Proceedings Paper DOI: 10.58190/icontas.2023.72

PROCEEDINGS OF

INTERNATIONAL CONFERENCE ON NEW TRENDS IN APPLIED SCIENCES

https://proceedings.icontas.org/

International Conference on New Trends in Applied Sciences (ICONTAS'23), Konya, December 1-3, 2023.

SPECTROGRAM BASED CLASSIFICATION OF ANIMAL SOUNDS WITH DEEP LEARNING AND MACHINE LEARNING METHODS

Omer UCKUN [1]

[1] *Graduate School of Natural and Applied Sciences, Selcuk University, Konya, TURKEY*

ORCID: 0000-0002-7402-828X

ouckun@selcuk.edu.tr,

Ilker Ali OZKAN [2]

[2] *Department of Computer Engineering, Selcuk University, Konya, TURKEY*

ORCID: 0000-0002-5715-1040

ilkerozkan@selcuk.edu.tr

ABSTRACT: The classification of animal sounds is extremely important in fields such as biology, ecology, and environmental science. These classification studies provide great convenience in determining animal populations, monitoring their behavior and defining species. While traditional methods often analyze audio features, deep learning and machine learning techniques have the ability to identify and understand more complex audio features using spectrograms. Deep learning and machine learning can work on spectrograms of audio data. These techniques offer a powerful approach to identifying and classifying animal sounds with more complex characteristics. This approach provides a faster and more automated sorting process than traditional methods, which are often time-consuming and complex. The aim of the study is to develop a model that can automatically classify animal sounds obtained from natural environments without human intervention. In this study, spectrogram images of 577 3-class animal voice data were used. Classification was made with machine learning methods, extracting features with Squeezenet and Inception v3 deep learning methods. The Support Vector Machines (SVM) machine learning model has distinguished animal sounds at a higher rate.

Key words: animal sound classification, deep learning, machine learning, spectrogram, sound analysis.

INTRODUCTION

Animals are an important part of the world's biodiversity. According to estimates, it is estimated that there are about 2.1 million species of animals in the world (U.S & Agarwal, 2015) . These species have different shapes, sizes, colors, sounds, and other characteristics. Each animal has its own unique sound and frequency. By listening to these sounds, people have the ability to determine, in some cases, which animal it belongs to. People with a lot of ear familiarity can recognize these sounds.

Animals communicate by making different sounds in various situations. Conditions such as state of activity, health, hunger, heat, illness and even pregnancy can be expressed through the sounds that animals make (Yeon et al., 2006). These sounds are used by veterinarians and specialists to assess the health status of animals, diagnose a specific condition, or have information about the general condition of the animal. For example, a cow might produce a distinct sound when unwell, or a cat could emit a tone different from its usual meowing when it is sick. These sounds can provide important clues to experts on animal health and welfare and help make the right intervention. In this way, animal sounds serve as an important communication tool for diagnosis and due diligence. Topics such as understanding animal behavior, monitoring of natural life, cultural and scientific research, and the recognition and protection of animals have revealed situations that require the discrimination of animal sounds.

Deep learning is a type of machine learning that is a sub-branch of artificial intelligence. Basically, it focuses on multi-layered artificial neural networks that can operate on large data sets in a structured and hierarchical manner. Deep learning has the capacity to automatically identify patterns and structures in complex data sets. In particular, multilayer neural networks learn representations at different levels when analyzing input data and can make sense of deep structures within the data. With appropriate data and computational facilities, deep learning methods can work with high accuracy in voice and image recognition, natural language processing, recommendation systems, game strategies and many other areas. Deep learning algorithms have been successfully applied to various data types and domains. In contemporary times, some of these domains include image processing (Akhtar & Mian, 2018), speech processing (Sprengel et al., 2016), machine translation (Cho et al., 2014), and natural language processing (Young et al., 2018). Deep learning is also employed in other fields. For instance, in recent years, deep learning algorithms have been utilized for cancer diagnosis (Pacal et al., 2022), acoustic surveillance (Aide et al., 2013), and the preservation of biological diversity (Salamon et al., 2016) (Bayat and Light, 2020).

Convolutional neural networks (CNNs) are deep learning models that mimic the human visual system and are widely used, especially in the field of image recognition. For audio signals, mel-spectrograms obtained by applying the Fourier transform are similar to visual data. Thanks to this similarity, CNN architecture is successfully used on mel-spectrograms (Thornton, 2019).

The first stage of diagnostic procedures to be performed with animal sounds is the successful differentiation of animal sounds of various species from each other. In this way, the sounds of different animals in the same environment can be identified automatically and precisely. In this study, convolutional neural network (CNN) architectures such as SqueezeNet and Inception v3 were used to extract features from spectrogram images of sounds belonging to different animal species with 3 classes. Then, it was aimed to determine which method can distinguish sounds most effectively and accurately by using machine learning algorithms such as K-NN (Nearest Neighbor), SVM (Support Vector Machines), ANN (Artificial Neural Networks), Logistic Regression and Random Forest to make predictions. Different machine learning algorithms were evaluated for the classification and recognition of audio data.

