A Deep Learning Approach for Urban Sound Classification

Sanjoy Barua
Department of Computer Science and Engineering
Port City International University
Chattogram, Bangladesh

Tahmina Akter
Department of Computer Science and Engineering
Port City International University
Chattogram, Bangladesh

Mahmud Abu Saleh Musa
Department of Computer Science and Engineering
Port City International University
Chattogram, Bangladesh

Muhammad Anwarul Azim
Department of Computer Science and Engineering
University of Chittagong
Chittagong-4331, Bangladesh

ABSTRACT
Urban sound classification is the task of identifying the type of sound present in a given recording, such as car honks, pedestrian footsteps, or construction noise. Accurate classification of urban sounds is important for a variety of applications, including environmental monitoring, traffic management, and public safety. To address this problem, we experiment with five different deep learning models: ANN, CNN, RNN, LSTM plus GRU combined model, and Bi-LSTM plus Bi-GRU model. These models are trained and evaluated on the Urban Sound 8K dataset, which consists of 8,000 urban sound recordings from 10 different classes. Our results show that the ANN model achieved the highest accuracy, reaching 95% on the test set. Overall, our results demonstrate the effectiveness of deep learning for urban sound classification and suggest that the ANN model is the most suitable for this task. This work has the potential to impact a variety of fields that rely on the accurate identification of urban sounds.

Keywords
Urban sound, environmental monitoring, deep learning, ANN, CNN, RNN, LSTM, LSTM plus GRUBi-LSTM plus Bi-GRU.

1. INTRODUCTION
Urban sound classification is a crucial task in modern cities, as it helps us understand and analyze the various types of sounds present in urban environments. This information can be used to optimize the acoustic environment of cities, improve public safety, and enhance the quality of life for residents[1].

Deep learning techniques have emerged as powerful tools for urban sound classification, as they can capture complex patterns in audio data and provide high accuracy in classification tasks.[2]

We propose a deep learning-based approach forurban sound classification using the Urban Sound 8k dataset and various models, including artificial neural networks (ANNs), convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and bidirectional LSTM (BILSTM) networks. Our approach involves the collection and preprocessing of urban sound data, the design and training of deep learning models, and the evaluation of the model's performance.

Our results showed that the ANN model achieved the highest accuracy of 95% on the urban sound classification task. This model outperformed the other models, including the CNN, LSTM, and BILSTM models, in terms of classification accuracy. We also conducted a detailed analysis of the model's performance andidentified the key factorsthatcontributed to its high accuracy. Later we tried LSTM plus GRU combined model and Bi-LSTM plus Bi-GRU model.

Overall, this work demonstrates the potential of deep learning techniques for urban sound classification and highlights the importance of carefully designing and evaluating deep learning models for this task. Our findings have implications for the optimization of urban acoustic environments and the development of intelligent systems for urban sound analysis.

2. LITERATURE REVIEW
Massoudi et. al., (2021) documented based on CNN. In this document Urban Sound dataset has used, which has 10 different classes. The audio samples firstly preprocessed in order to extract the MFCC features of each audio signal. Then the MFCC features vector is used as an input for the CNN mode for classification and to generate predictions. CNN model achieved 91% accuracy on the dataset.

Lezhenin et. Al., (2019) presented based on LSTM. The model is trained on magnitude mel-spectrograms extracted from UrbanSound8K dataset audio. The proposed network is evaluated using 5-Fold cross-validation and compared with the baseline CNN. LSTM network that takes magnitude mel-spectrograms provides 84.25% of average accuracy.

Demir et. al., (2020) documented is based on a deep Convolutional Neural Network (CNN). The proposed approach covers a bunch of stages such as pre-processing, deep learning-based feature extraction, feature concatenation, feature reduction, and classification. The proposed method produced 94.8%, 81.4%, and 78.14% accuracy scores for ESC-10, ESC-50, and UrbanSound8k datasets.

Krishan et. Al,. (2020) document showing that the long-short term memory (LSTM) shows a better result in classification accuracy compared to CNN for many features used. Some of the known datasets are used UrbanSound8k, ESC-10, ESC-50, and ESC-US. For the classification problem, authors have used spectral features where they convert the audio from the time domain to the frequency domain using Fourier transformation. There is a number of spectral features like MFCC, Spectral Centroid, Spectral Roll-off, Mel spectrogram, etc.

