JOURNAL OF BASIC SCIENCE AND ENGINEERING

CNN-LSTM Deep Learning Architecture for Animal Identification Using Acoustics

[1]Rohini H M, [2]Dr. Prabhavathi S

[1]Research Scholar, Department of Electronics & Communication Engineering, RYMEC Ballari – 58104, Karnataka, India | Department of Electronics & Communication Engineering, Central University of Karnataka Kalaburgi – 58567 Karnataka, India.
[2]Professor & Head, Department of Electronics & Communication Engineering, RYMEC Ballari – 58104, Karnataka, India.

DOI: https://doie.org/10.10399/JBSE.2026305789

ABSTRACT- Animal vocalizations provide valuable information regarding species identification, behavioral activities, physiological conditions, and environmental interactions. Automated analysis of animal sounds has gained considerable attention in precision livestock farming, veterinary healthcare, biodiversity monitoring, and smart agricultural systems. Conventional animal recognition approaches mainly rely on manual observation or handcrafted acoustic features, which regularly experience poor scalability and limited generalization capability. This paper presents a deep learning-based Animal Sound Recognition framework using Mel-Frequency Cepstral Coefficients (MFCC) and a hybrid Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) architecture. The proposed framework consists of audio acquisition, preprocessing, MFCC feature extraction, CNN-based spatial feature learning, and LSTM-based temporal sequence modeling. A dataset containing 1,942 audio recordings from four animal species, namely cats, dogs, cows, and chickens, was utilized for evaluation. Proposed CNN-LSTM model achieved an overall classification accuracy of 92.45%. The developed framework provides an effective and non-invasive solution for intelligent animal monitoring applications since the obtained precision, recall, and F1-score values confirmed that the proposed model is robustness against Noise. Proposed CNN-LSTM model outperformed several existing animal sound recognition approaches.

Keywords: Acoustic Classification, Animal Sound Recognition, Bioacoustics, CNN-LSTM, Deep Learning, Livestock Monitoring, MFCC.

1. INTRODUCTION

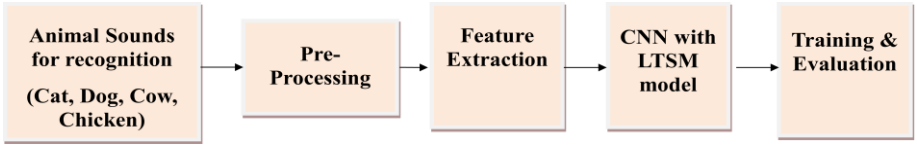
The increasing demand for intelligent livestock management and precision farming has accelerated the development of automated monitoring systems capable of continuously observing animal health and behavior. Animal vocalizations contain rich information related to species identity, emotional state, physiological conditions, reproductive status, stress levels, and disease symptoms. Consequently, automatic analysis of animal sounds has become an important research area in bioacoustics and artificial intelligence.

Traditional methods of animal monitoring rely heavily on visual inspection and manual observation by farm personnel. These approaches are labor-intensive, subjective, and often

impractical for large-scale livestock operations. Advancements in machine learning have enabled automated recognition of acoustic patterns directly from animal vocalizations.

Convolutional Neural Networks (CNNs), have shown remarkable success in audio and image classification tasks. CNNs are capable of automatically extracting hierarchical acoustic features from spectrogram representations. However, animal vocalizations are sequential signals that exhibit temporal dependencies, which cannot be fully captured by CNN architectures alone. Long Short-Term Memory (LSTM) networks are specifically designed to learn temporal relationships and long-range dependencies within sequential data. Figure 1 shows block diagram of animal Sound Recognition.

Figure.1: Block Diagram of Animal Sound Recognition.



Motivated by these advantages, this research proposes a CNN-LSTM framework for animal sound recognition. The proposed architecture integrates MFCC-based acoustic feature extraction, convolutional feature learning, and temporal sequence modeling to improve classification performance. The system was evaluated on a multiple animal sound dataset containing cat, dog, cow, and chicken vocalizations.

The major contributions of this work are:

• Development of a deep learning-based animal sound recognition framework.

• Integration of MFCC feature extraction with CNN-LSTM architecture.

• Evaluation using 1,942 animal sound recordings.

• Achievement of 92.45% classification accuracy.

• Comparative analysis with existing bioacoustic recognition approaches.