RELATED WORKS

Artificial intelligence is widely used in sound recognition. It is used in applications such as the recognition and detection of animal sounds, the identification of species, the monitoring of migratory movements and the protection of their habitats.

The convolutional neural network (CNN) was trained on spectrograms in this study. The dataset was partitioned into 5-second segments, and mel-scale spectrograms were extracted for each segment. Additionally, data augmentation was applied to all training examples, incorporating pitch and time shifts. The developed network architecture achieved an average precision score of 0.686 in predicting the main bird species for each sound file. The winner of the international BirdCLEF 2016 Competition was this approach.

(Kahl et al., 2017) conducted an investigation that explores the application of convolutional neural networks (CNNs) in voice recognition. This study specifically addresses the utilization of CNNs for the classification of 1500 distinct bird species from the Xeno-Canto database. The CNN model employed in the study featured dimensions of 512x256 pixels and was trained using spectrogram images, each lasting five seconds. The findings reveal that employing larger-sized spectrograms in the categorization process leads to superior outcomes.

(Grill & Schluter, 2017) investigated by comparing two different approaches used to detect bird sounds in sound recordings. These methods involve using mel spectrograms with convolutional neural networks. In the test set, an Area Under Curve (AUC) measure of 89% was achieved in the difficulty of signal processing performed on audio data from environmental sources.

The study conducted by (Salamon et al., 2017) aimed to categorize migratory birds' flight calls. Deep convolutional neural network combined with "shallow learning" and data augmentation research is being examined. A dataset containing 5428 flight calls from 43 different bird species was evaluated. Comparing both models to the MFCC baseline, the results were noticeably better. A similar performance was obtained between the models. The average classification accuracy was 96%.

(Bayat and Light, 2020) used deep learning techniques to recognize the sounds of bird species that are frequently observed in Iğdır Aras River Bird Sanctuary. This research includes acoustic observation studies in order to study and analyze biodiversity. Various analyzes were made on the raw audio recordings recorded through passive listener/recorder devices and these audio recordings were classified according to bird species using deep learning architectures. He dealt with the sound recordings of 22 bird species that are frequently observed in Aras Bird Sanctuary. The recorded sounds were divided into 10-second segments, and each track was converted into one-second log-mel spectrograms. Convolutional Neural Networks (CNN), Long-Short-Term Memory Neural Networks (LSTM) and Learning Transfer methods were used for classification. The input layers of the classifiers were created by extracting the high-level attribute vectors of the sounds from the pre-trained VGGish and YAMNet models for learning transfer. As a result of the experiments, accuracy rates and F1 scores were evaluated in sound recording classifications performed on four different architectures. The findings showed that the highest accuracy rate obtained with the classifier using the VGGish model was 94.2%, and the F1 score was 92.8%.

Convolutional neural networks (CNN) have made great advances in the field of voice recognition in recent years. CNNs are very successful in distinguishing different sounds from each other by learning the frequency, time, and spectral characteristics of sounds. One of the key advantages of using CNNs in voice recognition is that it replaces previously used methods. For example, methods such as Mel kepstrum coefficients and support vector machines were widely used in voice recognition before the advent of CNNs. However, CNNs have significantly outperformed the performance of these methods. Table 1 presents a comparative overview of recent studies pertaining to the classification of animal sounds.

Table 1. Comparison of related works

Authors	Dataset	Techniques	Results
(Á et al., 2018)	Xeno-canto bird songs	CNNs	Accuracy-0.85
(Bayat et al. 2020)	22 species of birds	CNN, LSTM, VGGish and YAMNET learning transfer	Accuracy: 94.2% F1 score - 92.8%
(Budiman et al., 2022)	400 bird species images	K-Nearest Neighbors	Accuracy: 95.5%
(Mhatre & Bhattacharjee, 2018)	10 species of birds	K-Nearest Neighbors	Accuracy - 82%
(Permana et al., 2022)	14 species of birds	CNNs	Accuracy: 96.45%
(Shriharsha et al. 2020)	20 species of birds	CNNs	Accuracy - 98%
(Xie & Zhu, 2019)	14 species of birds	CNNs	F1 rating - 95.95%
(Wang et al., 2022)	264 species of birds	LSTM	77.43% mAP
(Zhang et al., 2021)	18 species of birds	CNNs	Accuracy: 91.4%

These findings highlight the growing trend of deep learning-based research in the field of animal sound categorization. Furthermore, they show that they provide better results than traditional machine learning approaches.

