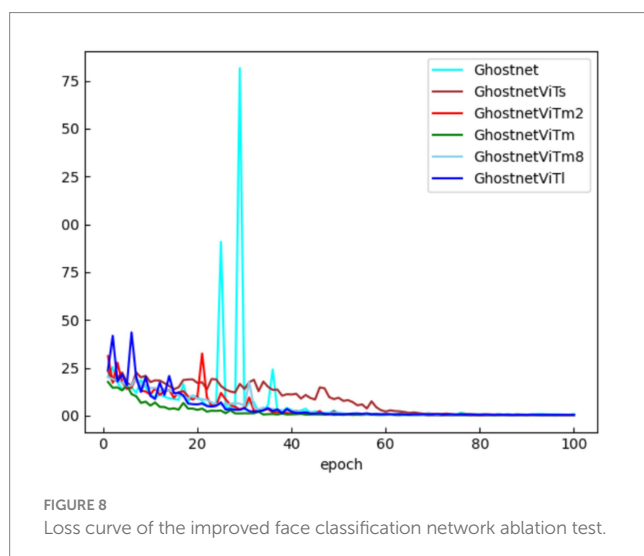
places more emphasis on the features in those regions. This suggests that these areas play a more significant role in the recognition process of the sheep's facial identity. The facial features of sheep are primarily focused around the bridge of the nose, with some concentration on the sides of the nose bridge. This highlights that the improved facial classification model primarily differentiates individual sheep based on the nose bridge area. For sheep with less distinct features in the nose bridge region, the model relies more on the sides of the nose bridge for identification.

3.3 Comparison of results of improved facial expression network model

In order to show the detection effect of the improved facial expression classification network, the model is trained and verified by using facial pictures of healthy sheep in the natural state and painful expression pictures in the pathological state. EfficientNet0.5 represents scaling of the depth of the network layers and the input image resolution size by 0.5 times, respectively. In Table 5, EfficientNet0.25 represents scaling for the depth of network layers and the resolution size of the input image by 0.25 times, respectively. EfficientNet0.5 facial expression classification network can achieve 99.5, 98.0%, 140Fps/s, 99.2%, and 2.64 M in terms of Precision, Recall, Fps, test set accuracy, and Weight size, respectively. Compared with EfficientNet, it has improved 5.4, 4.4 and 3.6% in Precision, Recall, test set accuracy and other metrics imitations. Compared to EfficientNet0.25, the improvement is 0.6, 2.3 and 0.5%, respectively. In terms of detection frame rate, EfficientNet0.25 has the fastest detection speed, up to 185Fps/s. EfficientNet0.5 has a detection speed of 140Fps/s, which is a significant improvement over EfficientNet and slightly lower than the detection speed of EfficientNet0.25. In terms of weight size, the EfficientNet0.25 model is the smallest, with a weight of only 0.55 M. The weight of EfficientNet0.5 is 2.64 M, which is a substantial reduction compared to EfficientNet.

To further verify the effectiveness of the network improvement, the loss curves of the above three network structures in the test set were plotted in this experiment, as shown in Figure 10A. EfficientNet_{0.5} loss curve has the best smoothness with increasing epoch, and the loss value is the lowest among the three curves. Therefore, for the sheep facial expression analysis network, appropriately reducing the depth of the network and the resolution of the input images helps improve both the detection accuracy and speed. It effectively reduces the risk of losing important features during the feature extraction process, enhancing the robustness of the model. In addition, this study selected facial images of sheep in healthy state and painful images in sick state, a total of 16 images. Heatmap visualization was performed on the facial expression analysis network EfficientNet0.5, as shown in Figure 10B.

The heat map reveals that the improved facial expression classification network primarily focuses on the eye region for sick sheep. In contrast, for facial images of sheep in a healthy state, the network emphasizes the region between the two eyes and the side face area. The distribution of important features in the facial visualizations indicates that the eye region plays a crucial role in distinguishing between the two conditions—healthy and sick sheep. This highlights the eye area as the key feature for identifying the different morphological states of the sheep.



