Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 418–428

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Environmental Sound Classification using Hybrid Ensemble Model

Anam Bansal[a], Naresh Kumar Garg[b]

Department of Computer Science and Engineering, GZS Campus College of Engineering and Technology, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

[a]*Research Scholar*
[b]*Professor*

Abstract

Environmental sound classification(ESC) is the most trending research areas. The sounds in the surroundings such as screaming, air conditioners, and rain droplets can help in the development of context-aware applications. It is complex to process the environmental sounds as compared to speech and music due to the unstructured essence of environmental sounds. In the past, certain preprocessing techniques, feature extraction, and classification algorithms are used for ESC. Several researchers have applied machine learning classifiers for ESC and certain ensemble classifiers are also used but the accuracy can be increased if instead of combining homogeneous classifiers, heterogeneous classifiers can be ensembled. In this paper, a hybrid ensemble classifier is used for ESC on the UrbanSound8k dataset and cepstral features Mel Frequency Cepstral Coefficients are used. Five different machine learning classifiers- Decision Tree, Support Vector Machine, Logistic Regression, K- Nearest Neighbour, and Naive Bayes are used to develop a hybrid ensemble model. The highest accuracy is obtained when all the five classifiers are combined. The proposed approach gives an accuracy of 79.4% and is compared with the benchmark results using individual classifiers and the former outperforms the latter. The results of the hybrid ensemble model on the UrbanSound8K dataset are also compared with the dataset ESC-10.

Keywords:
Environmental Sound Classification; Hybrid ensemble; Machine learning; Feature extraction.

1. Introduction

Classifying the sounds such as dog barking, gunshots and jackhammers can help in recognizing certain situations and environments such as crime, offices, parks, and many more. The classification of such sounds is termed Environ-

* Anam Bansal. Tel.: +91-8437800233
E-mail address: anambansal19@gmail.com

1877-0509
10.1016/j.procs.2023.01.024

mental Sound Classification(ESC). The sound signals are better than the videos as they can be captured from every direction and are not affected by occlusion[1]. Sound signals are even cheaper to capture as compared to videos.

ESC is used in many areas such as robots can be trained to navigate by recognizing sounds[2] [3]. Hearing aids can be prepared using ESC[4] [5]. ESC helps in wildlife monitoring [6] such as classifying birds[7], frogs[8], bats [9], and many more. Old age people staying alone at home can get the advantage of ESC as smart homes can be designed [10] [11]. ESC helps in automatic audio surveillance [12].

The process of ESC has four steps- data acquisition, preprocessing, feature extraction, and classification (Fig.1). The steps are explained as follows:

- **Data acquisition:** The very first step in classification problems is data acquisition. The data can be either primary or secondary depending upon the research and application.
- **Preprocessing:** The data acquired is not suitable to be used as it is. There is certain noise or silence in parts of the audio recorded. So, the data undergo preprocessing to remove noise or silence. Noise reduction techniques [13] [14] and silence detection techniques [15] make the data amenable for feature extraction. Sometimes, the spectrograms or the dimensions of the audios captured are quite long. The dimensionality reduction techniques can help in reducing the dimensions of the audio signals[16]. The signals can be enhanced using signal enhancement techniques[17]. The preprocessed data is appropriate for extracting features.
- **Feature extraction:** The features of audios may be categorized into four sorts -temporal functions, cepstral capabilities, spatial features, and image-based features. Temporal features are time-domain functions and spatial functions are acquired from temporal capabilities through some type of modifications such as Fourier transform [18] and Discrete Chirplet transforms[19]. Cepstral functions which include Mel Frequency Cepstral Coefficients have also been used broadly for ESC [20] [21]. Some of the past studies have used image-based features such as Cross Recurrence Plots, and Mel spectrograms that are extracted from audios [22]. The advantage of image-based features is that image-based deep neural networks(DNN) can be applied to those features.
- **Classification:** Finally, the classification of features is done either using machine learning or deep learning models. Machine learning classifiers such as Naive Bayes, Support Vector Machine(SVM), K-Nearest Neighbour(KNN) are used for classification [23]. Recently, the deep learning models such as Convolutional Neural Networks(CNN)[24] [25], Long short-term neural network [26], Recurrent Neural networks(RNN)[27], and combination of different models [28] have proved to be more accurate but the drawback is that they require more data and computational power.