MATERIALS AND METHODS

In this study, deep learning and machine learning were applied for the purpose of categorizing a spectrogram dataset containing animal sounds, specifically focusing on the classification of birds, cats, and dogs. The procedural flow of the study is illustrated in Figure 1. SqueezeNet and Inception v3 were the deep learning models used in this study.The classification of animal sounds was accomplished through the incorporation of various machine learning methodologies. These include, K-Nearest Neighbors (K-NN), Support Vector Machines (SVM), Artificial Neural Networks (ANN), Logistic Regression (LR), and Random Forests (RF).

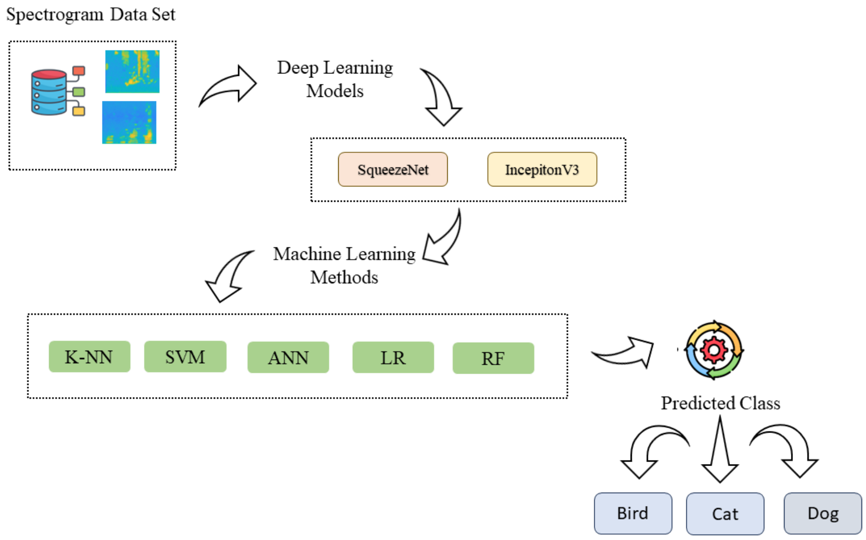
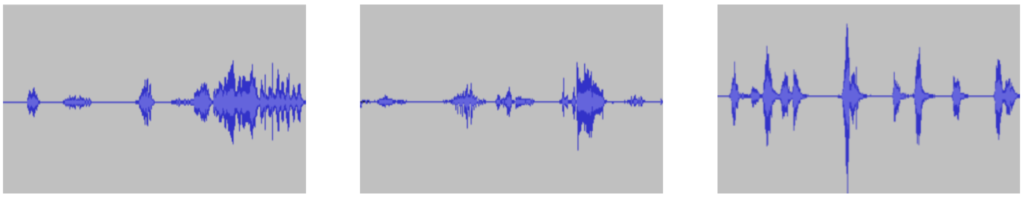


Figure 1. Flow Chart of the Classification Model

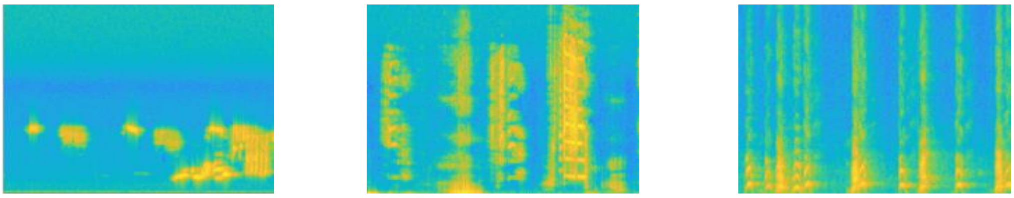
Animal Sounds Dataset

In this study, a dataset of animal sounds compiled by (Takahashi et al., 2016). This dataset included audio recordings ranging in length from 1 to 17 seconds and categorized as birds, cats, and dogs. The presence of noise in the audio files was mitigated through the utilization of the Audacity program, and the recordings, each lasting no more than 5 seconds, were divided into 5-second chunks. The resulting audio files were then converted into spectrogram images using codes specially developed in the MATLAB program. Figure 2 and Figure 3 shows representative bird sound files and its corresponding spectrogram images. Figure 4 displays the distribution of datasets classified according to the given animal classes.