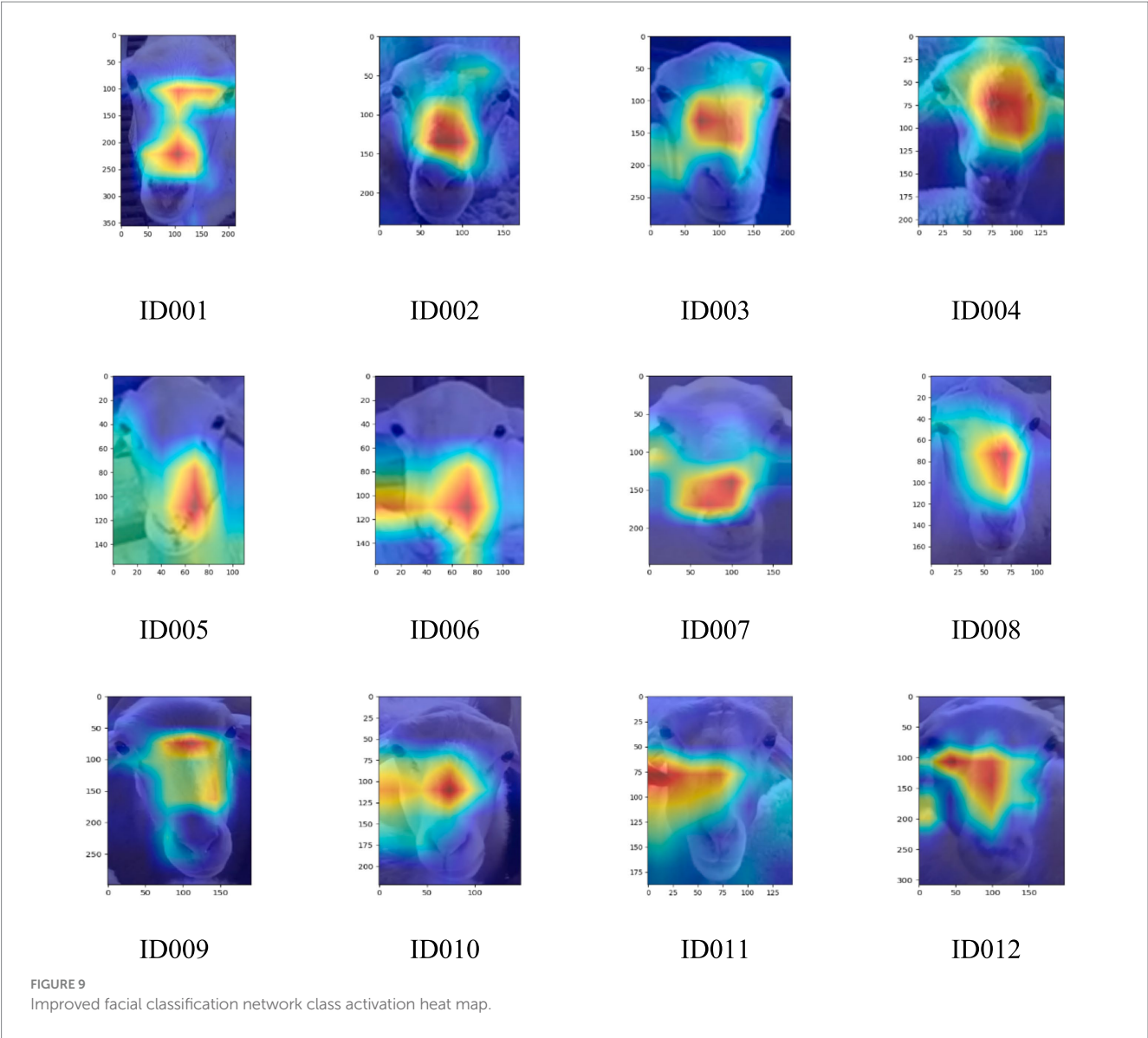


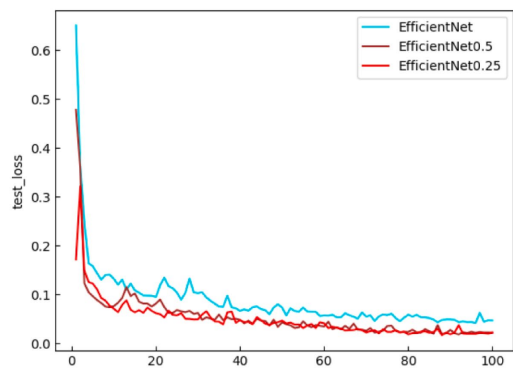
TABLE 5 Comparison of ablation test results of the improved facial expression classification network.