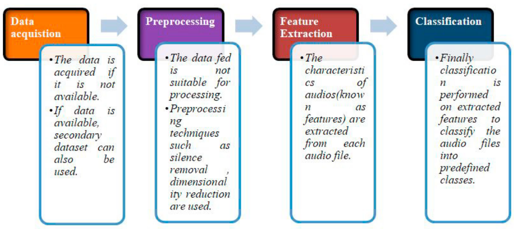

Figure 1. Process of ESC

1.1. Significance and Contributions of work

In this paper, hybrid ensemble classifiers are used in which several heterogeneous weak learners are combined to create an ensemble classifier. The approach is compared to baseline weak learners and higher accuracy is obtained in the proposed approach. The main aims of this paper are:

1. Studying and implementing the existing machine learning classifiers used for ESC.
2. Proposing a hybrid ensemble classifier for ESC.

3. Comparing the hybrid ensemble classifier with the weak learners.

The remaining paper is established as follows. In part 2, the research work performed by researchers in past for ESC is illustrated. Segment three describes the method and experimental setup. Results and analysis are finished in phase 4 and ultimately, the conclusion is presented in segment five.

2. Related Work

Several researchers have worked in the area of ESC. In a study[29], pre-trained networks Alexnet and VGG-16 are used to extract deep features from the spectrograms obtained through Optimum allocation sampling (OAS) from the ESC-10 dataset[30]. The deep features are given as input to the decision tree, linear discriminant analysis (LDA), k-nearest neighbor (KNN), bagged tree, support vector machine(SVM),and softmax classifier and an accuracy of 90.1%, 87.9%, 95.8%, 95.6%, 94.7%, and 92.4% is obtained respectively. Another usage of pre-trained neural networks for feature extraction led to accuracy scores of 94.8%, 81.4%, and 78.14% for ESC-10, ESC-50[30] and UrbanSound8K[31] datasets[32]. The feature selection is employed and the classifier used is SVM.

The computational cost for deep neural networks is quite high, so the authors in ref[33], proposed a low-cost method for ESC using SVM and the method proved to be more accurate as compared to the other approaches. In 2016[34], Gabor features are extracted from audio, and the dimensions of features are compressed using Linear Discriminant Analysis (LDA) and Principal Component Analysis(PCA). KNN is used to classify the sounds for surveillance applications and an accuracy of 96.1% is obtained.

The different features such as cepstral, temporal, and spectral features are used in the study[23]. Different machine learning classifiers k-Nearest Neighbour, Naive Bayes, Artificial Neural Network (ANN), Support Vector Machine, and Decision Tree give the accuracies of 50.48, 49.29, 50.6, 25.29, and 34.7% respectively. The authors in [35] experimented with different audio features and classifiers. KNN, SVM, Naive Bayes, Logistic Regression, C4.5,and ANN give an accuracy of 85.8, 82.63, 78.92, 76.37, 71.91, and 84.17% respectively.

The spectrogram features when fed to a feed-forward neural network give an accuracy of 85.66%[36]. The experiments are performed with the KNN classifier and a combination of features. A novel feature extraction approach named Matching Pursuit Algorithm is adopted by the researchers in the study [37] and SVM performs better as compared to the standard classifiers.

The unsupervised self-organizing map(SOM) is used for the frequency representation of audio signals[38]. For classification, the authors used event-based spiking neural network (SNN), and an accuracy of 95.1% is obtained. The study claims that the proposed network works well in noise-corrupted environments and gives accurate results even with partial presentation of inputs.

In 2022, the environmental noise in smart cities is classified into four categories- highway, railway, lawnmowers, and birds [39]. The authors collected 44 sound samples and fed them for feature extraction. MFCCs are used and then four machine learning classifiers are compared in terms of accuracy in classifying the noise. Random Forest, KNN, SVM, and bagging give the accuracy between 95-100%.

Table 1 abridges the past research in the field of ESC using machine learning classifiers.