2. RELATED WORK

Animal sound recognition has improved significantly from conventional signal-processing methods to advanced Machine learning architectures. Bardeli et al. presented similarity search techniques for large-scale animal sound databases and shown the effectiveness of acoustic signatures for automated retrieval and classification [1]. Their work established the foundation for bioacoustic information systems. Yeo et al. demonstrated dog voice recognition systems using spectral feature extraction and pattern recognition techniques [2]. The study was later extended toward generalized animal voice recognition systems capable of recognizing multiple animal species [3]. Li and Wu proposed animal sound recognition based on double-feature spectrogram representations and demonstrated improved robustness

under noisy environmental conditions [4]. Lin et al. investigated classification of overlapping animal sounds using sparse representation techniques [5]. Their work addressed the challenge of simultaneous animal vocalizations frequently encountered in real-world environments. Narasimhan et al. introduced CNN-based frameworks for bird-song segmentation and classification [6]. The study demonstrated the ability of convolutional architectures to effectively model temporal and spectral structures of animal vocalizations.

Tang et al. explored birdsong identification using spectrogram-based feature representations and machine learning techniques [7]. Their findings highlighted the significance of frequency-domain information for species classification. Ruff et al. proposed an automated CNN-based animal sound recognition workflow that significantly reduced dependence on handcrafted acoustic features [8]. Sun et al. applied data augmentation strategies combined with CNN architectures for rainforest animal classification and demonstrated improved generalization performance [9]. Şaşmaz and Tek reported superior performance of CNN-based animal sound classifiers compared with traditional machine-learning approaches [10]. Singh et al. demonstrated that spectrogram-based CNN models can automatically learn discriminative acoustic patterns from animal vocalizations without extensive feature engineering [11].

Almost all existing studies focused on CNN-based spatial feature extraction. Combining spatial and temporal acoustic modeling for multi-class animal sound recognition was not explored. Therefore, this work introduces CNN-LSTM architecture capable of simultaneously learning spectral and temporal characteristics of animal vocalizations.

3. MATERIALS AND METHODS

3.1 Dataset Description

The proposed study utilizes animal vocalization datasets collected from authenticated online repositories, Kaggle datasets, livestock farming recordings, and publicly available animal sound databases. The dataset consists of acoustics from four animal species, namely cat, dog, cow, and chicken. Total of 1,942 audio samples were collected and standardized before model development. Table 1 shows number of Animal sound datasets used for research.

Table 1. Animal Sound Dataset.

Animal Species	Number of Samples
Cat	893
Dog	675
Cow	262
Chicken	122

The collected recordings represent a variety of natural vocalizations recorded under different environmental conditions. To ensure consistency, Acoustics were converted into a standardized audio format (wav) prior to feature extraction and model training.

3.2 Audio Preprocessing

Animal vocalizations recorded in practical environments are often contaminated with background noise originating from machinery, human activities, ventilation systems, and other animals. Therefore, preprocessing is essential to improve signal quality and enhance acoustic feature extraction.

The preprocessing stage includes:

• Pre-emphasis filtering

• Framing

• Windowing

• Signal normalization

• Spectrogram generation

• Audio augmentation

3.2.1 Pre-emphasis Filtering

Pre-emphasis filtering strengthens high-frequency components of acoustic signals while reducing low-frequency dominance. This process improves spectral balance and enhances discriminative acoustic features.

The pre-emphasis operation is represented as:

$y(n) = x(n) - \alpha\, x(n-1)$…… Equation 3.1

where x(n) denotes the original signal, y(n) represents the filtered signal, and α is the pre-emphasis coefficient.

3.2.2 Framing and Windowing

Animal sounds are non-stationary signals whose statistical properties change over time. Therefore, each audio recording is segmented into short-duration frames where stationarity can be assumed.

A Hamming window is applied to each frame to minimize spectral leakage and improve frequency-domain representation.

3.2.3 Spectrogram Generation

Vol. 23, No. 1, (2026)
ISSN: 1005-0930

Spectrograms provide time-frequency representations of audio signals and enable deep learning models to identify acoustic patterns effectively. Mel spectrograms and MFCC spectrograms were generated for further analysis. Figure 2 shows Me Mel Spectrogram of Animal Sound Signal.