Tripathi et. al., (2021) document based on residual neural network. The proposed method is evaluated on two widely used Environmental Sound Classification datasets: ESC-10 and DCASE 2019 Task-1(A) datasets. The proposed model yields an accuracy of92% and 82% for theESC10 and DCASE 2019 Task-1(A) datasets.

Wenjie et. al. (2021) In this paper, a temporal-frequency attention-based convolutional neural network (TFCNN) is proposed in this paper to get better classification results. In the attention mechanism, the Log-Mel spectrogram extracted from the original audio data is generated. The backbone network part consists of 3 main layers of CNN. In this paper, ESC 50 & UrbanSound8k dataset is used and get the best accuracy of avg in ESC 50 and urbansound8k is 83.80% and 92.91%.

FATIH DEMIR et. al. (2020) In this paper, they trained the model in an end-to-end fashion with the spectrogram images using CNN. They also use the KNN ensembles classifier. First, they input an audio signal then make a spectrogram image and run it in the CNN model. they have used the DCASE-2017 ASC and the UrbanSound8K datasets, and show that the proposed CNN model achieves classification accuracies of 96.23% and 86.70%, respectively. They also try different CNN models to compare.

Chen et. al., (2019) The paper proposed dilated CNN model . the work mainly compares the result of using maxpooling & dilated CNN and finds out which gives the best accuracy. The dataset is not mentioned. They obtain an accuracy of 78% by using the proposed dilated CNN model.

Zilong et. al., (2020 they proposed a 2- Order Dense Convolutional network to overcome the limitations of a single-feature network. In this work, they use MFCC and GFCC for feature extraction. UrbanSound8K and Dcase2016 datasets are used. The experimental result shows that the accuracy of the network is respectively 84.83% and 85.17%, which has increased up to 13.81% and 7.07% compared with the baseline.

Guo et. al. (2022) presented a recognition method based on multi-feature parameters and time–a frequency attention module. This paper works on the betterment of classifying the audios which have multi features in it. They use a two-dimensional CNN model for this. And used ESC-10, ESC 50, and UrbanSound8K dataset for the work, The phase spectrogram is added to the input feature as the phase feature supplement of the Log-Mel spectrogram, which enhances the feature expression ability and improves the robustness and generalizability of the network.

3. METHODOLOGY

3.1 Dataset collection

Urban sound 8k is a well-known dataset for urban sound classification which is available on the internet.

3.2 Dataset Description

The Urban Sound 8K dataset is a collection of 8,000 urban sound recordings, covering 10 different classes of sounds. These classes include air conditioning, car horn, children playing, dog bark, drilling, engine idling, gunshot, jackhammer, siren, and street music.The dataset was collected from field recordings made in avariety of urban environments, and each recording is one second long.

3.3 Preprocessing Steps

3.3.1 Librosa

Librosa is a Python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems [13]. In the context of urban sound classification, it can be used to extract features from audio recordings of city sounds, such as traffic noise, and then use those features to train a machine learning model for classifying new audio recordings.

3.3.2 MFCC

An MFCC is made up of a number of coefficients known as mel-frequency cepstral coefficients (MFCCs). They were created using an audio clip's cepstral representation (a nonlinear "spectrum-of-a-spectrum")[14]. The mel-frequency cepstrum (MFC) differs from the cepstrum in that the frequency bands are evenly spaced on the mel scale, which more closely resembles the response of the human auditory system than the linearly-spaced frequency bands used in the conventional spectrum. When used in audio compression, for instance, this frequency warping can improve the representation of sound and potentially lower the transmission bandwidth and storage needs of audio signals [15].

Feature extraction is a special form of reduction of the dataset. Using feature extraction techniques used for extracting specific features from the speech, these features carry the characteristics of the specific speech which are useful for Differentiating the different speech, so these features will play a major role in speech recognition [16].

Compressing a voice signal into streams of acoustic feature vectors, also known as speech feature vectors, is the first step in speech recognition. The idea of feature extraction is actually divided into two steps: first, the speech signal is transformed into feature vectors; second, the useful characteristics that are impervious to changes in the surroundings and speech variation are selected. In speech recognition systems, however, where accuracy has drastically declined in the case of their existence, changes in ambient variables and variances in speech are significant. The Mel Frequency Cepstral Coefficients (MFCC) features, which are the most popular and reliable due to their precise estimation of the speech parameters and effective computational model of speech, are unquestionably the most often utilized speech features.[17]

3.4 Splitting into train and test data

After the Feature extraction part the next step is to split the dataset into train and test data. Train data is for training the model and the test data is to check the performance of the model. In this research, 80% of the data are taken fortraining purposes.