Model	Precision	Recall	Fps/s	Train_acc	Test_acc	Weight/M
EfficientNet	94.1	93.6	76	93.7	95.6	15.50
EfficientNet _{0.5}	99.5	98.0	140	98.2	99.2	2.64
EfficientNet _{0.25}	98.9	95.7	185	98.0	98.7	0.55

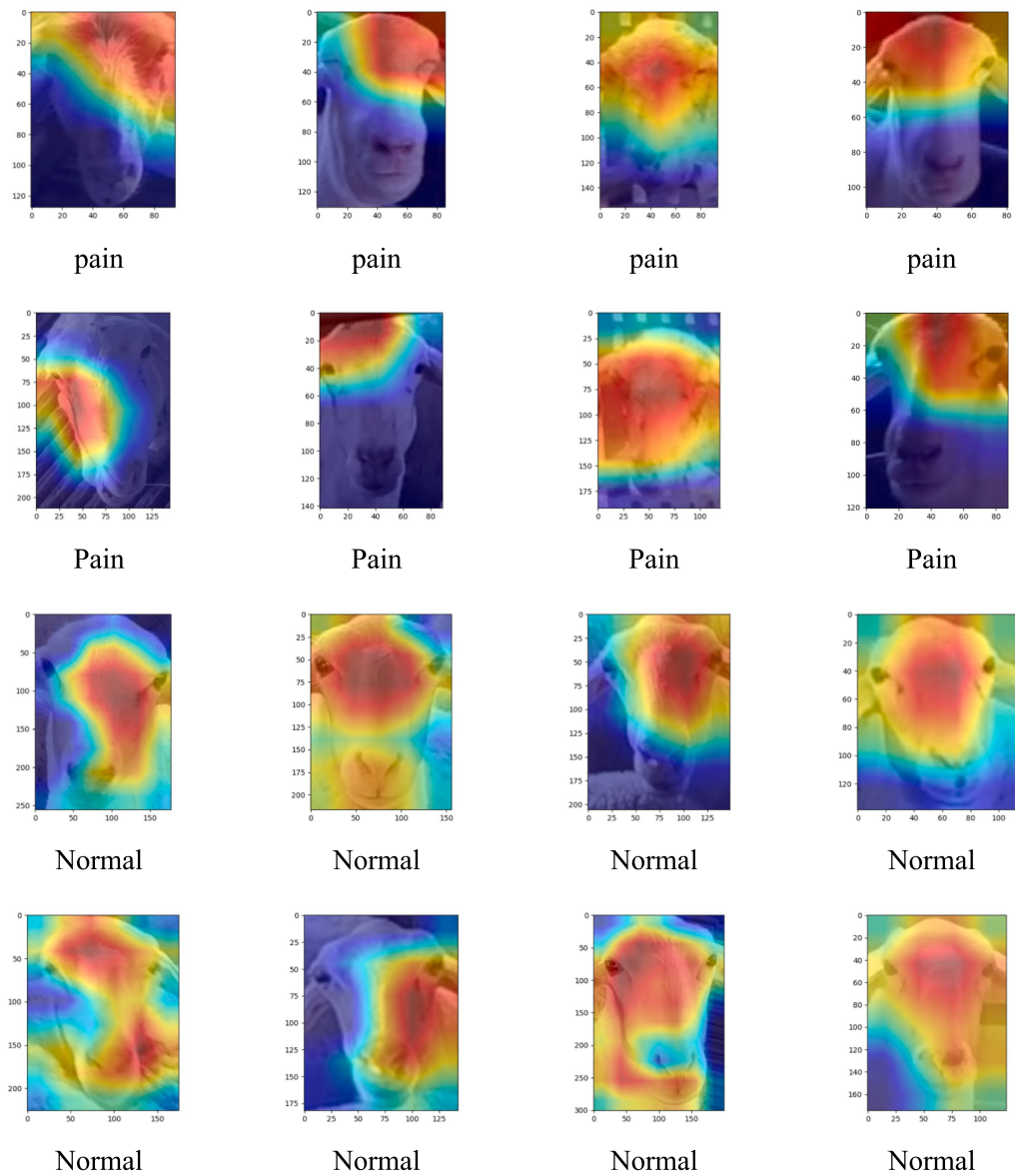
4 Discussion

This study focused on the analysis and summary of the detection results of individual multi-part recognition network, facial classification network and facial expression network of sheep. It was found that the improved individual multi-part recognition model focused more on the back and lamb features of sheep, but lacked focus on the facial region of sheep from the results of the improved individual multi-part recognition network. The improved multi-part recognition network for individual flock sheep may be missing when physical, behavioral and side faces are included. However, considering

the robustness of the individual multi-part recognition network, the network should selectively discard some images with obviously incomplete faces, so as to reduce the risk of misidentification and improve the accuracy of the facial classification network Guo et al. (31). Therefore, this balancing issue needs to be discussed further. In conclusion, the improved individual multi-part recognition network of sheep needs to be improved the detection effect of face as much as possible in all aspects, as well as back and pup detection performance. The facial classification network is mainly responsible for judging the individual facial images captured by the multi-part recognition network to realize individual identity recognition in



(a) Network loss curve of improved facial expression analysis



(b) Improved facial expression class activation heat map

FIGURE 10
Network loss curve and activation heat map. (A) Network loss curve of improved facial expression analysis. (B) Improved facial expression class activation heat map.

the herd environment. The detection results of the improved facial classification network showed that the network mainly discriminates the nasal bridge region of sheep to achieve the identification of individual sheep. And when individual nasal region features are not obvious, the model needs to make critical distinctions by features on both sides of the nasal bridge. Whether or not individual sheep identity can be reliably determined by features on the nasal bridge is still to be verified in future work, considering whether it is influenced by sheep species and the number of individual sheep Hitehman et al. (32).