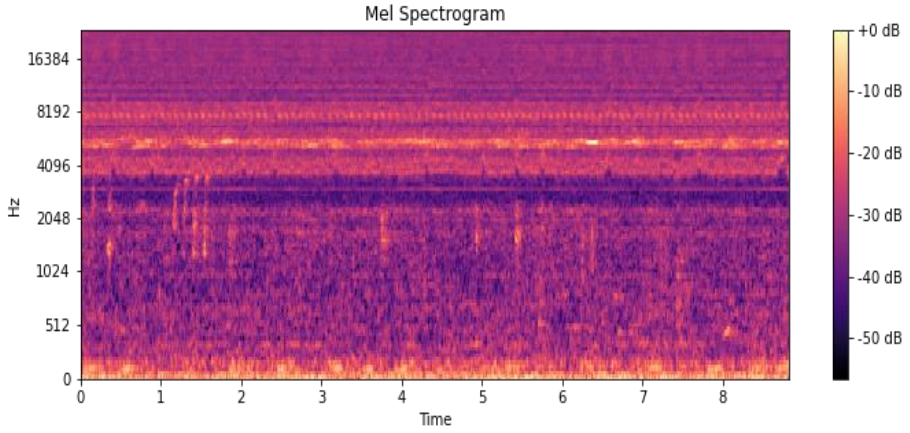


Figure.2: Mel Spectrogram of Animal Sound Signal.

3.3 MFCC Feature Extraction

Mel-Frequency Cepstral Coefficients (MFCCs) are acoustic descriptors for speech and bioacoustic analysis.

The MFCC extraction process consists of:

Step 1: Fourier Transform

Step 2: Mel-scale Filter Bank Generation

Step 3: Logarithmic Compression

Step 4: Discrete Cosine Transform

From each audio sample Forty MFCC coefficients were extracted. These coefficients capture spectral envelope characteristics, frequency distribution patterns, and temporal acoustic information relevant to animal vocalizations.

3.4 CNN-LSTM Architecture

The proposed framework combines Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks as shown in Figure 3.

.

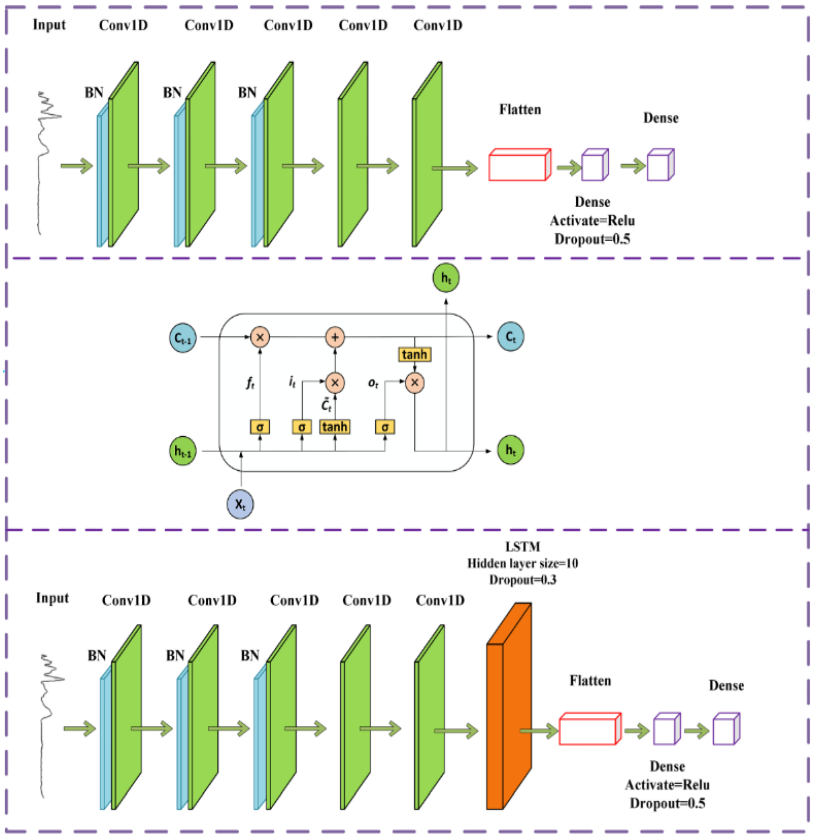


Figure.3: CNN-LSTM Architecture for Animal Identification.

CNN layers perform:

• Local feature extraction

• Spectral pattern learning

• Time-frequency representation analysis

LSTM layers perform:

• Sequential pattern learning

• Temporal dependency modeling

• Long-range contextual information extraction

The network architecture contains:

• Conv1D Layers

Vol. 23, No. 1, (2026)
ISSN: 1005-0930

• Batch Normalization Layers

• Max Pooling Layers

• LSTM Layers

• Dense Layers

• Dropout Regularization

The CNN component learns discriminative acoustic features while the LSTM component captures temporal relationships present in sequential animal vocalizations.