Table 3.1: The train and test data size

Dataset	TrainData	TestData
UrbanSound8k	6985	1747

3.5 Model Building

The purpose of this section is to introduce the Deep Learning models that were applied in this research.

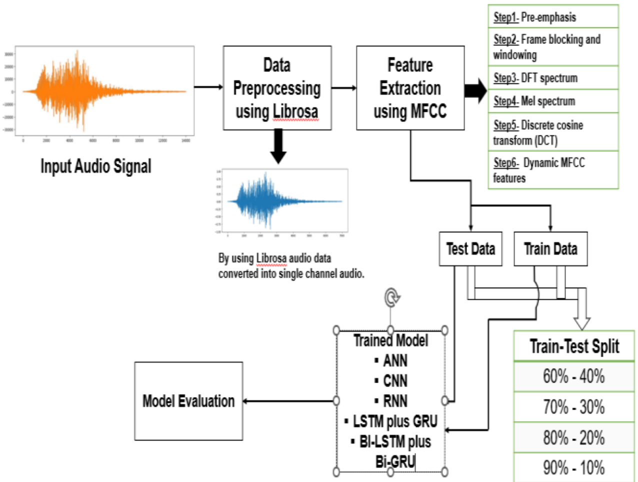

Figure 3.1: The overview of the proposed method

3.6 Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) are algorithms that simulate brain activity and are used to model complex patterns and make predictions about future events.
The Artificial Neural Network (ANN) is a deep learning technique that was inspired by the idea of biological neural networks seen in the human brain.[18]

An effort to mimic the functions of the human brain led to the creation of ANN.

Although they are not exactly the same, ANN and biological neural networks have very similar workings. The ANN algorithm only takes structured and numeric data.

3.6.1 Dense Layer

A dense layer is a layer that is deeply connected with its preceding layer which means the neurons of the layer are connected to every neuron of its preceding layer.[19]

The first dense layer which is receiving the input shape consists of 1000 nodes. Other dense layer consists of 750 nodes, 500 nodes, 250 nodes, and 100 nodes respectively. In this five-dense layer 'ReLU' has been used as an activation function.

3.6.2 Dropout Layer

Dropout is a regularization technique used to exacerbate the overfitting problem in neural networks. A significant portion of the neurons in a certain layer will be restricted during the preparation step. The layer is then forced to familiarize itself with the same concept using various neurons. This layer is turned off during the anticipation phase. The dropout rate for all of the ANN engineering layers was 30%.

3.6.3 Output Layer:

The output layer having activation function 'softmax' will consist of 10 nodes in our model that refer to the possible classification numbers. The model then predicts the choice with the highest probability.
Figure 2 demonstrates the architecture of the Artificial Neural Network (ANN) model for our proposed system which represents the concrete view of the ANN model. We look at the architecture of CNN at a glance,which can be divided into the following steps:

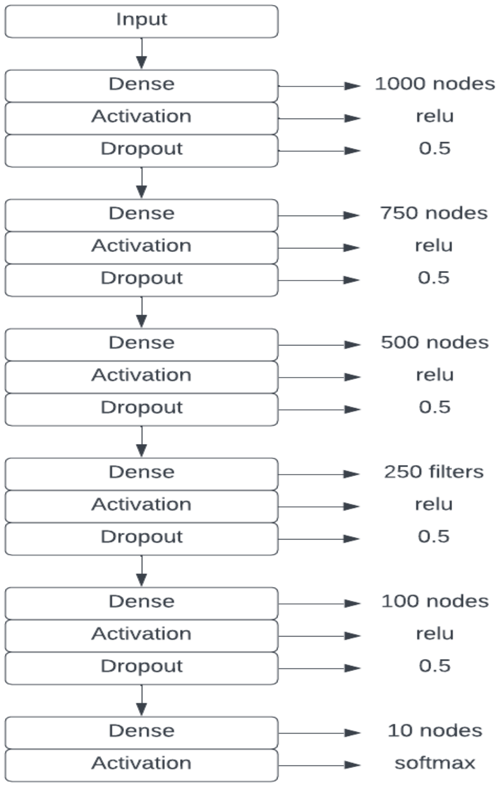

Figure 3.2: Block diagram of ANN model for our system

3.7 Convolutional Neural Network (CNN)

Convolutional neural network (CNN) is a subclass of deep neural networks in the field of deep learning. In a typical CNN, there is a series of various kind of layers that are combined in the overall architecture.[20]