Table 1: Literature Review in ESC

Authors	Dataset Used	Features Used	Technique	Accuracy
Silvia Liberata Ullo et al.	ESC-10	deep features extracted using pre-trained networks Alexnet and VGG-16	Decision Tree, KNN, SVM, LDA, Bagged tree	90.1%, 95.8%, 94.7%, 87.9%, 95.6% and 92.4% respectively.
Fatih Demir et al.	ESC-10, ESC-50, UrbanSound8k	SVM	features extracted from pre-trained networks	94.8%, 81.4% and 78.14% respectively.
Jia-Ching Wang et al.	Self-collected	Wavelet subspace based feature extraction	SVM	90.63%
Yingjie Li et al.	sound data from JDAE TUKE database [40]	Gabor features	KNN	96.1%
Bruno da Silva et al.	BDLib Dataset, ESC-10 Dataset and UrbanSound Dataset	spectral, temporal and cepstral features	KNN, Naive Bayes, ANN, SVM and decision tree	50.48, 49.29, 50.6, 25.29, 34.7% respectively

Authors	Dataset Used	Features Used	Technique	Accuracy
Vasileios Bountourakis et al.	BDLib	spectral, temporal, and cepstral features	KNN, Naive Bayes, SVM, C4.5 Logistic Regression, and ANN	85.8, 78.92, 82.63, 71.91, 76.37, and 84.17% respectively
Peerapol Khunarsal et al.	Twenty types of sound from BBC and Sound Ideas databases	spectrogram features	feed forward neural networks and KNN	85.66%
Yong Li et al.	sounds from the Freesound Project	Feature extraction based on Matching Pursuit Algorithm	SVM	92%
Jibin Wu et al.	RWCP environmental sound	mel features	SOM-SNN	95.1%
Yaseen Hadi Ali et al.	self-collected- 44 sound samples	MFCCs	SVM, KNN, Bagging, Random Forest	95-100%

2.1. Previous Research outcomes

Most of the researchers have used numerous techniques for ESC. Following are the research outcomes that are obtained by examination of the past literature.

1. In the past, researchers have used either self-collected datasets or smaller datasets.
2. Different features are used for ESC but MFCCs are more reliable and give higher accuracy without compromising the computational power
3. Machine learning algorithms or combinations of algorithms are used but hybrid ensemble models can give more accuracy.

3. Methodology and Experimental setup

In this section, the methodology used for the experiments is explained. The methodology can be split into the following phases:

3.1. Dataset

In this study, the secondary dataset UrbanSound8K is used. It is available openly. The dataset consists of 10 classes and 8732 recordings of four seconds each that are amassed from the free sound repository[41] and categorised manually. The 10 classes are - air conditioner, children playing, car horn, drilling, dog bark, gunshot, engine idling, street music, jackhammer, and siren. For comparison, another dataset 'ESC-10' is used. It is also benchmark dataset and available openly. ESC-10 dataset [30] consists of four hundred recordings belonging to 10 classes (rain, dog bark, child cry, sea waves, person sneeze, clock tick, helicopter, rooster, chainsaw, fire crackling). There are forty recordings of five seconds each belonging to each category.

The dataset is already preprocessed. So, the recordings need not be modified to remove noise. It is suitable for feature extraction.

3.2. Feature Extraction

The Mel Frequency Cepstral Coefficients(MFCCs) are extracted from the audios. MFCCs are cepstral features that are obtained by following a few steps:

1. The audio signals are divided into overlapping frames.
2. The frames undergo a Fourier transformation.
3. The powers of the spectrum are mapped to mel scale using triangular or cosine overlapping windows.
4. Log of the powers is taken at each mel frequency(Equation 1).
5. Finally, discrete cosine transformation of log mel powers is performed and the amplitudes of the resultant spectrum are the MFCCs.

$$Mel(f) = 2595 log(1 + (f/700)) \quad (1)$$

MFCCs are obtained through the python library function librosa.

3.3. Classification

The features are fed to a hybrid ensemble classifier. In the usual ensemble classifier, similar weak learners are combined but in hybrid ensemble classifiers, diverse weak learners are combined to get the classification result. Hence, the term hybrid is used. In this study, five weak learners- Decision Tree, Logistic Regression, K-Nearest Neighbor, Support Vector Machine, and the Naive Bayes Models are combined to get the final model and the final prediction is done using a hybrid model (Refer Fig. 2).

