BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

ANIMAL TRANSLATOR USING ARTIFICIAL INTELLIGENCE: SMART TECHNOLOGY TO INTERPRET ANIMAL BEHAVIOR AND SOUNDS

Nayan Kalamkar

Artificial Intelligence and Data Science Department, Ajeenkya DY Patil School of Engineering, Maharashtra, India. kalamkarnayan204gmail.com

Prof. Poonam Pawar

Artificial Intelligence and Data Science Department, Ajeenkya DY Patil School of Engineering, Maharashtra, India. pawar.punam08@gmail.com

ABSTRACT

Animal communication takes place via vocalizations, body posture, movement, facial expressions, and context. Although there are some animal communications known to humans, decoding the internal state of an animal correctly may be challenging, especially if there are different species of animals. An idea of an AI Animal Translator is introduced in this paper. It is supposed that the system would perform audio analysis, computer vision, machine learning, and other context-related operations to classify the animals' sounds and behaviour into understandable human messages like hunger, fear, pain, stress, comfort, alertness, or any social activity. The proposed system would include data collection, preprocessing, features extraction, multimodal classification, confidence assessment, and alerting for users. This device should be used as an auxiliary system to help people to better comprehend animals' communication, but not as a direct translation from one language to another. The review of related scientific trends, the problem statement, the architecture of the proposed system, evaluation methodology, applications, limitations, ethical considerations, and further research are covered in the paper.

Keywords – Artificial Intelligence, Animal Communication, Bioacoustics, Machine Learning, Deep Learning, Computer Vision, Sound Classification, Behavior Recognition, Multimodal Learning

I. INTRODUCTION

Among the aspects that characterize animal behavior, communication stands out as one of the most critical ones. Communication can involve various calls, barks, meows, chirps, growls, whistles, etc., as well as body postures, tail movements, ear positioning, facial expressions, and movements of the whole body. The signals can indicate social behavior, alarm, mating, foraging, territorial behavior, and distress, among others. Understanding from the perspective of human perception largely depends on the species, situation, and experience; the same sound can have different meaning depending on the context.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

The recent development in the field of artificial intelligence makes it possible to analyze extensive datasets of audio and video recordings. Machine learning algorithms will help to recognize acoustic signals, while computer vision algorithms will identify visual actions and postures. Utilization of both modalities together with the context can improve the recognition accuracy compared to the case when either sound or picture alone is analyzed. Thus, the aim of the AI Animal Translator is to analyze the observed signals and give the interpretation of the state of an animal.

In the term "translator," the idea is understood in a limited and pragmatical sense. First of all, this does not imply a decoding of the universal language of animals or the ability to understand the thoughts of animals. It just implies the learning of the patterns that had been identified in the past by researchers or experts in this particular area. These results should be treated as probabilistic interpretations and not diagnoses. This is especially important when a signal of disease or pain is recognized.

II. LITERATURE SURVEY

The literature on bioacoustic analysis has proven that animal sounds could be detected, segmented and classified using signal processing and machine learning methods. The existing methods include such features as Mel-frequency cepstral coefficients, spectral centroid, energy and pitch-related features, as well as classifiers such as SVM and Random Forests. While these techniques work well enough in structured data sets, their performance depends on good quality recordings and proper choice of features.

The technique of deep learning includes usage of convolutional neural networks for analyzing spectrograms and recurrent or attention networks for temporal modeling. Usage of audio embeddings and transformers might help perform transfer learning when there is a shortage of labeled data. Still, deep learning models might perform worse when they are used in a different setting from the training one: another species, environment, microphone and recording conditions.

Computer vision technologies were applied for recognition of actions and postures of animals based on images or video footage. Methods such as pose estimation help detect body keypoints, while image classifiers are able to detect broad visual states. Multimodal systems might help classify sounds using both audio and visual features in order to reduce ambiguity in case sound alone is not sufficient. Existing research proves the possibility of signal classification, but there are still many challenges left for the creation of an animal translator due to difference in labels, meanings and schemes of communication between species.

Ref.	Study	Approach / Focus	Main Contribution
[1]	Stowell (2022)	Computational bioacoustics and deep learning	Summarised recent developments in deep-learning-based acoustic analysis of animal sounds and outlined future research needs.
[2]	Piczak (2015)	Convolutional Neural Networks (CNNs) for sound classification	Showcased how CNNs could be used to learn sound representations in form of a spectrogram for classification.
[3]	Hershey et al. (2017)	Different CNN Architectures for audio recognition	Explored deep CNN architectures for audio recognition tasks, which have applications in animal sound classification.
[4]	Salamon, Jacoby and Bello (2014)	Urban sound datasets and classification	A developed dataset and taxonomy of urban sounds that help build and evaluate sound-classification systems.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

[5]	Stowell et al. (2019)	Bird audio detection using deep learning	Created a benchmark challenge for assessing automated bird audio detection algorithms based on real-world data.
[6]	Bello et al. (2019)	Monitoring and analysing acoustic environment	Presented a system for monitoring and analysing sounds in the urban environment.