The training of a CNN requires different kinds of decisions that need to be taken in terms of both architectural patterns such as the number of convolution and pooling layers, input data format, filter dimension etc, as well as hyperparameters such as learning rate, dropout probability, number of epochs, batch size, etc.[21]

Our CNN model is consisting of three Conv1D layers followed by one dense layer and ReLU is used as an activation function. The first Conv1D layer which is receiving the input shape consists of 500 filters, a kernel size of 5, and every Conv1D layer followed by the MaxPooling1D layer. The second Conv1D layer consists of 300 filters, a kernel of size 5, and the third Conv1D layer consists of 150 filters, a kernel of size 5. The dense layer used 150 nodes and dropout is 50% to reduce overfitting. The output layer having the activation function 'sigmoid' will consist of 10 nodes in our model that refer to the possible classification numbers.

3.8 Recurrent Neural Network (RNN)

A feed-forward neural network with internal memory is referred to as a recurrent neural network. RNN is recurrent in nature as it performs the same function for every input of data while the output of the current input depends on the past computation. RNN is best suited for sequential data. It can handle arbitrary input/output lengths. RNN uses its internal memory to process arbitrary sequences of inputs. RNN has memory capabilities. It memorizes previous data. It considers the current input as well as the lessons it has learned from the inputs it has previously received before reaching a decision. A feedback loop is created when the output from the previous step is used as the input for the next step. As a result, it uses the set of current input and the past state to determine its current state. The information loops around in this manner.

In the RNN model, we used two SimpleRNN layers and a dense layer as an output layer. The SimpleRNN layer consists of 300 nodes and 200 nodes respectively.

3.9 LSTM plus GRU Combined Model

A combined LSTM and GRU model is a deep learning model that utilizes both Long Short-term Memory (LSTM) and Gated Recurrent Unit (GRU) layers in its architecture. LSTM and GRU are both types of recurrent neural networks that are able to process sequential data such as time-series data or natural language [22].

LSTM and GRU have similar functionality, but they have different architectures and work slightly differently. LSTM networks have memory cells, input gates, output gates, and forget gates that allow them to selectively retain or discard information over time. On the other hand, GRU networks have update and reset gates that control the flow of information [23]. In a combined LSTM and GRU model, the LSTM layers are typically used to capture long-term dependencies in the data, while the GRU layers are used to capture short-term dependencies. The combination of these two types of layers allows the model to learn both long-term and short-term patterns in the data, making it more robust and accurate.

The architecture of a combined LSTM and GRU model can vary depending on the specific task and dataset. Typically, the LSTM and GRU layers are stacked on top of each other and are followed by one or more dense layers for classification or regression. Dropout and other regularization techniques can also be added to prevent overfitting [24].

In the LSTM plus GRU model, we used one LSTM layer, one GRU layer, and one dense layer as an output layer. The first LSTM layer which is receiving the input shape consists of 128 nodes and is followed by a GRU layer which consists of 64 nodes. Reducing the overfitting dropout layer is used after LSTM and the dropout is 50%. After using the dropout layer a flatten layer has used. The output layer having the activation function 'sigmoid' will consist of 10 nodes in our model that refer to the possible classification numbers.

3.10 Bi-LSTMplusBi-GRU Combined Model

A bi-directional LSTM (Bi-LSTM) and bi-directional GRU (Bi-GRU) combined model is a deep learning model that utilizes both bi-directional Long Short-term Memory (Bi-LSTM) and bi-directional Gated Recurrent Unit (Bi-GRU) layers in its architecture.

A bi-directional RNN (BRNN) is a type of recurrent neural network (RNN) that processes the input sequence in both forward and backward directions. This allows the BRNN to capture information from the past and future context of a given input element. A bi-directional LSTM or a Bi-GRU, therefore, is a BRNN that uses LSTM or GRU units respectively [25].