a) Bird b) Cat c) Dog

Figure 2. 5-second images of audio files



a) Bird b) Cat c) Dog

Figure 3. Spectrogram images of audio files

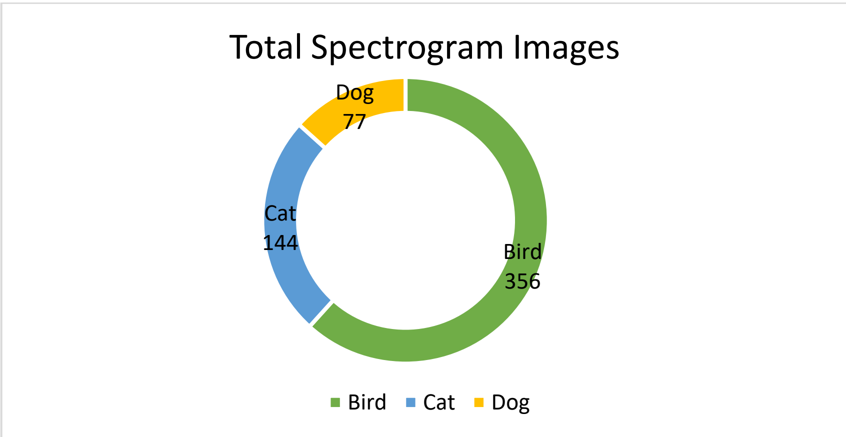


Figure 4. Dataset distribution by categories

Deep Learning Models

Convolutional Neural Networks (CNNs) used in image processing and analysis are artificial neural network structures specifically designed for the analysis of image data represented in the form of matrices formed by pixels. CNNs are notable for their ability to maintain correlation between pixels while processing the image data they receive as input (Singh et al., 2022). This capability facilitates the execution of mathematical operations encompassing convolution, pooling, and classification stages, thereby enabling the identification and detection of salient features within the input image (Taspinar et al., 2022; Unal et al., 2022). Within the scope of this study, two different models were emphasized and it was envisaged to examine these models. The convolution step is used to define the features in the input image, and in this study, two different models are mentioned that we will apply.

a) **SqueezeNet Model:** SqueezeNet is a very compact CNN architecture that uses fewer parameters compared to other CNN models. This architecture includes 15 layers of 5 different types: 2 convolution layers, 3 maximum pooling layers, 8 "fire" layers, 1 global average pooling layer, and 1 softmax layer, which is an output layer. Specifically, SqueezeNet consists of "wastage" layers that are compressed with only 1x1 size filters. These "wastage" layers perform data compression and expansion between convolution layers. This architecture reduces the size of models by using fewer parameters compared to others and can enable faster learning (Unal et al., 2022).

b) **Inception v3 Model:** The Inception v3 model has an architecture that works with images with a size of 299 x 299 pixels and consists of 48 layers. The architecture, which includes the symmetrical and asymmetrical combination of convolution, maximum pooling, average pooling, drop-off, and fully connected layers, is both symmetrical and asymmetrical in nature. Based on the "Inception module", the Inception-v3 architecture can capture various scales of information simultaneously using convolutions of different kernel sizes. This network uses features at multiple levels of abstraction to extract different visual patterns in the input image (Unal et al., 2022).

Machine Learning Algorithms

Machine learning is a type of algorithm that can perform specific tasks without coding instructions. This type of algorithm is a branch of science that studies algorithms and statistical models in the context of computer programs. It has several applications, such as data mining, computer vision, and predictive analytics. One of the key benefits of machine learning is the ability for algorithms to automatically perform their work after learning patterns by analyzing data. In this study, five different algorithms were selected for the classification process and the details of each algorithm were presented in the study. These algorithms are often tested on different data sets and the advantages and disadvantages of each are examined. This process is done to determine which algorithm performs better for a particular task.

- **K-NN:** A frequently used model in machine learning is the K-Nearest Neighbor (KNN) algorithm. Instead of learning the training data, this algorithm is used to perform classification tasks by memorizing this data. To classify new data points, the model identifies the closest neighbors of those data points. The K value represents the number of these neighbors and is determined when running the algorithm. The model evaluates the test data point against its nearest k neighbor and assigns it to the appropriate class based on that. KNN classifies based on similarities between data points, and this is a proximity-based model (Cinar & Koklu, 2022; Koklu et al., 2022; Koklu et al., 2021).

- **SVM:** Support Vector Machines (SVMs) were originally developed to solve binary classification problems. It classifies data by creating a hyperplane between the two classes. However, when dealing with datasets with multiple classes, more than one hyperplane may be required. In this case, more than one hyperplane is created for the classification of multiple classes. Many methods can be used to create these multiple hyperplanes. These strategies are used to identify relationships between multiple classes and to aggregate the results to obtain overall classification results. In this way, SVMs can effectively classify multi-class datasets (Taspinar et al., 2022).

- **ANN:** Generally, neural networks have three layers: an input layer that receives input data, a hidden layer that sits between the input and output layers, and an output layer that houses neurons equal to the number of classes in the problem. This model is designed to learn and generate predictions based on the connections between the input, hidden, and output layers, using input data (Koklu, 2016). The architecture of the neural network model is based on the connections between the input, hidden, and output layers created through learning from the data. These connections form the complex structures of the neural network in the data processing and learning process and enable predictions to be made (Koklu et al., 2022).