4. EXPERIMENTAL SETUP

4.3 Training Configuration

The dataset was separated into training and validation subsets using an 80:20 split.

Training Dataset Distribution:

Cat: 740

Dog: 542

Cow: 207

Chicken: 104

Validation Dataset Distribution:

Cat: 152

Dog: 133

Cow: 55

Chicken: 18

The model was trained using:

• Adam Optimizer

• Batch Size = 32

• Epochs = 40–100

• Early Stopping

• Learning Rate Optimization

5. RESULTS AND DISCUSSION

5.1 Confusion Matrix

Confusion matrix in Figure 4 demonstrated strong classification ability of the proposed CNN-LSTM model. Out of 358 validation samples, 331 samples were correctly classified, resulting in an overall classification accuracy of 92.45%.

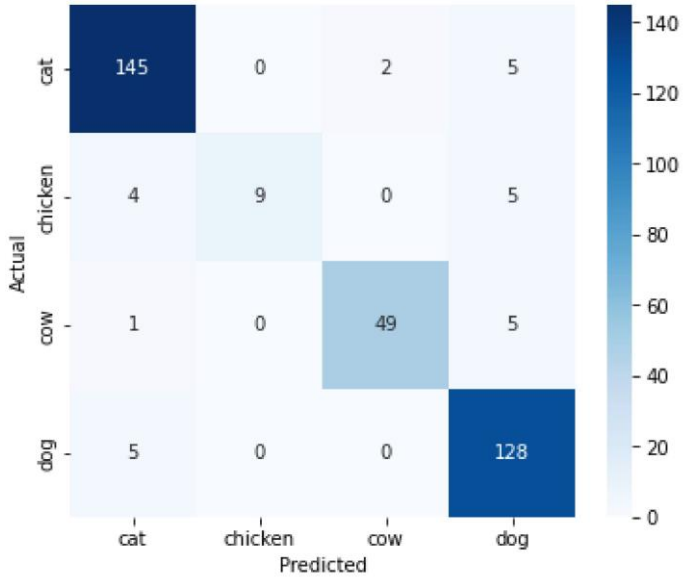


Figure.4: Confusion Matrix for Animal Identification using Acoustics.

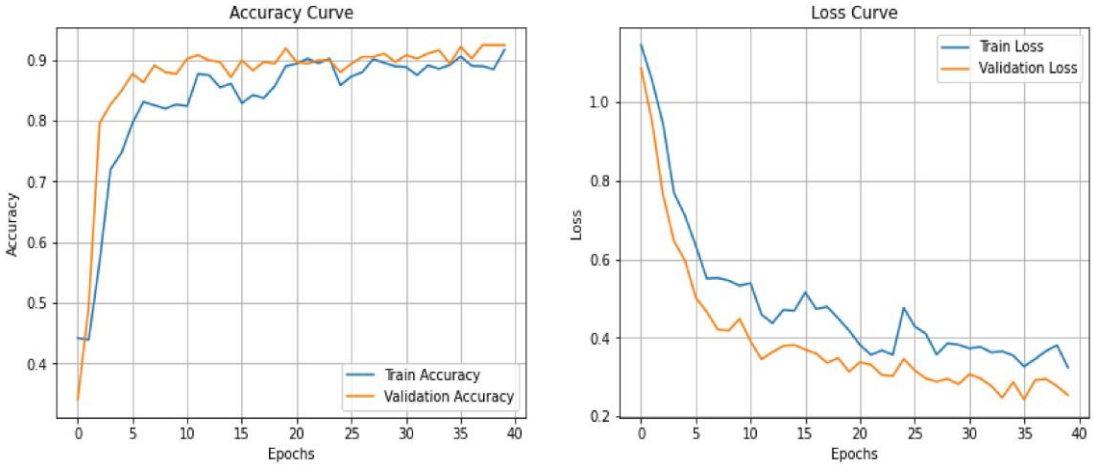
The Cat class achieved 145 correct classifications and Dog class achieved 128 correct classifications. The Cow class recorded 49 correct classifications. The Chicken category showed comparatively lower performance with 9 correctly classified samples.