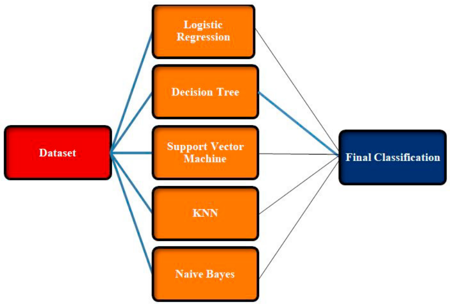

Figure 2. Illustration of the proposed method

Every weak learner is used five times. Since there are five weak learners so overall 25 weak learners are applied. The dataset is fed to the weak learners. The final result is calculated by majority voting by finding which class of the audio is predicted by the maximum number of weak learners.

The experiments are also performed by feeding the features to the five classifiers individually. The classification accuracy of a hybrid model is compared with the individual performance of weak learners in benchmark research. The proposed model is also compared with the different combinations of weak learners and the former outperforms the latter. For combinations, first, the two best performing classifiers are combined. Then the three best three classifiers are combined and finally, the best four classifiers are combined. Following combinations of weak learners are used:

1. **SVM+ Logistic Regression:** First, SVM and logistic regression are combined and the accuracy attained is compared with the individual SVM and logistic regression with the features.
2. **SVM+ Logistic Regression+ KNN:** The above model is extended by adding KNN five times and the accuracy attained is compared with the individual classifiers.
3. **SVM+ Logistic Regression+ KNN+ Naive Bayes:** Finally, the four classifiers- SVM, KNN, Logistic Regression, and Naive Bayes are combined and results are computed. The accuracy is again compared with the four individual classifiers.
4. **Hybrid Model:** In this model, all the five classifiers are combined and compared with all the combinations of weak learners. The results are also compared with the hybrid ensemble model on the ESC-10 dataset.

4. Results and Discussions

The dataset is split into train and test sizes of 0.7 and 0.3 respectively. The most important python library that is used is sklearn.

4.1. Evaluation metrics

Evaluation metrics help in evaluating the performance of the classification model. Different evaluation metrics can be used. In this paper, accuracy score and confusion matrices are used for assessing the performance of the hybrid model, individual weak learners, and combinations of weak learners.

- **Accuracy Score:** Accuracy is the proportion of the number of observations predicted correctly to the total number of observations.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{2}$$

  where TP, FN, FP, and TN are True Positive, False Negative, False Positive, and True Negative respectively [42].
- **Confusion Matrix:** The confusion matrix is a table that gives a vivid description of the performance of a model [43]. It consists of four terms- True positive, True negative, False positive, and False negative. Figure 3 illustrates the confusion matrix.

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Figure 3. Confusion Matrix

Initially, the features are fed to the five classifiers individually. Table 2 and Figure 4 show the accuracy obtained using the classifiers.

Table 2. Accuracy of different classifiers

Classifier	Accuracy
Logistic Regression	56.14%
Decision Tree	29.04%
SVM	59.2%
KNN	81.46%
Naive Bayes Model	48.10%

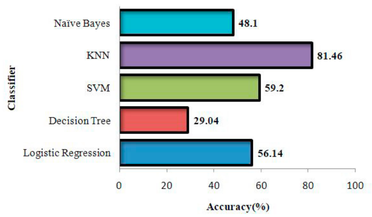

Figure 4. Accuracy of machine learning classifiers

SVM+ Logistic Regression The two most accurate classifiers- SVM and Logistic Regression are combined and an accuracy of 58.92% is obtained which is slightly less than using SVM alone and higher than Logistic Regression. Table 3 and Figure 5 illustrate the outcomes of the combination compared with the accuracy of individual classifiers.

Table 3. SVM+ Logistic Regression

Classifier	Accuracy
Logistic Regression	56.14%
SVM	59.2%
SVM+ Logistic Regression	58.92%

SVM+ Logistic Regression+ KNN: The three best performing classifiers are combined to get an accuracy of 67.6% which is greater than Logistic Regression and SVM individually but less than KNN. Table 4 and Figure 6 illustrate the results of the combination compared with the accuracy of individual classifiers.