In a bi-directional LSTM and bi-directional GRU combined model, the bi-directional LSTM layers are typically used to capture long-term dependencies in the data, while the bi-directional GRU layers are used to capture short-term dependencies. The combination of these two types of layers allows the model to learn both long-term and short-term patterns in the data, making it more robust and accurate.

The architecture of a bi-directional LSTM and bi-directional GRU combined model can vary depending on the specific task and dataset.Typically, the Bi-LSTM and Bi-GRU layers are stacked on top of each other and are followed by one or more dense layers for classification or regression. Dropout and other regularization techniques can also be added to prevent overfitting [26].

In a Bi-LSTM and Bi-GRU combined model, the input sequence is processed by both the layers in both forward and backward directions, leading to the model learning both past and future context for each element of the input sequence.It can be useful in tasks such as named entity recognition, sentiment analysis, and machine translation.

In Bi-LSTM and Bi-GRU combined model, we have used one Bidirectional LSTM layer, one Bidirectional GRU layer, and one dense layer as an output layer. The first Bidirectional LSTM layer which is receiving the input shape consists of 128 nodes and is followed by the Bidirectional GRU layer which consists of 64 nodes. Dropout is 50% to reduce overfitting. After using the dropout layer a flatten layer has used. The output layer having the activation function 'softmax' will consist of 10 nodes in our model that refers to the possible classification numbers.

4. RESULT AND ANALYSIS

Parameter tuning based on MFCC value in the proposed ANN model:

Table 4.1: Results of different MFCC value

MFCC Value	Accuracy	Precision	Recall	F1 score
10	89.98	90	89	90
20	94.10	94	93	94
30	94.16	94	93	94
40	94.56	94	95	94
50	94.47	96	95	95
60	94.73	95	94	95
70	94.73	95	95	95
80	95.13	96	94	95
90	94.96	95	94	95
100	95.30	95	95	95
110	95.24	96	95	95
120	95.47	96	95	95

In this section, we have shown the result of ANN, CNN, RNN, LSTM plus GRU and Bi-LSTM plus Bi-GRU model performance on UrbanSound8k. Here are the accuracy, Precision, Recall & F1 scores of all models using MFCC value 120, and all models used 1000 epochs and 400 batch size.

Table 4.2: Results of different models

Models	Results			
	Accuracy	Precision	Recall	F1 score
ANN	**95**	**95**	**95**	**95**
CNN	93	93	93	93
RNN	88	87	87	88
LSTM plus GRU	92	92	91	92
Bi-LSTM plus Bi-GRU	93	94	93	93

By this result table, we can say that based on our work ANN model perform the best in urban sound classification.

4.1 Accuracy based on train test split value

Table 4.3: Results based on train test split of different models.

Models	Results				
	ANN	CNN	RNN	LSTM plus GRU	Bi-LSTM plus Bi-GRU
60% - 40%	94.27	92.02	89.92	84.89	84.74
70% - 30%	94.65	92.13	90.57	83.96	85.83
80% - 20%	**95.47**	**93.70**	**91.92**	84.88	88.03
90% - 10%	94.50	92.79	90.16	**87.46**	**88.32**

4.2 Training and validation losses of the ANN model

Figures 3 and 4 show that training and validation losses of the ANN model have decreased over time as epochs progressed

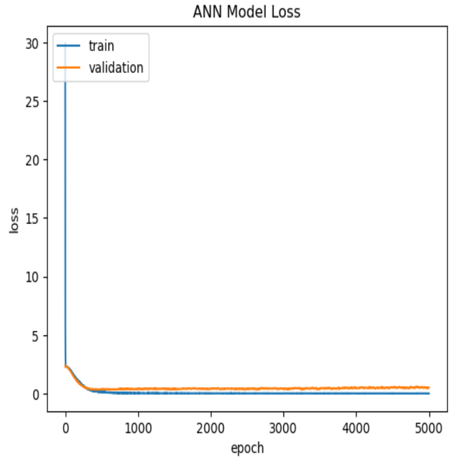

Figure 4.1: Train accuracy and validation loss

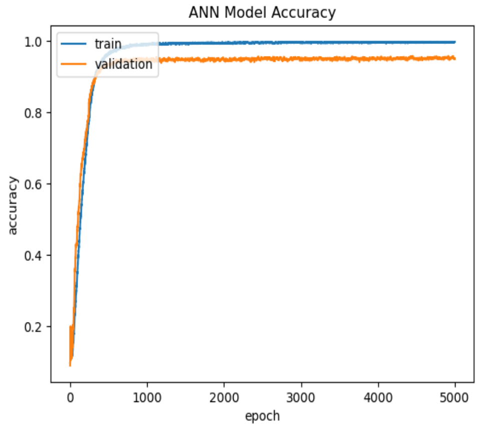

Figure 4.2: Train accuracy and validation accuracy

4.3 Classification report on the ANN model

In machine learning, a classification report is a performance evaluation metric. It is used to display the trained classification model's precision, recall, F1 Score, and support.