For the facial expression analysis network, this study focuses on the differences and connection between the natural expression of healthy sheep and the painful expressions of sick sheep to achieve the initial health management of group sheep. Through the visualization of the training process, it was observed that the eyes of healthy sheep are more lively, while the eyes of sick sheep appear dull Fitzpatrick et al. (33). Therefore, it is feasible to judge the health status of sheep by using the features of sheep eye region in a certain principle.

5 Conclusion

- (1) For the problems that the target recognition algorithm with large number of network parameters has low detection speed but it has low detection accuracy with small network parameters, an improved multi-part recognition algorithm was proposed based on YOLOv5s as the prototype. A multi-link convolution residual feature fusion structure was introduced into the YOLOv5s path aggregation network structure to improve the screening ability of fine-grained features of objects of different sizes. In order to further improve the detection effect of dense targets and the detection ability of small targets, a layer of Repconv convolution structure was added to the head part on the basis of the introduction of the multi-convolution residual feature fusion structure in the sheep individual multi-part recognition algorithm, so as to realize the detection of small targets in complex environments.
- (2) Aiming at the characteristics of sheep with similar facial texture, small intra-class feature variability, and fine-grained features that were not easy to distinguish, improvements were made on the basis of the Ghostnet facial classification network. In order to effectively replace the sixth stage network structure in GhostNet, the four-layer SSSA was used to enhance the feature extraction ability of sheep facial features and improve the detection accuracy of the facial classification network.
- (3) For the improvement of the detection speed and effectiveness of the facial expression classification network, model compression tests were performed for the expression analysis network EfficientNet. The test results showed that certain compression EfficientNet helped to improve the model detection accuracy and enhance the model robustness and generalization performance.

Data availability statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Ethics statement

Ethics review and approval and written informed consent were not required as per local legislation and institutional requirements.