Table I. Summary of major research directions relevant to an AI Animal Translator.

III. PROBLEM FORMULATION AND OBJECTIVES

Let us take A to be an audio clip, V to be an image or video clip, and C to be context in terms of species, time, location, and recent history. Then, the system is used to estimate a set of behavioral labels Y along with the probabilities of those predictions: P(Y | A, V, C). Examples of these predictions may be that the animal vocalizations correspond to alarm calls or that certain postures correspond to the animals being uncomfortable. This label prediction is only an estimation based on the training.

The primary goals are:

(1) detection/classification of animal sounds;

(2) identification of behaviors/posture if there is video;

(3) modality integration and contextual inputs;

(4) intuitive explanation with confidence;

(5) multi-species targets using species-specific models;

(6) recording and user privacy protection; and

(7) accuracy assessment through expert-labeled samples.

IV. MOTIVATION FOR AI-BASED ANIMAL COMMUNICATION

The owners of pets can have trouble differentiating between the signals that the animal wants attention, or it responds to an unusual sound, or it feels uncomfortable. The farmers and other people who take care of animals are watching many of them and cannot track all their behaviors. For the veterinarians, the structured logging of how vocalizations and activities vary over time will be useful.

This tool can provide assistance in recognizing patterns, structuring observations, and realizing that expert consultation is required. It does not lie in the fact that this will substitute animal-care experts but rather in additional evidence and easier re-reviewing of repeated observations. The system should not confuse the user by giving uncertain predictions as a fact, especially in case of outputs related to medicine or animal well-being.

V. PROPOSED SYSTEM AND METHODOLOGY

A. Audio/Video Recordings Collection and Labelling

The system will use microphone for audio and camera for video capturing. Each record should contain metadata regarding the species and recording settings along with context information and labels confirmed

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

by observers or experts in the field. Different subjects, age, context, and background noise should be used in the recordings. It is important to consider the issues of consent and well-being of the animals when collecting recordings.

B. Preprocessing and Features Extraction

For audio preprocessing we might use the steps of noise reduction, removing silences, normalization of amplitude, and segmentation of the sound into equal-length windows. We can transform the audio signal into a waveform representation which can further be converted into a spectrogram or log-Mel spectrogram. Other potential audio features are MFCCs, pitch, energy, duration, and spectral bandwidth. For video preprocessing, we might select frames, resize, adjust the background, and detect the presence of the animal.