- **Logistic Regression:** Logistic Regression is a machine learning technique that allows data to be classified numerically or categorically. This process involves the use of a function called the logistic or sigmoid function. Logistic Regression can be used without the requirement of normal distribution if conclusions can be drawn from one or more variables. Instead of directly predicting outcomes, this algorithm is used to predict the probability that a particular data point belongs to a particular category. The logistic function is a mathematical function that compresses this probability between 0 and 1, thus allowing to determine a threshold value for the classification process. In this way, the probability that a data point belongs to a particular category can be calculated and then classified (Ahmed et al., 2020; Taspinar et al., 2022).

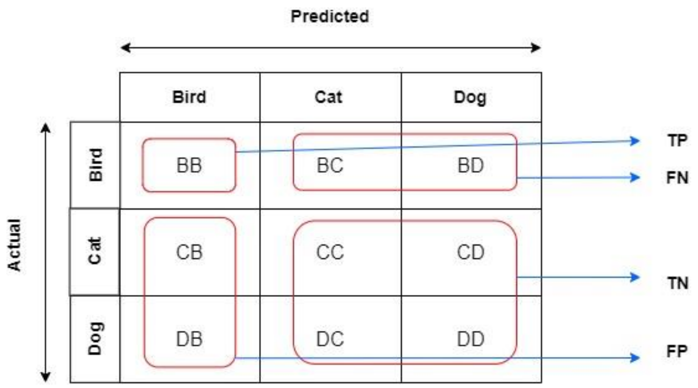
- **Random Forest:** Random Forest (RF) algorithm, is an algorithm that combines the classifications provided by multiple interrelated Decision Trees. The RF takes a classification from each decision tree for the input data and creates a structure consisting of a combination of these classifications. Within this structure, the RF algorithm determines which classification is the most popular and makes new classifications. The RF performs a "voting" process based on classifications from decision trees. It aggregates the classification results provided by each tree and selects the most preferred class as a result of this voting. This algorithm is known for its capacity to handle large datasets and is also highly effective in predicting incomplete data. RF can deal with missing data and use other information from the dataset to fill in those gaps. This feature allows RF to be used effectively in cases where there are deficiencies in the data set (Cinar & Koklu, 2019).

Confusion Matrix

Confusion matrices show how well a classification model performs by comparing predicted and actual values (Butuner et al., 2023). The confusion matrix is an important tool for evaluating the performance of classification models. By visualizing the relationship between actual and predicted classifications, it helps us understand how accurately the model predicted which classes and made errors in which classes. Table 2 shows the confusion matrix table for 3 class types (Wabang et al., 2022).
The confusion matrix usually consists of 4 parts:
True Positive (TP): Refers to situations in which samples that are actually positive are correctly predicted as positive.
False Positive (FP): Refers to situations where samples that are actually negative are falsely predicted as positive.
True Negative (TN): Refers to situations in which samples that are actually negative are correctly predicted as negative.
False Negative (FN): Refers to situations in which samples that are actually positive are falsely predicted as negative.

The confusion matrix is just one of many model evaluation and improvement methods used to identify the weaknesses of the model and optimize classification performance. This valuable tool provides the ability to drill down into the classification results to understand the overall performance of the model. This is an important step in the process of developing and improving classification models.

Table 2. 3x3 Confusion matrix



Accuracy measures the percentage of correct predictions within the total number of predictions. It's a commonly used measure of performance, but it may not always be the optimal measure of model performance, especially when data is unbalanced or the cost of misclassification isn't equal for all classes (Butuner et al., 2023; Dogan et al., 2023).

Accuracy = (TP + TN) / (TP + FP + TN + FN)×100 (1)

Precision measures the proportion of true positive predictions relative to all positive predictions generated by the model. It measures the model's ability to accurately identify the positive class. Precision Formula:

Precision = TP / (TP + FP) (2)

The analysis of the recall of a dataset is also referred to as sensitivity or true positive rate; This is the percentage of true positive predictions compared to all true positive samples in the dataset. It measures the model's ability to accurately identify the positive class.

Recall=TP/ (TP+FN) (3)

The F1 score is a statistical measure that combines precision and recall in a balanced way using the harmonic mean. It is frequently used as the sole performance measure to evaluate the effectiveness of classification models. F1 score formula:

F1-score=2 ×(Precision ×Recall)/(Precision+Recall) (4)

EXPERIMENTAL RESULTS

In this study, the classification of the features obtained after using SqueezeNet and Inception v3 deep learning methods is included here. The proposed study employs the Orange data mining program (Demšar et al., 2013). The research used a computer equipped with an Intel(R) Xeon(R) Gold 6226R CPU processor clocked at 2.90 GHz and 32 GB of RAM. Table 3 shows the parameters of k-NN, SVM, ANN, LR, and RF machine learning methods used in this study. Different parameters were assigned to each machine learning algorithm, and their respective values are detailed in the table 3. The chosen parameters for the machine learning techniques are based on the default settings provided in the Orange Data Mining tool (Atalan, 2022; Itsari & Budi, 2022).