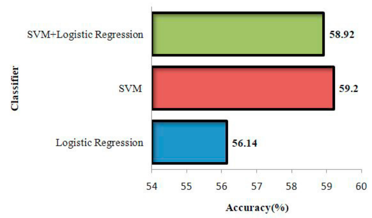

Figure 5. SVM+ Logistic Regression

Table 4. SVM+ Logistic Regression+ KNN

Classifier	Accuracy
Logistic Regression	56.14%
SVM	59.2%
KNN	81.46%
SVM+ Logistic Regression+ KNN	67.60%

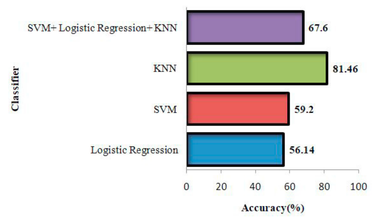

Figure 6. SVM+ Logistic Regression+ KNN

SVM+ Logistic Regression+ KNN+ Naive Bayes: Further, the Naive Bayes model is combined with the above combination and an accuracy of 67.75% is attained which is greater than all the other classifiers individually except for KNN. Table 5 and Figure 7 depict the results of the combination compared with the accuracy of individual classifiers.

Table 5. SVM+ Logistic Regression+ KNN+ Naive Bayes

Classifier	Accuracy
Logistic Regression	56.14%
SVM	59.2%
KNN	81.46%
Naive Bayes	48.10%
SVM+ Logistic Regression+ KNN+ Naive Bayes	67.75%

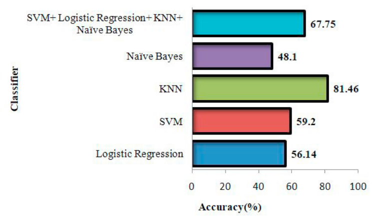

Figure 7. SVM+ Logistic Regression+ KNN+ Naive Bayes

Different combinations of weak learners differ in accuracies for ESC.

Hybrid Model: Finally, a hybrid classifier is obtained by combining all the five classifiers including the Decision Tree. The hybrid classifier with five weak learners gives an accuracy of 79.4% which is the highest among all the classifiers and combinations of classifiers. This can be verified by observing the accuracy score and confusion matrix.

Table 6 and Figure 8 describe the comparison of the accuracy of the hybrid model with all the classifiers individually.

Table 6. Comparison of hybrid model with individual classifiers

Classifier	Accuracy
Logistic Regression	56.14%
Decision Tree	29.04%
SVM	59.2%
KNN	81.46%
Naive Bayes Model	48.10%
Hybrid Model	79.40%

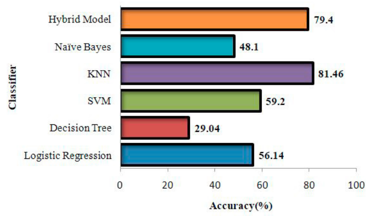

Figure 8. Comparison of hybrid model with individual classifiers

The hybrid model is also compared with the combinations in terms of accuracy. This comparison is depicted in Table 7 and Figure 9.

Table 7. Comparison of hybrid model with different combinations

Classifier	Accuracy
SVM+ Logistic Regression	558.92%
SVM+ Logistic Regression+ KNN	67.60%
SVM+ Logistic Regression+ KNN+ Naive Bayes	67.75%
Hybrid Model	79.40%

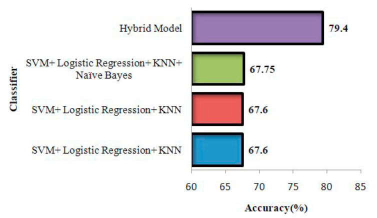

Figure 9. Comparison of hybrid model with different combinations

The result of the proposed hybrid classifier is compared with the results of the classifiers for ESC in past research. The classifiers Logistic Regression Model, Support Vector Machine, Decision Tree, Naive Bayes Model, and K-Nearest Neighbor Model are already used by the researchers for ESC on UrbanSound8K and certain accuracy is obtained. Table 8 compares the result of the hybrid ensemble classifier with other benchmark results for ESC. The pictorial representation is done in Fig. 4.1.

The hybrid ensemble model on UrbanSound8K is compared with the same model on the ESC-10 dataset. It is found that the model on the ESC-10 dataset gives an accuracy o 70% which is less than the results on the UrbanSound8K dataset. Table 9 and Figure 4.1 depict the comparison of accuracies of the hybrid ensemble model on different datasets. Figure 4.1 and Figure 4.1 illustrate the confusion matrices of the hybrid ensemble model with UrbanSound8k and ESC-10 respectively.