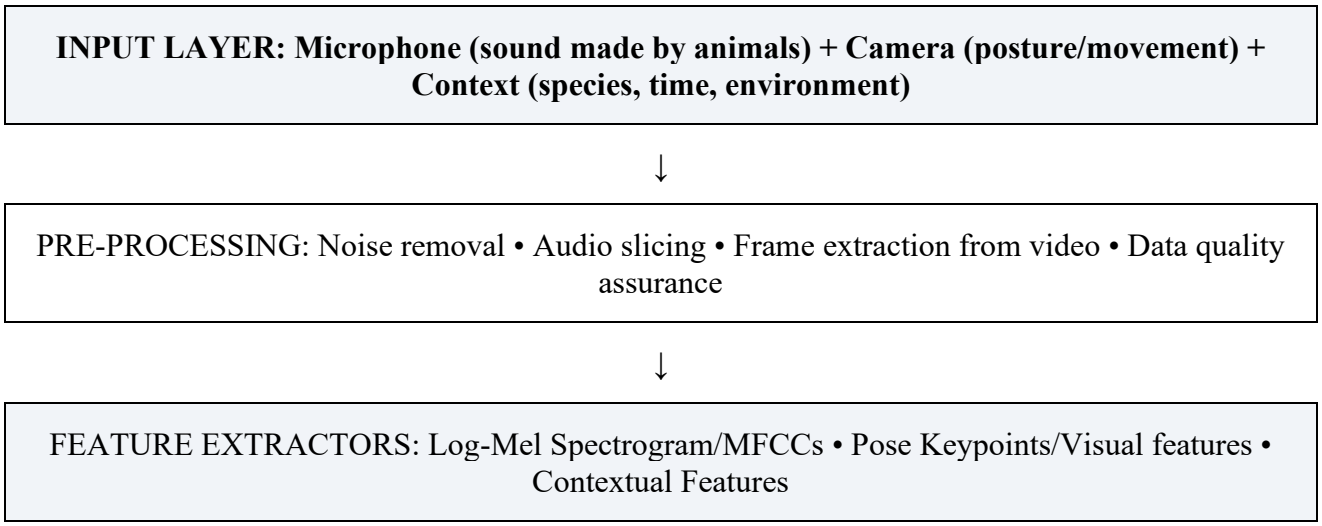
C. Model Training and Multimodal Fusion

We can use CNN-based or any pre-trained audio model for recognition of sounds. For recognizing the behavior from the video, we can use a CNN to detect animal's posture/movement. Two models will generate two feature sets which can be fused at the feature level, decision level, or using an attention mechanism. Species identifier or species-specific classifier will be useful for avoiding mistakes due to different mechanisms of sound generation in other species. The training process should use separate datasets for training, validation, and testing, and it would be best to divide individual animals between these datasets in order to avoid data leakage.

D. Interpretation and UI design

The output statements should include precise terms like "This pattern might indicate alertness" or "This behavior relates to possible discomfort". The detected sound/behavior, certainty of the output, timestamp of the event, and a short description are required. If the certainty of the classification is less than some threshold or an unknown species was provided, the system should say "Uncertain" instead of trying to make a classification. Consistent signs of pain or illness require consulting a veterinarian.

VI. PROPOSED SYSTEM ARCHITECTURE



BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

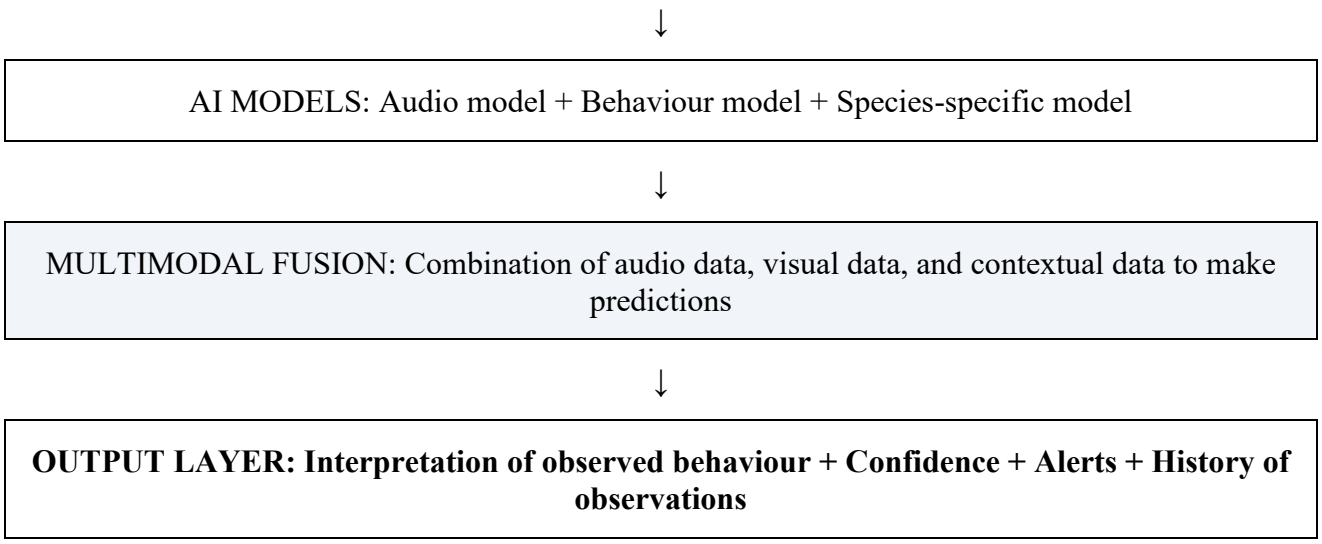


Fig. 1. Conceptual architecture of the proposed AI Animal Translator.

VII. EVALUATION METRICS

Evaluation of the performance of the AI Animal Translator can be performed based on the extent to which the proposed system can recognize sounds and comprehend animal behavior. Factors such as sound classification, spectrogram analysis, vocalization detection, and behavior recognition can be used to prove whether the system performs well in recognizing sounds and possibly some states, including hunger, fear, stress, alertness, or discomfort. Audio recordings, video observations, and labeled datasets can be used for evaluation purposes.

An effective AI Animal Translator system should work with high accuracy in different animal species, individual animals, and conditions of their environment. Such metrics as accuracy, precision, recall, F1 score, and confusion matrix can be used for the assessment of the system's performance. Sound classification can be evaluated with different vocalizations, and behavior recognition can be performed by taking into consideration animal's posture, facial expressions, and movements. Evaluation of the AI Animal Translator under noisy conditions and with an unfamiliar animal will show the performance of the system.