Table 3. Machine learning model parameters

Models	**Parameters**
k-NN	Number of neighbors: 5 -Metric: Euclide-Weight: Uniform
SVM	Cost(C): 1.00- Regression loss epsilon(Ɛ): 0.10-Numerical tolerance: 0.0010-Iteration limit: 100
ANN	Neurons in hidden layers: 100-Activation: ReLu-Solver: Adam Edit, α= 0.0001 - Maximum number of iterations: 200
LR	Arrangement type: Ridge (L2)- Power C= 1
RF	Number of trees: 10 - Do not divide subsets smaller than: 5

Analysis of spectrogram data often forms an important component in audio or signal processing projects. Processing and classifying data sets requires the use of a variety of machine learning techniques. In this context, in the study carried out on spectrogram data, 20% of the data set was reserved for testing, while the remaining 80% was used for training. Table 4 shows the confusion matrix results of SqueezeNet and Table 5 shows the confusion matrix results of Inception V3 deep learning models.

Table 4. Confusion Matrix of the SqueezeNet Model

k-NN		Predicted		
		Bird	Cat	Dog
Actual	Bird	691	23	6
	Cat	13	273	4
	Dog	4	21	125

SVM		Predicted		
		Bird	Cat	Dog
Actual	Bird	704	12	4
	Cat	19	263	8
	Dog	7	16	127

ANN		Predicted		
		Bird	Cat	Dog
Actual	Bird	691	18	11
	Cat	16	257	7
	Dog	6	16	128

LR		Predicted		
		Bird	Cat	Dog
Actual	Bird	698	15	7
	Cat	25	255	10
	Dog	5	20	125

RF		Predicted		
		Bird	Cat	Dog
Actual	Bird	689	28	3
	Cat	60	215	15
	Dog	11	24	115

Table 5. Confusion Matrix of the Inception v3 Model

k-NN		Predicted		
		Bird	Cat	Dog
Actual	Bird	702	16	2
	Cat	45	225	20
	Dog	9	10	131

SVM		Predicted		
		Bird	Cat	Dog
Actual	Bird	702	15	3
	Cat	29	252	9
	Dog	8	8	134

ANN		Predicted		
		Bird	Cat	Dog
Actual	Bird	689	23	8
	Cat	27	247	16
	Dog	7	9	134

LR		Predicted		
		Bird	Cat	Dog
Actual	Bird	704	16	0
	Cat	25	250	15
	Dog	8	11	131

RF		Predicted		
		Bird	Cat	Dog
Actual	Bird	687	26	7
	Cat	76	203	11
	Dog	19	28	103

The SqueezeNet model gave more successful results when used in conjunction with the SVM method. The SVM method correctly classified 704 of the bird spectrogram images, and misclassified 12 as cats and 4 as dogs, while the remaining 16 actually belonged to the bird data type. Among the cat spectrogram images, 263 were classified as correct prediction, while the remaining 27 actually belonged to the cat data type, while 19 were misclassified as birds and 8 as dogs. Regarding the dog spectrogram images, 127 were correctly classified, while the remaining 23 actually belonged to the dog data type, while 7 were misclassified as birds and 16 as cats.

The Inception v3 model gave more successful results when used in conjunction with the Logistic Regression (LR) method. The LR method correctly classified 704 of the bird spectrogram images, and misclassified 16 as cats, while the remaining 16 actually belonged to the bird data type. Among the cat spectrogram images, 250 were classified as correct predictions, while the remaining 40 actually belonged to the cat data type, while 25 were misclassified as birds and 15 as dogs. Regarding the dog spectrogram images, 131 were correctly classified, while the remaining 19 actually belonged to the dog data type, while 8 were misclassified as birds and 11 as cats.

In Table 6, the results obtained by using machine learning methods together with the SqueezeNet model were obtained with the SVM method compared to other machine learning methods and classified the spectrogram images with a higher accuracy rate. In Table 7, the Inception v3 model, the LR machine learning method obtained better results than other methods and classified spectrogram images with a higher accuracy rate.

Table 6. Performance measurement algorithms with the SqueezeNet Model

Algorithms	Accuracy (%)	Precision	Recall	F-1 score
k-NN	94.5	0.941	0.939	0.939
SVM	**99.1**	**0.943**	**0.943**	**0.943**
Random Forest	96.2	0.876	0.878	0.876
ANN	98.5	0.927	0.928	0.927
Logistic Regression	98.8	0.929	0.929	0.929

Table 7. Performance measurement algorithms with the InceptionV3 Model

Algorithms	Accuracy (%)	Precision	Recall	F-1 score
k-NN	90.7	0.911	0.912	0.910
SVM	99	0.937	0.938	0.937
Random Forest	94.9	0.855	0.856	0.852
ANN	98.6	0.922	0.922	0.922
Logistic Regression	**99.1**	**0.935**	**0.935**	**0.935**