Table 8. Comparison of accuracy with benchmark research

Algorithm	**Accuracy**
SVM [31]	68%
Logistic Regression[44]	75.38%
RandomForest [31]	66%
Decision Tree [31]	48%
Naive Bayes [23]	53.50%
Hybrid Ensemble Model (proposed approach)	79.4%

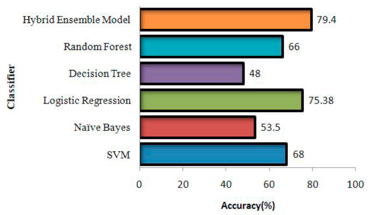

Figure 10. Comparison with benchmark results

Table 9. Comparison of accuracy of hybrid ensemble model on different datasets

Dataset	**Accuracy**
ESC-10	70%
UrbanSound8K	79.4%

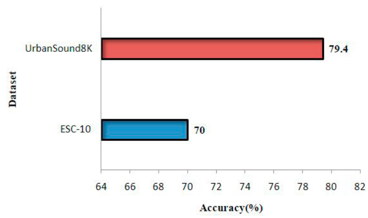

Figure 11. Comparison of hybrid ensemble model on different datasets

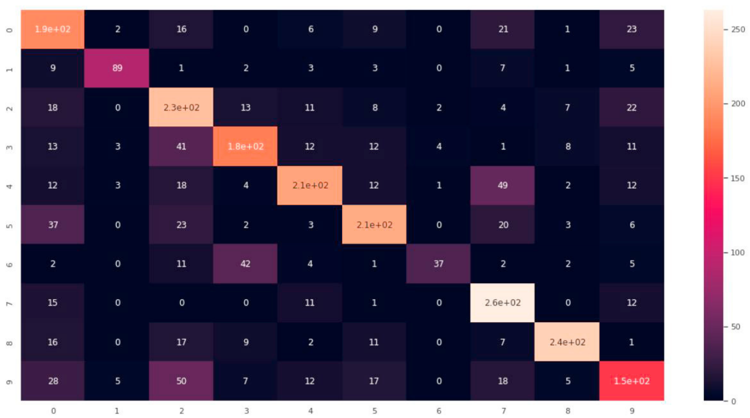

Figure 12. Confusion matrix of the hybrid model with UrbanSound8K

4.2. Analysis and Discussion

It is analyzed that the hybrid ensemble classifier performs better as compared to individual weak learners, combinations of weak learners and weak learners in the past benchmark research. An accuracy of 79.4 %, though it is less

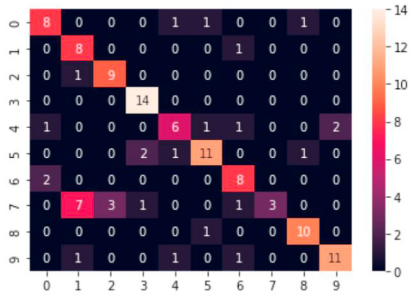

Figure 13. Confusion matrix of the hybrid model with ESC-10

as compared to recent approaches using deep learning techniques but the hybrid ensemble model is computationally less costly as it uses machine learning models only. After analyzing the literature carpingly, it is found that accuracy of almost 100% is also achieved using machine learning models but either the dataset is small or the application is different. The five machine learning classifiers which have given the best accuracy in the past are chosen for hybridization. When the dataset is changed to a smaller dataset (ESC-10), the accuracy of the model drops due to less training data. The chances of overfitting are higher in the ESC-10 dataset as the number of classes is more but the number of recordings in each class is less [45]. The accuracy of the hybrid model can further be improved by ensembling other machine learning classifiers or choosing other feature sets or combinations.

5. Conclusion

In this study, a hybrid ensemble learning model is developed for ESC. UrbanSound8K dataset is used and MFCCs features are used. Five weak learners- Logistic Regression Model, Decision Tree, Support Vector Machine, K-Nearest Neighbor Model, and the Naive Bayes Model are combined to create a hybrid ensemble model. The proposed approach for ESC gives 79.4% accuracy for classification. The proposed method is compared with the classification accuracy obtained using individual weak learners in past research and it is found that the hybrid ensemble classifier is the best among all. The proposed model is also compared with the different combinations of weak learners. Further, the variation in accuracy is observed by changing the dataset. ESC-10 dataset is used for comparison. In the future, other ensemble classifiers can be explored and other feature combinations can be tried to get better accuracy.

