iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4137

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Sound Recognition Using Recurrent Neural Network

Mrs Deepa R[1], Atharva M Kavitkar[2], V. Soumya[3]
[1]*CSE Department, Assistant Professor, SRM Institute of Science and Technology, Chennai, India*
[2]*CSE Department, Student, SRM Institute of Science and Technology, Chennai, India*
[3]*CSE Department, Student, SRM Institute of Science and Technology, Chennai, India*

Abstract: ***For intelligent systems to be more efficient and effective, it is not enough that they can recognize just speech and music, but also general sounds in everyday environment. In this paper, Recurrent Neural Network will be implemented to classify different sounds present in the environment .Sound classification is an important feature that allows a robot to recognize its auditory environment without an intervention from a human operator. It helps the robot to explore, identify and map its surroundings. By recognizing sounds, the pertaining algorithms and behavior of the robot can be studied and improved upon. The various physical properties of a sound wave are used for recognizing each sound uniquely.***
Index Terms: ***Sound Recognition, Recurrent Neural Network, MFCC, Spectogram***

I. INTRODUCTION

E ver since the birth of Artificial Intelligence(AI) in the early 19th century there has been a tremendous growth in its application area. Converting the complex process of human thinking into a much simpler mechanical manipulation of symbols has become the need of the hour. Increase in knowledge and experience has led to the opening of much wider scope for AI and its applications.
Automatic Sound Recognition (ASR) and Music Information Retrieval(MIR) are such applications where AI can be put to test. Sound and Music are just two out of many types of sound available in our environment and in order to function effectively like a human being it is important for the system to be able to recognize other sounds as well. These Sounds can be incorporated along with videos or simply as an individual entity. The information obtained from this audio analysis can further be used for predicting events or actions. . The various physical properties of a sound wave are used for recognizing each sound uniquely.
Artificial Neural Network(ANN) called Recurrent Neural Network(RNN) can help in detecting unique properties and classifying sounds from various entities A Recurrent Neural Network is a class of Artificial Neural Network where associations between units shape a coordinated diagram along a grouping. This enables it to display dynamic transient conduct for a period succession. Dissimilar to feed forward Recurrent Neural Networks, RNNs can utilize their inner state (memory) to process successions of sources of info. It acts like a memory stockpiling unit which stores for a brief time of time. This makes them relevant to assignments.
In this paper, the sound from various sources found in day to day life such as dog bark, siren, gunshot ,street music etc will be fed into the system where the system first creates an auditory image of the sound which helps us of differentiate sounds in terms of their properties and then feeds the vectored input into the RNN which further classifies the sound into their respective classes. We have used 10 classes of sounds namely dog bark ,children playing, street music ,siren ,gunshot, drilling, air conditioner, car horn, jack hammer and engine idling.

II. BACKGROUND

The programmed acknowledgment of sound by PCs is an essential part of rising applications, for example, mechanized reconnaissance, machine hearing and sound-related event understanding. Late advances in machine learning, and additionally in computational models of the human sound-related framework, have added to progresses in this undeniably well known research field. Question sound grouping, the capacity to perceive sounds under certifiable loud conditions, is a particularly difficult errand. Order strategies deciphered from the discourse acknowledgment space, utilizing highlights, for example, Mel-recurrence cepstral

coefficients, have been appeared to perform sensibly well for the sound grouping errand; in spite of the fact that spectrogram-based or sound-related picture investigation systems apparently accomplish prevalent execution in clamor. Execution is assessed on a standard characterization undertaking in various levels of defiling clamor, and with a few framework upgrades, and appeared to contrast exceptionally well and current cutting edge arrangement systems.

III. PROPOSED SYSTEM

A. Objective

Focus of the proposed system is to provide optimized and robust Recurrent Neural Network. The suggested Recurrent Neural Network will be capable of classifying datasets with high amount of correlation .The system would be able to operate in real life conditions with noise, unlike the existing system.

B. Scope

The proposed system will be built upon the existing system, generalizing the concept to be implemented on a larger dataset. The proposed system has vast applications due to its Universal Nature. Our project acts as a small contribution to the ongoing research of Machine Listening and Auditory Analysis which will bring humans closer to the ultimate goal. We see it becoming an integral part of nature inspired robots, they will be more intelligent with respect to their surroundings and will actually be able to 'feel' their environment. Hence this project has tremendous scope and possibility of advancement in the near future.