CONCLUSION

In this study, a data set consisting of spectrogram images of three different animal species was used to determine whether they belong to bird, cat or dog species. SqueezeNet and Inception v3 were preferred as deep learning models, and various machine learning methods were applied for each model (k-NN, SVM, ANN and LR). Deep learning models were used to extract the features of the data set, and the results obtained were then trained with machine learning methods. The SVM method on the SqueezeNet model provided a 99.1% success rate compared to other methods, and the Logistic Regression method achieved a 99.1% success rate for the Inception v3 model compared to other methods. The results of the study show that the SVM method has achieved a high rate of success, especially on the SqueezeNet model, and the Logistic Regression method has achieved a similar success in the Inception v3 model. These findings suggest that it may be more effective to use deep learning models and traditional machine learning methods together to successfully classify animal species through spectrogram images.

These findings have demonstrated the potential of deep learning and machine learning techniques that can provide a faster and more automated classification process than traditional methods. However, in the continuation of the study, it is aimed to use more data sets to increase the accuracy rates of deep learning models with larger data sets. Such techniques have great potential to obtain more precise results in important applications in fields such as biology, ecology, and environmental science, such as identifying animal species and monitoring their populations.

REFERENCES

Á, I., Jancsó, H. B., Szilágyi, Z., Farkas, A., & Sulyok, C. (2018, 13-15 Sept. 2018). Bird Sound Recognition Using a Convolutional Neural Network. 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY),

Ahmed, A., Jalal, A., & Kim, K. (2020). A Novel Statistical Method for Scene Classification Based on Multi-Object Categorization and Logistic Regression. *Sensors*, *20*(14), 3871. https://www.mdpi.com/1424-8220/20/14/3871

Aide, T. M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., & Alvarez, R. (2013). Real-time bioacoustics monitoring and automated species identification. *PeerJ*, *1*, e103. https://doi.org/10.7717/peerj.103

Akhtar, N., & Mian, A. (2018). Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey. *Ieee Access*, *6*, 14410-14430. https://doi.org/10.1109/Access.2018.2807385

Atalan, A. (2022). Forecasting drinking milk price based on economic, social, and environmental factors using machine learning algorithms. *Agribusiness*, *39*. https://doi.org/10.1002/agr.21773

Budiman, I., Ramdania, D. R., Gerhana, Y. A., Putra, A. R. P., Faizah, N. N., & Harika, M. (2022, 20-21 Sept. 2022). Classification of Bird Species using K-Nearest Neighbor Algorithm. 2022 10th International Conference on Cyber and IT Service Management (CITSM),

Butuner, R., Cinar, I., Taspinar, Y. S., Kursun, R., Calp, M. H., & Koklu, M. (2023). Classification of deep image features of lentil varieties with machine learning techniques. *European Food Research and Technology*, *249*(5), 1303-1316. https://doi.org/10.1007/s00217-023-04214-z

Cinar, I., & Koklu, M. (2019). Classification of Rice Varieties Using Artificial Intelligence Methods. *International Journal of Intelligent Systems and Applications in Engineering*, *7*, 188-194. https://doi.org/10.18201/ijisae.2019355381

Cinar, I., & Koklu, M. (2022). *IDENTIFICATION OF RICE VARIETIES USING MACHINE LEARNING ALGORITHMS*. https://doi.org/10.15832/ankutbd.862482

Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., & Starič, A. (2013). Orange: data mining toolbox in Python. *the Journal of machine Learning research*, *14*(1), 2349-2353.

Dogan, M., Taspinar, Y. S., Cinar, I., Kursun, R., Ozkan, I. A., & Koklu, M. (2023). Dry bean cultivars classification using deep cnn features and salp swarm algorithm based extreme learning machine. *Computers and Electronics in Agriculture*, *204*, 107575. https://doi.org/https://doi.org/10.1016/j.compag.2022.107575

Grill, T., & Schluter, J. (2017). *Two convolutional neural networks for bird detection in audio signals*. https://doi.org/10.23919/EUSIPCO.2017.8081512

Itsari, M. Y. I., & Budi, I. (2022, 24-25 Aug. 2022). Classification of Complaint Categories in E-Commerce: A Case Study of PT Bukalapak. 2022 5th International Conference on Information and Communications Technology (ICOIACT),

Kahl, S., Wilhelm-Stein, T., Hussein, H., Klinck, H., Kowerko, D., Ritter, M., & Eibl, M. (2017). *Large-Scale Bird Sound Classification using Convolutional Neural Networks*.