References

1. Bansal, Anam et al. (2018). "An off the shelf CNN features based approach for vehicle classification using acoustics". In: *International Conference on ISMAC in Computational Vision and Bio-Engineering*. Springer, pp. 1163–1170.
2. Eliakim, Itamar et al. (2018). "A fully autonomous terrestrial bat-like acoustic robot". In: *PLoS computational biology* 14.9, e1006406.
3. Song, Liuyang, Huaqing Wang, and Peng Chen (2020). "Automatic patrol and inspection method for machinery diagnosis robot—Sound signal-based fuzzy search approach". In: *IEEE Sensors Journal* 20.15, pp. 8276–8286.
4. Humes, Larry E et al. (2018). "The acoustic environments in which older adults wear their hearing aids: Insights from datalogging sound environment classification". In: *American Journal of Audiology* 27.4, pp. 594–603.
5. Habib, Md et al. (2022). "Sound classification using deep learning for hard of hearing and deaf people". PhD thesis. Brac University.
6. Thangavel, Surya and Chitra Selvi Shokkalingam (2021). "The IoT based embedded system for the detection and discrimination of animals to avoid human–wildlife conflict". In: *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17.
7. Xie, Jie et al. (2019). "Investigation of different CNN-based models for improved bird sound classification". In: *IEEE Access* 7, pp. 175353–175361.
8. Brodie, Sheryn et al. (2020). "Automated species identification of frog choruses in environmental recordings using acoustic indices". In: *Ecological Indicators* 119, p. 106852.
9. Moss, Cynthia F (2018). "Auditory mechanisms of echolocation in bats". In: *Oxford Research Encyclopedia of Neuroscience*.
10. Vafeiadis, Anastasios et al. (2020). "Audio content analysis for unobtrusive event detection in smart homes". In: *Engineering Applications of Artificial Intelligence* 89, p. 103226.
11. Udupa, Prathiba and Siva S Yellampalli (2018). "Smart home for elder care using wireless sensor". In: *Circuit World*.
12. Shreyas, N et al. (2020). "Trends of sound event recognition in audio surveillance: a recent review and study". In: *The Cognitive Approach in Cloud Computing and Internet of Things Technologies for Surveillance Tracking Systems*, pp. 95–106.
13. Awad, Ali (2019). "Impulse noise reduction in audio signal through multi-stage technique". In: *Engineering Science and Technology, an International Journal* 22.2, pp. 629–636.