C. Feature Extraction

Feature extraction starts from a basic course of action of assessed data and makes deduced values (features) anticipated that would be helpful and non-dull, empowering the subsequent learning and hypothesis steps, and now and again inciting better human interpretations.The Waveplot[1], Spectogram[2] and log power Spectogram[3] provide a more intuitive representation of the audio signal. Sound consists of following features

1) *Mel-frequency Cepstrum (MFC)*-It is a portrayal of the fleeting force range of a sound, in light of a straight cosine change of a log power range on a nonlinear mel size of frequency. Mel-recurrence cepstral coefficients (MFCCs) are coefficients that all things considered make up MFC. They are gotten from a sort of ceptral portrayal of the sound clasp (a nonlinear "range of-a-range"). The contrast between the cepstrum and the mel-recurrence cepstrum is that in the MFC, the recurrence groups are similarly divided on the mel scale, which approximates the human sound-related framework's reaction more intently than the straightly separated recurrence groups utilized as a part of the ordinary cepstrum. This recurrence distorting can take into consideration better portrayal of sound, for instance, in sound compression. MFCCs are generally inferred as takes after:-

2) The Fourier transform of (a windowed excerpt of) a signal is taken.

3) The spectrum values obtained are mapped ont the mel scale, with the help of triangular overlapping windows.

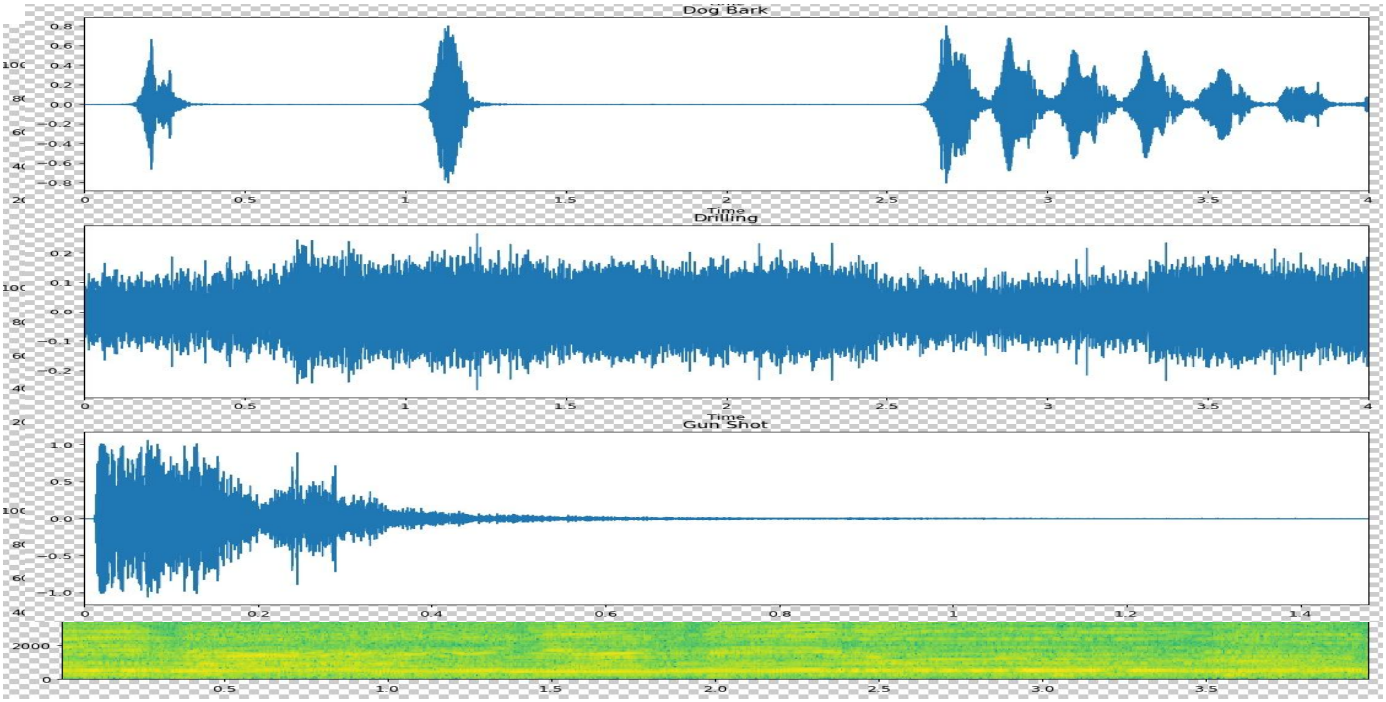
4) The mel frequency's log values are taken into account.

5) The discrete cosine transform is considered as a signal and its mel log powers.

6) The MFCCs are the peak displacement from normalized values of the resulting spectrum.