Evaluation of the AI Animal Translator in challenging environments, including background noises, unclear vocalizations, changes in behavior, and species' difference, can be used to reveal incorrect predictions of the system. It is crucial to have the uncertainty detection and confidence scores because it is essential for the system to avoid providing uncertain interpretations of animal behavior or emotions as confirmed intentions or emotions of the animal.

There are several practical aspects of the system that can be used for comparison of different models of the AI Animal Translator including training time, computational and memory resources, prediction speed, size of the dataset, accuracy of the classification, and the number of supported species. The quality of audio and video, robustness in the real environment, and interpretation quality can be also used.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

VIII. CHALLENGES AND LIMITATIONS

Another obstacle in the development of the tool is that one cannot assume that the same acoustic pattern has the same meaning across different animals. Vocalizations may differ depending on the age, gender, health, individuality, social context and habitat of the animal. Public databases are small, uneven and unbalanced (some well-known species are studied more often).

Outdoor sounds are associated with wind, vehicles, insects, animals and microphone artifacts. In video data, low light, occlusions, perspective and distance can affect the performance of the system. Labels like fear, pain and happiness cannot be observed and require a lot of expert consideration. Misclassification results may lead people to overlook the real problem or to react to the false situation inappropriately.

Privacy and ethical data usage issues should be considered as well. Unintentional sound recording may contain humans' voice or location of the user. The system should collect the least amount of data, secure its storage, enable record deletion and inform about limitations. It should not be sold as a universal animal language decoder and substitute veterinarian.

IX. APPLICATIONS

Pet care: summarize vocalizations and activities and help the owner identify any changes that might be worth observing. Veterinary help: give the professional an audio/video recording with timestamp. Agriculture: monitor the animals' activities and vocalizations on farms or in the barns. Wildlife conservation: record the vocalizations of different species. Animal behavior study: organize recordings and select the possible patterns for research. Animal shelters: help observe the stress behaviors of the animals under structured conditions.

X. FUTURE DIRECTIONS

Going forward, balanced and expert-labeled datasets should be collected for different species and conditions, and approaches for individualized adaptation without overfitting need to be developed. Self-supervised learning and transfer learning could decrease dependency on costly manual labeling. Multimodal foundation models might combine sound, video, and other context data, whereas edge AI will allow low-latency inference on mobile devices. Federated learning might facilitate cooperation without centralized recordings.

The other crucial research area would be explainability – the system needs to specify what acoustic or visual feature led to the certain output. Calibration, out-of-distribution detection, external validation, and independent animal welfare evaluation should be considered as key elements of development. Studies over time are required to see whether the application makes a positive difference to animal treatment and if its notifications are useful to the user.

XI. CONCLUSION

In this paper, we propose the design of a multimodal decision support model called the AI Animal Translator, which aims at interpreting the sound and behavior of animals. The AI Animal Translator makes use of techniques such as audio preprocessing, machine learning, computer vision, context, and confidence-awareness to provide potential behavioral interpretations in an understandable format. Applications could range from pet ownership, veterinarian practice, livestock management, to wildlife studies. Nevertheless,

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

animal communication varies according to species and context, and the model should not make assumptions about what the animals think or feel. In order to use this tool effectively, various precautions are needed.

REFERENCES

[1] D. S. S. A. Stowell, "Computational bioacoustics with deep learning: a review and roadmap," arXiv preprint, 2022.
[2] J. Salamon, C. Jacoby, and J. P. Bello, "A Dataset and Taxonomy for Urban Sound Research," in Proc. ACM Multimedia, 2014.
[3] K. J. Piczak, "Environmental sound classification with convolutional neural networks," in Proc. IEEE International Workshop on Machine Learning for Signal Processing, 2015.
[4] S. Hershey et al., "CNN Architectures for Large-Scale Audio Classification," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.
[5] J. P. Bello et al., "SONYC: A System for Monitoring, Analysis and Mitigation of Urban Noise Pollution," Communications of the ACM, vol. 62, no. 2, pp. 68–77, 2019.
[6] A. G. Z. R. et al., "Animal behaviour recognition and tracking using computer vision: a review," relevant review literature; check exact references before final submission.
[7] World Organisation for Animal Health (WOAH), "Animal Welfare," topic resources, official website, access date 2026.
[8] Cornell Lab of Ornithology, "Bioacoustics and passive acoustic monitoring" resources.
[9] D. Stowell, M. D. Wood, H. Pamuła, Y. Stylianou, and H. Glotin, "Automatic acoustic detection of birds through deep learning: the first Bird Audio Detection challenge," Methods in Ecology and Evolution, 2019.