14. Ni, Aoxin and Nasser Kehtarnavaz (2022). "Smartphone-Based Hearing Aid Compression and Noise Reduction". In: *Sensors* 22.9, p. 3306.
15. Chhaglani, Bhawana et al. (2022). "FlowSense: Monitoring Airflow in Building Ventilation Systems Using Audio Sensing". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6.1, pp. 1–26.
16. Esmaeilpour, Mohammad, Patrick Cardinal, and Alessandro Lameiras Koerich (2019). "A robust approach for securing audio classification against adversarial attacks". In: *IEEE Transactions on information forensics and security* 15, pp. 2147–2159.
17. Hua, Guang and Haijian Zhang (2019). "ENF signal enhancement in audio recordings". In: *IEEE Transactions on Information Forensics and Security* 15, pp. 1868–1878.
18. Salah, Euschi et al. (2021). "A Fourier transform based audio watermarking algorithm". In: *Applied Acoustics* 172, p. 107652.
19. Ghosh, Samit Kumar et al. (2020). "Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals". In: *Computers in biology and medicine* 118, p. 103632.
20. Sharma, Jivitesh, Ole-Christoffer Granmo, and Morten Goodwin (2020). "Environment Sound Classification Using Multiple Feature Channels and Attention Based Deep Convolutional Neural Network." In: *Interspeech*, pp. 1186–1190.
21. Das, Joy Krishan, Amitabha Chakrabarty, and Md Jalil Piran (2022). "Environmental sound classification using convolution neural networks with different integrated loss functions". In: *Expert Systems* 39.5, e12804.
22. Boddapati, Venkatesh et al. (2017). "Classifying environmental sounds using image recognition networks". In: *Procedia computer science* 112, pp. 2048–2056.
23. Silva, Bruno da et al. (2019). "Evaluation of classical machine learning techniques towards urban sound recognition on embedded systems". In: *Applied Sciences* 9.18, p. 3885.
24. Piczak, Karol J (2015a). "Environmental sound classification with convolutional neural networks". In: *2015 IEEE 25th international workshop on machine learning for signal processing (MLSP)*. IEEE, pp. 1–6.
25. Esmaeilpour, Mohammad, Patrick Cardinal, and Alessandro Lameiras Koerich (2022). "From environmental sound representation to robustness of 2D CNN models against adversarial attacks". In: *Applied Acoustics* 195, p. 108817.
26. Lezhenin, Iurii, Natalia Bogach, and Evgeny Pyshkin (2019). "Urban sound classification using long short-term memory neural network". In: *2019 federated conference on computer science and information systems (FedCSIS)*. IEEE, pp. 57–60.
27. Sang, Jonghee, Soomyung Park, and Junwoo Lee (2018). "Convolutional recurrent neural networks for urban sound classification using raw waveforms". In: *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, pp. 2444–2448.
28. Bahmei, Behnaz, Elina Birmingham, and Siamak Arzanpour (2022). "CNN-RNN and Data Augmentation Using Deep Convolutional Generative Adversarial Network for Environmental Sound Classification". In: *IEEE Signal Processing Letters* 29, pp. 682–686.
29. Ullo, Silvia Liberata et al. (2020). "Hybrid computerized method for environmental sound classification". In: *IEEE Access* 8, pp. 124055–124065.
30. Piczak, Karol J (2015b). "ESC: Dataset for environmental sound classification". In: *Proceedings of the 23rd ACM international conference on Multimedia*, pp. 1015–1018.
31. Salamon, Justin, Christopher Jacoby, and Juan Pablo Bello (2014). "A dataset and taxonomy for urban sound research". In: *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 1041–1044.
32. Demir, Fatih et al. (2020). "A new pyramidal concatenated CNN approach for environmental sound classification". In: *Applied Acoustics* 170, p. 107520.
33. Wang, Jia-Ching et al. (2015). "Robust environmental sound recognition with fast noise suppression for home automation". In: *IEEE Transactions on Automation Science and Engineering* 12.4, pp. 1235–1242.
34. Li, Yingjie and Gang Liu (2016). "Sound classification based on spectrogram for surveillance applications". In: *2016 IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC)*. IEEE, pp. 293–297.
35. Bountourakis, Vasileios, Lazaros Vrysis, and George Papanikolaou (2015). "Machine learning algorithms for environmental sound recognition: Towards soundscape semantics". In: *Proceedings of the Audio Mostly 2015 on Interaction With Sound*, pp. 1–7.
36. Khunarsal, Peerapol, Chidchanok Lursinsap, and Thanapant Raicharoen (2013). "Very short time environmental sound classification based on spectrogram pattern matching". In: *Information Sciences* 243, pp. 57–74.
37. Li, Yong and Ying Li (2010). "Eco-environmental sound classification based on matching pursuit and support vector machine". In: *2010 2nd International Conference on Information Engineering and Computer Science*. IEEE, pp. 1–4.
38. Wu, Jibin et al. (2018). "A spiking neural network framework for robust sound classification". In: *Frontiers in neuroscience* 12, p. 836.
39. Ali, Yaseen Hadi, Rozeha A Rashid, and Siti Zaleha Abdul Hamid (2022). "A machine learning for environmental noise classification in smart cities". In: *Indonesian Journal of Electrical Engineering and Computer Science* 25.3, pp. 1777–1786.
40. Pleva, Matús et al. (2011). "The joint database of audio events and backgrounds for monitoring of urban areas". In: *Journal of Electrical and Electronics Engineering* 4.1, p. 185.
41. Font, Frederic, Gerard Roma, and Xavier Serra (2013). "Freesound technical demo". In: *Proceedings of the 21st ACM international conference on Multimedia*, pp. 411–412.
42. Gans, Joshua S et al. (2022). "False-positive results in rapid antigen tests for SARS-CoV-2". In: *JAMA* 327.5, pp. 485–486.
43. Heydarian, Mohammadreza, Thomas E Doyle, and Reza Samavi (2022). "MLCM: multi-label confusion matrix". In: *IEEE Access* 10, pp. 19083–19095.
44. Esmaeilpour, Mohammad, Patrick Cardinal, and Alessandro Lameiras Koerich (2020). "Detection of adversarial attacks and characterization of adversarial subspace". In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 3097–3101.
45. Mushtaq, Zohaib and Shun-Feng Su (2020). "Environmental sound classification using a regularized deep convolutional neural network with data augmentation". In: *Applied Acoustics* 167, p. 107389.