Table 4.4: Classification report on the ANN model

Class Name	Precision	Recall	F1-Score
0	0.98	0.99	0.99
1	0.97	0.93	0.95
2	0.93	0.92	0.93
3	0.95	0.94	0.94
4	0.96	0.94	0.95
5	0.98	0.98	0.98
6	0.97	0.90	0.93
7	0.95	0.96	0.96
8	0.98	0.98	0.98
9	0.89	0.9	0.92

Accuracy	0.95
Macro avg	0.95
Weighted avg	0.95

4.4 Confusion Matrix for ANN Model

A confusion matrix is a table that is used to define the performance of a classification algorithm. A confusion matrix visualizes and summarizes the performance of a classification algorithm.

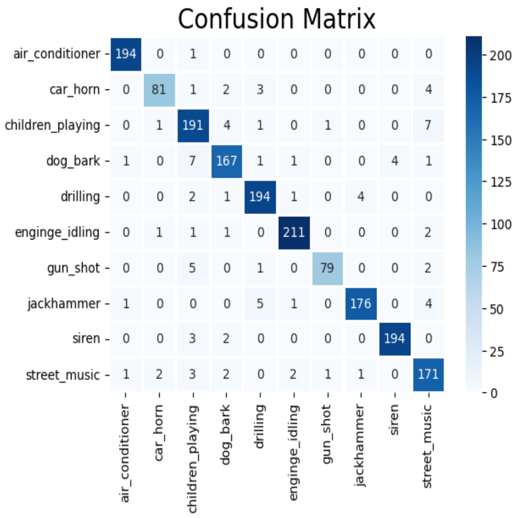

Figure 4.3: Confusion Matrix for ANN Model

5. CONCLUSION

In conclusion, this study aimed to classify urban sounds using deep learning algorithms and the Urban Sound 8k dataset. The MFCC feature was extracted from the audio samples and used as input for the models. A total of five algorithms were tested, including ANN, CNN, RNN, LSTM, and B-iLSTM. The results showed that ANN had the highest accuracy of 95%, followed by CNN with 93.70% accuracy. RNN and LSTM had lower accuracies of 88%, and 87% respectively. We also tried LSTM plus GRU combined Model and Bi-LSTM plus Bi-GRU Combined model. These findings suggest that ANN is the most effective algorithm for urban sound classification, followed by CNN when using the MFCC feature. However, further research is needed to determine the optimal feature representation and the most suitable algorithm for this task under different circumstances.

6. REFERENCES

[1] Chachada, S. and Kuo, C.C.J., 2014. Environmental sound recognition: A survey. APSIPA Transactions on Signal and Information Processing, 3, p.e14.

[2] Sharma, Jivitesh, Ole-ChristofferGranmo, and Morten Goodwin. "Environment Sound Classification Using Multiple Feature Channels and Attention Based Deep Convolutional Neural Network." Interspeech. Vol. 2020. 2020.

[3] Massoud Massoudi, Siddhant Verma &Riddhima Jain "Urban Sound Classification using CNN" Published in 2021, 6th International Conference on Inventive Computation Technologies (ICICT)

[4] IuriiLezhenin, Natalia Bogach&EvgenyPyshkin "Urban Sound Classification using Long Short-Term Memory Neural Network"Published in 2019, Federated Conference on Computer Science and Information Systems (FedCSIS)

[5] Fatih Demir, Muammer Turkoglu, Muzaffer Aslan & Abdulkadir Sengur "A new pyramidal concatenated CNN approach for environmental sound classification " Published in 2020 by ELSEVIER .Applied Acoustics Vol. 170