The lower performance observed in the Chicken class may be attributed to limited training data, class imbalance, and overlapping acoustic characteristics with other animal vocalizations.

5.2 Training and Evaluation

The training and validation curves in Figure 5 indicate stable convergence throughout the training process.

Figure.5: Accuracy and Loss Curves for Animal Recognition.



The proposed CNN-LSTM model exhibited:

• Rapid convergence

• Reduced validation loss

• Strong generalization capability

• Minimal overfitting

Batch normalization and dropout layers significantly contributed toward stabilizing learning and improving model robustness.

The overall classification accuracy achieved by the proposed model was 92.45%. Figure 6 shows Precision, Recall & F1-Score for Animal Sound Recognition.

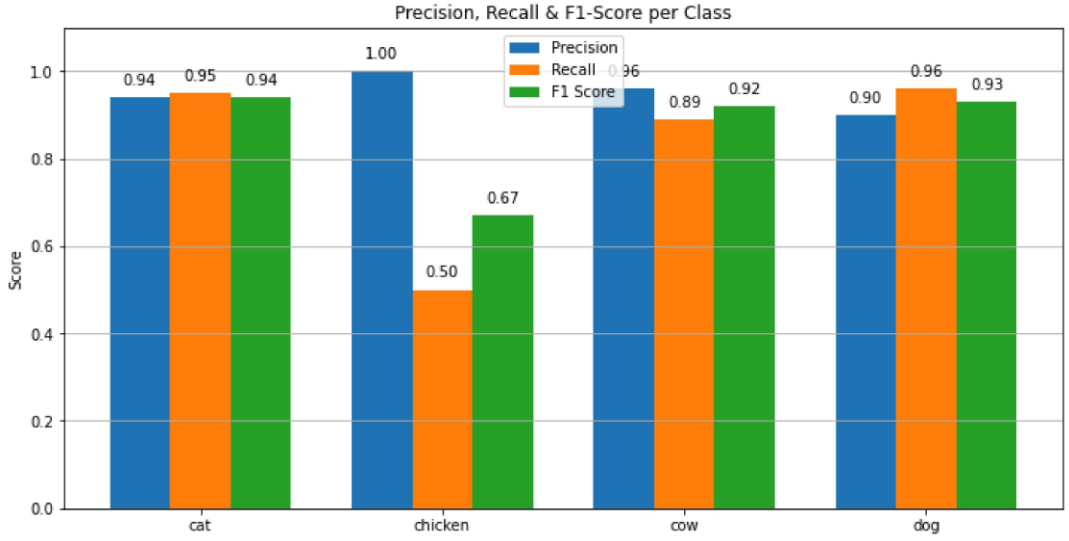


Figure.6: Precision–Recall–F1 for Animal Sound Recognition

Cat and Dog classes achieved the highest recognition performance because of their larger sample distributions and distinctive vocal patterns. The Cow class demonstrated reliable classification performance with high precision and recall values. Although the Chicken class achieved perfect precision, its lower recall suggests that several chicken vocalizations were misclassified into other animal categories.

5.4 Comparative Analysis of Animal sound Recognition with Existing Studies

Table 2 Gives Accuracy comparison of Animal Recognition with Existing Literature.

Table.2: Comparison with Existing Literature

Author	Methodology	Accuracy
Bardeli et al. [1]	Similarity Search	81.30%
Tang et al. [7]	SNM + RF	75.10%
Singh et al. [11]	CNN	88.00%
Pan et al. [12]	DNN	83.00%
Das et al. [13]	AI-based Model	91.32%
Bhumika et al. [14]	CNN	90.00%
Proposed Method	CNN-LSTM	92.45%

The proposed CNN-LSTM framework for Animal Sound Recognition achieved superior performance compared with several previously reported animal sound recognition systems.

The improvement can be attributed to:

• MFCC-based acoustic representation

• CNN-based feature extraction

• LSTM-based temporal modeling

• Deep hierarchical learning

These factors collectively contributed to improved classification accuracy and robustness.

6. CONCLUSION

This study presented a CNN-LSTM-based animal sound recognition framework capable of automatically classifying cat, dog, cow, and chicken vocalizations. MFCC features were extracted from preprocessed audio signals and supplied to a hybrid CNN-LSTM architecture for classification.

Experimental evaluation conducted on 1,942 audio recordings established an overall classification accuracy of 92.45%. The model achieved strong precision, recall, and F1-score values across most animal categories, confirming its ability to effectively learn acoustic representations from animal vocalizations.