Author contributions

FZ: Conceptualization, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. XZ: Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. SW: Methodology, Investigation, Data curation, Writing - review & editing. YQ: Software, Data curation, Writing - review & editing. SF: Conceptualization, Resources, Writing - review & editing, Supervision. YZ: Writing - review & editing, Supervision.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was support by the National Natural Science Foundation of China (No. 52075149), and Frontier Exploration Projects of Longmen Laboratory (Grant No. LMQYTSKT032), the National Key Research and Development Program of China project (No. 2017YFD0301106), the Scientific and Technological Project of Henan Province (Grant No. 242102110337), Graduate Education Reform Project of Henan Province (2023SJGLX180Y), Colleges and Universities of Henan Province Youth Backbone Teacher Training Program (Grant No. 2017GGJS062).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Wang L, Hu B, Hou Y, Wu H. Lightweight sheep head detection and dynamic counting method based on neural network. *Animals*. (2023) 13:3459. doi: 10.3390/ani13223459
- Li X, Zhang Y, Li S. SheepFaceNet: a speed–accuracy balanced model for sheep face recognition. *Animals*. (2023) 13:1930. doi: 10.3390/ani13121930
- McLennan KM, Rebelo CJ, Corke MJ, Holmes MA, Leach MC, Constantino-Casas F. Development of a facial expression scale using footrot and mastitis as models of pain in sheep. *Appl Anim Behav Sci*. (2016) 176:19–26. doi: 10.1016/j.applanim.2016.01.007
- Tsai DM, Huang CY. A motion and image analysis method for automatic detection of estrus and mating behavior in cattle. *Comput Electron Agric*. (2014) 104:25–31. doi: 10.1016/j.compag.2014.03.003
- Xu BB, Wang WS, Guo LF, Chen GP. A review and future prospects on cattle recognition based on non-contact identification. *J Agric Sci Technol*. (2020) 22:79–89. doi: 10.5555/20220037493
- Yao Z, Tan H, Tian F. Research progress of computer vision technology in wisdom sheep farm. *China Feed*. (2021) 7:7–12. doi: 10.15906/j.cnki.cn11-2975/s.20210704
- Barron UG, Corkery G, Barry B, Butler F, McDonnell K, Ward S. Assessment of retinal recognition technology as a biometric method for sheep identification. *Comput Electron Agric*. (2008) 60:156–166. doi: 10.1016/j.compag.2007.07.010
- Leslie E, Hernández-Jover M, Newman R, Holyoake P. Assessment of acute pain experienced by piglets from ear tagging, ear notching and intraperitoneal injectable transponders. *Appl Anim Behav Sci*. (2010) 127:86–95. doi: 10.1016/j.applanim.2010.09.006
- Adell N, Puig P, Rojas-Olivares A, Caja G, Carné S, Salama AA. A bivariate model for retinal image identification in lambs. *Comput Electron Agric*. (2012) 87:108–112. doi: 10.1016/j.compag.2012.05.011
- Wei Z. Research on imperfect Bovine Iris recognition based on combination of local features and global features. Nanjing: Southeast University (2017).
- Wan Z, Tian F, Zhang C. Sheep face recognition model based on deep learning and bilinear feature fusion. *Animals*. (2023) 13:1957. doi: 10.3390/ani13121957
- Zhang C, Zhang H, Tian F, Zhou Y, Zhao S, Du X. Research on sheep face recognition algorithm based on improved AlexNet model. *Neural Comput Applic*. (2023) 35:24971–24979. doi: 10.1007/s00521-023-08413-3
- Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. Face recognition: a literature survey. *ACM Comput Surveys*. (2003) 35:399–458. doi: 10.1145/954339.954342
- Zhang HM, Zhou LX, Li YH. Sheep face recognition method based on improved MobileFaceNet. *Trans Chinese Soc Agric Mach*. (2022) 53:267–274.
- Gadekallu TR, Rajput DS, Reddy MPK, Lakshmana K, Bhattacharya S, Singh S, et al. A novel PCA-whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. *J Real-Time Image Proc*. (2021) 18:1383–1396. doi: 10.1007/s11554-020-00987-8