[6] Joy Krishan Das, Arka Ghosh, Abhijit Kumar Pal, Sumit Dutta & Amitabha Chakrabarty "Urban Sound Classification Using Convolutional Neural Network and Long Short Term Memory Based on Multiple Features" Published in 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)

[7] Achyut Mani Tripathi &Aakansha Mishra "Environment sound classification using an attention-based residual neural network "Published In 2021 by ELSEVIER . Neurocomputing Vol. 460

[8] Wenjie Mu, BoYin, Xianqing Huang, Jiali Xu &Zehua Du "Environmental sound classifcation using temporal-frequency attention based convolutional neural network" Published in 2021 by Scientific Reports

[9] Fatih Demir, Daban Abdulsalam Abdullah & Abdulkadir Sengur "A New Deep CNN model for Environmental Sound Classification" Published in 2020 by IEEE Access Vol. 8

[10] Chen, Y., Guo, Q., Liang, X., Wang, J. and Qian, Y., 2019. Environmental sound classification with dilated convolutions. Applied Acoustics, 148, pp.123-132.

[11] Huang, Z., Liu, C., Fei, H., Li, W., Yu, J. and Cao, Y., 2020. Urban sound classification based on 2-order dense convolutional network using dual features. Applied Acoustics, 164, p.107243.

[12] Guo, J., Li, C., Sun, Z., Li, J. and Wang, P., 2022. A Deep Attention Model for Environmental Sound Classification from Multi-Feature Data. Applied Sciences, 12(12), p.5988.

[13] Raguraman, Preeth, R. Mohan, and Midhula Vijayan. "Librosa based assessment tool for music information retrieval systems." 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE, 2019.

[14] Majeed, S.A., Husain, H., Samad, S.A. and Idbeaa, T.F., 2015. MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC) FEATURE EXTRACTION ENHANCEMENT IN THE APPLICATION OF SPEECH RECOGNITION: A COMPARISON STUDY. Journal of Theoretical & Applied Information Technology, 79(1).

[15] Hossan, Md Afzal, Sheeraz Memon, and Mark A. Gregory. "A novel approach for MFCC feature extraction." 2010 4th International Conference on Signal Processing and Communication Systems. IEEE, 2010.

[16] Gaikwad, S., Gawali, B., Yannawar, P. and Mehrotra, S., 2011, December. Feature extraction using fusion MFCC for continuous marathi speech recognition. In 2011 Annual IEEE India Conference (pp. 1-5). IEEE.

[17] Murty, K. Sri Rama, and BayyaYegnanarayana. "Combining evidence from residual phase and MFCC features for speaker recognition." *IEEE signal processing letters* 13.1 (2005): 52-55.

[18] Abiodun, Oludare Isaac, et al. "Comprehensive review of artificial neural network applications to pattern recognition." IEEE Access 7 (2019): 158820-158846.

[19] Bashar, Syed K., Abdullah Al Fahim, and Ki H. Chon. "Smartphone based human activity recognition with feature selection and dense neural network." 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020.

[20] Chauhan, Rahul, Kamal Kumar Ghanshala, and R. C. Joshi. "Convolutional neural network (CNN) for image detection and recognition." 2018 first international conference on secure cyber computing and communication (ICSCCC). IEEE, 2018.

[21] Kattenborn, Teja, Jens Leitloff, Felix Schiefer, and Stefan Hinz. "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing." ISPRS journal of photogrammetry and remote sensing 173 (2021): 24-49.

[22] Zeng, Chao, et al. "Parking occupancy prediction method based on multi factors and stacked GRU-LSTM." IEEE Access 10 (2022): 47361-47370.

[23] Shahid, F., Zameer, A. and Muneeb, M., 2020. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos, Solitons & Fractals, 140, p.110212.

[24] Islam, M.S. and Hossain, E., 2021. Foreign exchange currency rate prediction using a GRU-LSTM hybrid network. Soft Computing Letters, 3, p.100009.

[25] Bohan, H. and Yun, B., 2019, July. Traffic flow prediction based on BRNN. In 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC) (pp. 320-323). IEEE.

[26] Mouthami, K., Anandamurugan, S. and Ayyasamy, S., 2022, December. BERT-BiLSTM-BiGRU-CRF: Ensemble Multi Models Learning for Product Review Sentiment Analysis. In 2022 6th International Conference on Electronics, Communication and Aerospace Technology (pp. 1514-1519). IEEE.