Koklu, M. (2016). Estimation of Credit Card Customers Payment Status by Using kNN and MLP. *International Journal of Intelligent Systems and Applications in Engineering*, *4*, 249-251. https://doi.org/10.18201/ijisae.2016SpecialIssue-146983

Koklu, M., Cinar, I., Taspinar, Y. S., & Kursun, R. (2022, 7-10 June 2022). Identification of Sheep Breeds by CNN-Based Pre-Trained Inceptionv3 Model. 2022 11th Mediterranean Conference on Embedded Computing (MECO),

Koklu, M., Kursun, R., Taspinar, Y. S., & Cinar, I. (2021). Classification of Date Fruits into Genetic Varieties Using Image Analysis. *Mathematical Problems in Engineering*, *2021*, 4793293. https://doi.org/10.1155/2021/4793293

Mhatre, M., & Bhattacharjee, S. (2018). Birds Voice Classification using ResNet.

Pacal, I., Karaman, A., Karaboga, D., Akay, B., Basturk, A., Nalbantoglu, U., & Coskun, S. (2022). An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets. *Comput Biol Med*, *141*, 105031. https://doi.org/10.1016/j.compbiomed.2021.105031

Permana, S. D. H., Saputra, G., Arifitama, B., Yaddarabullah, Caesarendra, W., & Rahim, R. (2022). Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm. *Journal of King Saud University - Computer and Information Sciences*, *34*(7), 4345-4357. https://doi.org/https://doi.org/10.1016/j.jksuci.2021.04.013

Salamon, J., Bello, J. P., Farnsworth, A., & Kelling, S. (2017, 5-9 March 2017). Fusing shallow and deep learning for bioacoustic bird species classification. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),

Salamon, J., Bello, J. P., Farnsworth, A., Robbins, M., Keen, S., Klinck, H., & Kelling, S. (2016). Towards the Automatic Classification of Avian Flight Calls for Bioacoustic Monitoring. *PLoS One*, *11*(11), e0166866. https://doi.org/10.1371/journal.pone.0166866

Singh, D., Taspinar, Y., Kursun, R., Cinar, I., Koklu, M., Ozkan, I. A., & Lee, H.-N. (2022). Classification and Analysis of Pistachio Species with Pre-Trained Deep Learning Models. *Electronics*, *11*, 1-14. https://doi.org/10.3390/electronics11070981

Sprengel, E., Jaggi, M., Kilcher, Y., & Hofmann, T. (2016). Audio Based Bird Species Identification using Deep Learning Techniques. Conference and Labs of the Evaluation Forum,

Takahashi, N., Gygli, M., Pfister, B., & Van Gool, L. (2016). *Deep Convolutional Neural Networks and Data Augmentation for Acoustic Event Recognition*. https://doi.org/10.21437/Interspeech.2016-805

Taspinar, Y., Doğan, M., Cinar, I., Kursun, R., Ozkan, I. A., & Koklu, M. (2022). Computer vision classification of dry beans (Phaseolus vulgaris L.) based on deep transfer learning techniques. *European Food Research and Technology*, *248*. https://doi.org/10.1007/s00217-022-04080-1

Thornton, B. (2019). Audio recognition using mel spectrograms and convolution neural networks.

U.S, R., & Agarwal, N. (2015). Biodiversity: Concept, Threats and Conservation. *Environment Conservation Journal*, *16*, 19-28. https://doi.org/10.36953/ECJ.2015.16303

Unal, Y., Taspinar, Y. S., Cinar, I., Kursun, R., & Koklu, M. (2022). Application of Pre-Trained Deep Convolutional Neural Networks for Coffee Beans Species Detection. *Food Analytical Methods*, *15*(12), 3232-3243. https://doi.org/10.1007/s12161-022-02362-8

Wabang, K., Nurhayati, O. D., & Farikhin. (2022). Application of The Naïve Bayes Classifier Algorithm to Classify Community Complaints. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*.

Wang, H., Xu, Y., Yu, Y., Lin, Y., & Ran, J. (2022). An Efficient Model for a Vast Number of Bird Species Identification Based on Acoustic Features. *Animals*, *12*(18), 2434. https://www.mdpi.com/2076-2615/12/18/2434

Xie, J., & Zhu, M. (2019). Handcrafted features and late fusion with deep learning for bird sound classification. *Ecological Informatics*, *52*, 74-81. https://doi.org/https://doi.org/10.1016/j.ecoinf.2019.05.007

Yeon, S. C., Jeon, J. H., Houpt, K. A., Chang, H. H., Lee, H. C., & Lee, H. J. (2006). Acoustic features of vocalizations of Korean native cows (Bos taurus coreanea) in two different conditions. *Applied animal behaviour science*, *101*(1-2), 1-9.

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. *IEEE Computational Intelligence Magazine*, *13*, 55-75. https://doi.org/10.1109/MCI.2018.2840738

Zhang, F., Zhang, L., Chen, H., & Xie, J. (2021). Bird species identification using spectrogram based on multi-channel fusion of DCNNs. *Entropy*, *23*(11), 1507.