- Hou J, He Y, Yang H, Connor T, Gao J, Wang Y, et al. Identification of animal individuals using deep learning: a case study of giant panda. *Biol Conserv*. (2020) 242:108414. doi: 10.1016/j.biocon.2020.108414
- Khaldi Y, Benzaoui A, Ouahabi A, Jacques S, Taleb-Ahmed A. Ear recognition based on deep unsupervised active learning. *IEEE Sensors J*. (2021) 21:20704–20713. doi: 10.1109/JSEN.2021.3100151
- Salama AYA, Hassanien AE, Fahmy A. Sheep identification using a hybrid deep learning and bayesian optimization approach. *IEEE Access*. (2019) 7:31681–31687. doi: 10.1109/ACCESS.2019.2902724
- Xie Q, Wu M, Bao J, Yin H, Liu H, Li X, et al. Individual pig face recognition combined with attention mechanism. *J Agric Eng*. (2022) 38:180–188. doi: 10.6041/j.issn.1000-1298.2022.05.027
- Wang R, Shi ZF, Gao RH. Pig individual recognition based on multi-scale convolutional network in variable environment. *Acta Agric Univ Jiangxiensis*. (2020) 42:391–400. doi: 10.13836/j.jjau.2020046
- Li XY, Li YH. Feature point detection method of pig face based on convolutional neural network. *J Jilin Univ*. (2022) 60:609–616. doi: 10.13413/j.cnki.jdxblxb.2021272
- Wang Z, Liu T. Two-stage method based on triplet margin loss for pig face recognition. *Comput Electron Agric*. (2022) 194:106737. doi: 10.1016/j.compag.2022.106737
- Yang SQ, Liu YQH, Wang Z. Recognition based on improved YOLOv4 model with fused coordinate information. *Trans Chinese Soc Agric Eng*. (2021) 37:129–135. doi: 10.11975/j.issn.1002-6819.2021.15.016
- Zhu ML, Zhao LL, He SJ. Research and realization on cattle face recognition system model based on CNN combined with SVM and ResNet. *J Chongqing Univ Technol*. (2022) 36:155–161. doi: 10.3969/j.issn.1674-8425(z).2022.07.020
- Xu B, Wang W, Guo L, Chen G, Wang Y, Zhang W, et al. Evaluation of deep learning for automatic multi-view face detection in cattle. *Agriculture*. (2021) 11:1062. doi: 10.3390/agriculture11111062
- Zhang X, Xuan C, Ma Y, Su H, Zhang M. Biometric facial identification using attention module optimized YOLOv4 for sheep. *Comput Electron Agric*. (2022) 203:107452. doi: 10.1016/j.compag.2022.107452
- Song S, Liu T, Wang H, Hasi B, Yuan C, Gao F, et al. Using pruning-based YOLOv3 deep learning algorithm for accurate detection of sheep face. *Animals*. (2022) 12:1465. doi: 10.3390/ani12111465
- Billah M, Wang X, Yu J, Jiang Y. Real-time goat face recognition using convolutional neural network. *Comput Electron Agric*. (2022) 194:106730. doi: 10.1016/j.compag.2022.106730
- Han D, Wang B, Wang L. Individual pain recognition method of sheep based on improved VGGNet. *Trans Chinese Soc Agric Mach*. (2022) 53:311–317. doi: 10.6041/j.issn.1000-1298.2022.06.033
- Noor A, Zhao Y, Koubaa A, Wu L, Khan R, Abdalla FY. Automated sheep facial expression classification using deep transfer learning. *Comput Electron Agric*. (2020) 175:105528. doi: 10.1016/j.compag.2020.105528
- Guo Y, Yu Z, Hou Z, Zhang W, Qi G. Sheep face image dataset and DT-YOLOv5s for sheep breed recognition. *Comput Electron Agric*. (2023) 211:108027. doi: 10.1016/j.compag.2023.108027
- Hitelman A, Edan Y, Godo A, Berenstein R, Lepar J, Halachmi I. Biometric identification of sheep via a machine-vision system. *Comput Electron Agric*. (2022) 194:106713. doi: 10.1016/j.compag.2022.106713
- Fitzpatrick J, Scott M, Nolan A. Assessment of pain and welfare in sheep. *Small Rumin Res*. (2006) 62:55–61. doi: 10.1016/j.smallrumres.2005.07.028

Frontiers in Veterinary Science

Transforms how we investigate and improve
animal health

The third most-cited veterinary science journal,
bridging animal and human health with a
comparative approach to medical challenges. It
explores innovative biotechnology and therapy for
improved health outcomes.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