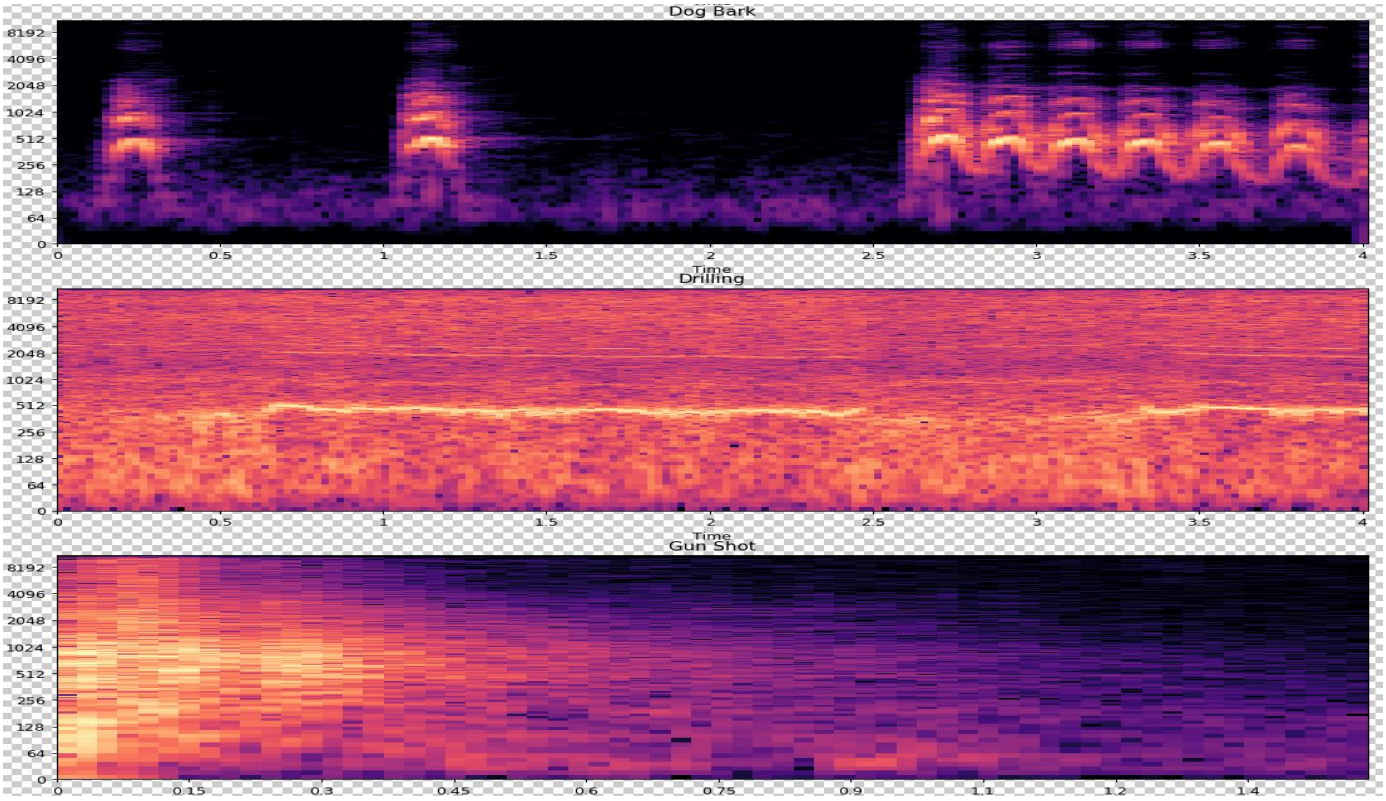
7) Chroma feature or Chromagram nearly identifies with the twelve diverse pitch classes. Chroma-based highlights, which are likewise alluded to pitch class profiles are a capable apparatus for examining music whose pitches can be definitively sorted (regularly into twelve classifications) and whose tuning approximates to the equivalent tempered scale. One fundamental property of chroma highlights is that they catch symphonious and melodic qualities of music, while being vigorous to changes in timbre and instrumentation.

Figure 1 Waveplot



It is a applied latticediagram speaking to tonal space initially depicted by Leonhard Euler (1739). Different visual portrayals of the Tonnetz can be utilized to indicate customary symphonious connections in European traditional music.