The findings indicate that integrating convolutional feature extraction with temporal sequence modeling significantly enhances bioacoustic classification performance. The proposed model can serve as an intelligent and non-invasive tool for livestock monitoring, veterinary assistance, animal welfare assessment, and smart farming applications.

7. FUTURE WORK

Future research can focus on:

• Increasing dataset diversity and size.

• Inclusion of additional animal species.

• Development of real-time monitoring systems.

• Integration with IoT-based livestock management platforms.

• Investigation of Transformer and Attention-based deep learning architectures.

• Development of ensemble deep learning models for further performance improvement.

• Deployment on edge computing devices for field applications.

REFERENCES

1. Rolf Bardeli, “Similarity Search In Animal Sound Databases.” IEEE Transactions On Multimedia, Vol. 11, No. 1, 2009, Pp 68-76, https://doi.org/10.1109/TMM.2008.2008920
2. Che Yong Yeo, S. A. R. Al-Haddad and Chee Kyun Ng, "Dog voice identification (ID) for detection system," 2012 Second International Conference on Digital Information Processing and Communications (ICDIPC), Klaipeda, Lithuania, 2012, pp. 120-123, https://doi.org/10.1109/ICDIPC.2012.6257264
3. Che Yong Yeo, S. A. R. Al-Haddad and C. K. Ng, "Animal voice recognition for identification (ID) detection system," 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, Penang, Malaysia, 2011, pp. 198-201, https://doi.org/10.1109/CSPA.2011.5759872
4. Ying Li, Zhibin Wu. “Animal Sound Recognition Based on Double Feature of Spectrogram in Real Environment.” Chinese Journal of Electronics Vol.28, No.4, July 2019. https://doi.org/10.1109/WCSP.2015.7341003
5. N. Lin, H. Sun and X. -P. Zhang, "Overlapping Animal Sound Classification Using Sparse Representation," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 2156-2160, https://doi.org/10.1109/ICASSP.2018.8462058
6. R. Narasimhan, X. Z. Fern and R. Raich, "Simultaneous segmentation and classification of bird song using CNN," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 146-150, https://doi.org/10.1109/ICASSP.2017.7952135

7. Tang, Y., Liu, C. And Yuan, X., “Recognition Of Bird Species With Birdsong Records Using Machine Learning Methods”. Plos One, 19(2), P.E 0297988. 2024 https://doi.org/10.1371/journal.pone.0297988
8. Zachary J. Ruff, Damon B. Lesmeister, Cara L. Appel, Christopher M. Sullivan, “Workflow and convolutional neural network for automated identification of animal sounds.” Ecological Indicators, Volume 124, 2021, 107419, ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2021.107419
9. Yuren Sun, Tatiana Midori Maeda, Claudia Solís-Lemus, Daniel Pimentel-Alarcón, Zuzana Bunvalova, “Classification of animal sounds in a hyperdiverse rainforest using convolutional neural networks with data augmentation.” Ecological Indicators, Volume 145, 2022, 109621, ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2022.109621
10. E. Şaşmaz and F. B. Tek, "Animal Sound Classification Using A Convolutional Neural Network," 2018 3rd International Conference on Computer Science and Engineering (UBMK), Sarajevo, Bosnia and Herzegovina, 2018, pp. 625-629, https://doi.org/10.1109/UBMK.2018.8566449
11. Singh, N., “Classification Of Animal Sound Using Convolutional Neural Network.” Masters Dissertation. Technological University Dublin. 2020 https://doi.org/10.21427/7pb8-9409
12. Pan, W., Li, H., Zhou, X., Jiao, J., Zhu, C. And Zhang, Q., “Research On Pig Sound Recognition Based On Deep Neural Network And Hidden Markov Models. Sensors” 24(4), 2024 P.1269. https://doi.org/10.3390/s24041269
13. Das, N., Padhy, N., Dey, N., Paul, H. And Chowdhury, S.,. “Exploring Explainable AI Methods For Bird Sound-Based Species Recognition Systems.” Multimedia Tools And Applications, 2024 Pp.1-31. https://doi.org/10.1007/s11042-023-17982-3
14. Bhumika, K., Radhika, G. And Ellaji, C.H.,. “Detection Of Animal Intrusion Using CNN And Image Processing.” World Journal Of Advanced Research And Reviews, 16(3), (2022) Pp.767-774. https://doi.org/10.30574/wjarr.2022.16.3.1393