Figure 2 Spectogram



D. Recurrent Neural Network

The essential component of RNN is that the system contains no less than one information affiliation, so the incitations can stream around, which enables the frameworks to do transient taking care of and learn progressions, e.g., perform game plan affirmation/age or passing association/desire .Recurrent Neural Network structures can have an extensive variety of structures. One essential difference containing out of a standard feed forward network is not withstanding previous iterations. These can abuse the powerful non-coordinate mapping capacities of the network, and besides have some sort of memory. Other models have more stable organisation, with every neuron related with all the others, and may in like manner have stochastic start limits. For straightforward designs and deterministic actuation capacities, learning can be accomplished utilizing comparable angle plunge methodology to those prompting the back-engendering calculation for encourage forward systems. At the point when the enactments are stochastic, reenacted strengthening methodologies might be more fitting. The Recurrent Neural Network sources are info and yields are the vectors. When all is said in done, the condition of a non-static framework is an arrangement of qualities that abridges all the data about the past conduct of the framework that is important to give a special portrayal of its future conduct, aside from the impact of any outer variables. In this case the state is characterized by the arrangement of concealed unit initiations. Thus, along with the internal and external spaces, it also includes state space. The order of the system is the dimensionality of the state space and the number of hidden units. The extracted features are fed into the Recurrent Neural Network for classification. Each neuron of the deep Recurrent Neural Network updates its weights with successive iterations to mold the Recurrent Neural Network according to the training data. In the event that the Recurrent Neural Network data sources and yields are the vectors x(t) and y(t), the association weight grids are WIH, WHH and WHO, and the covered up and yield unit enactment capacities are fH and fO, the conduct of the intermittent system can be depicted by the match of non-straight lattice conditions:

$$h(t) = f_H\,(W_{IH}\,x(t)+W_{HH}h(t-1))$$

$$y(t) = f_O(W_{HO}h(t))$$

As a rule, the state [h(t)] of a dynamical framework is an arrangement of qualities that compresses all the data about the past conduct of the framework that is important to give a one of a kind depiction of its future conduct, aside from the impact of any outer components.

The Cost function is the mathematical function of the error generated by the Recurrent Neural Network. The aim is to minimize the cost function for achieving maximum accuracy. To prepare our system we require an approach to gauge the blunders it makes. We call this the loss function L, and our objective is discover the parameters U,V and W that limit the loss function for our preparation information. A typical decision for the loss function is the cross-entropy loss. On the off chance that we have N preparing cases (words in our content) and C classes (the measure of our vocabulary) at that point the loss as for our forecasts o and the genuine marks y is given by:

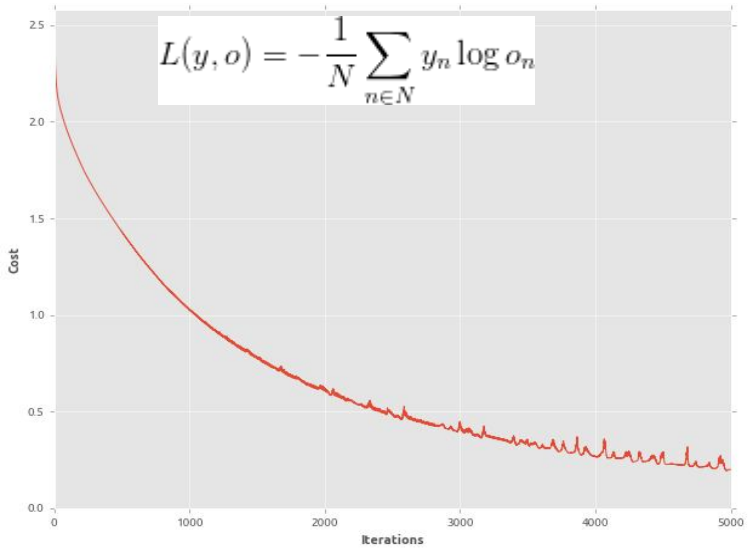


Figure 3 log power Spectogram

IV. CHALLENGES

The tunable parameters related to the DNN classifier were not fully optimized to produce the best results. System trained only in high level of noise, hence sacrifices performance in ideal conditions. Works well with data having very less correlation.

Segmentation of the foreground and background signals while feature extraction was a particularly tricky portion to implement. As the number of neurons in the Recurrent Neural Network increases the possibility of the network discarding the semantic structure and meaning of the training data also increases. Making the Recurrent Neural Network follow the semantic and sequential structure of the data was one of the challenges faced. The dataset used for training and testing had more than 8000 audio files. Unfortunately some of them were corrupt due to which their features could not be extracted. This reduced the size of effective training dataset and eventually decreasing in the training accuracy of the Recurrent Neural Network.

V. RESULT AND CONCLUSION

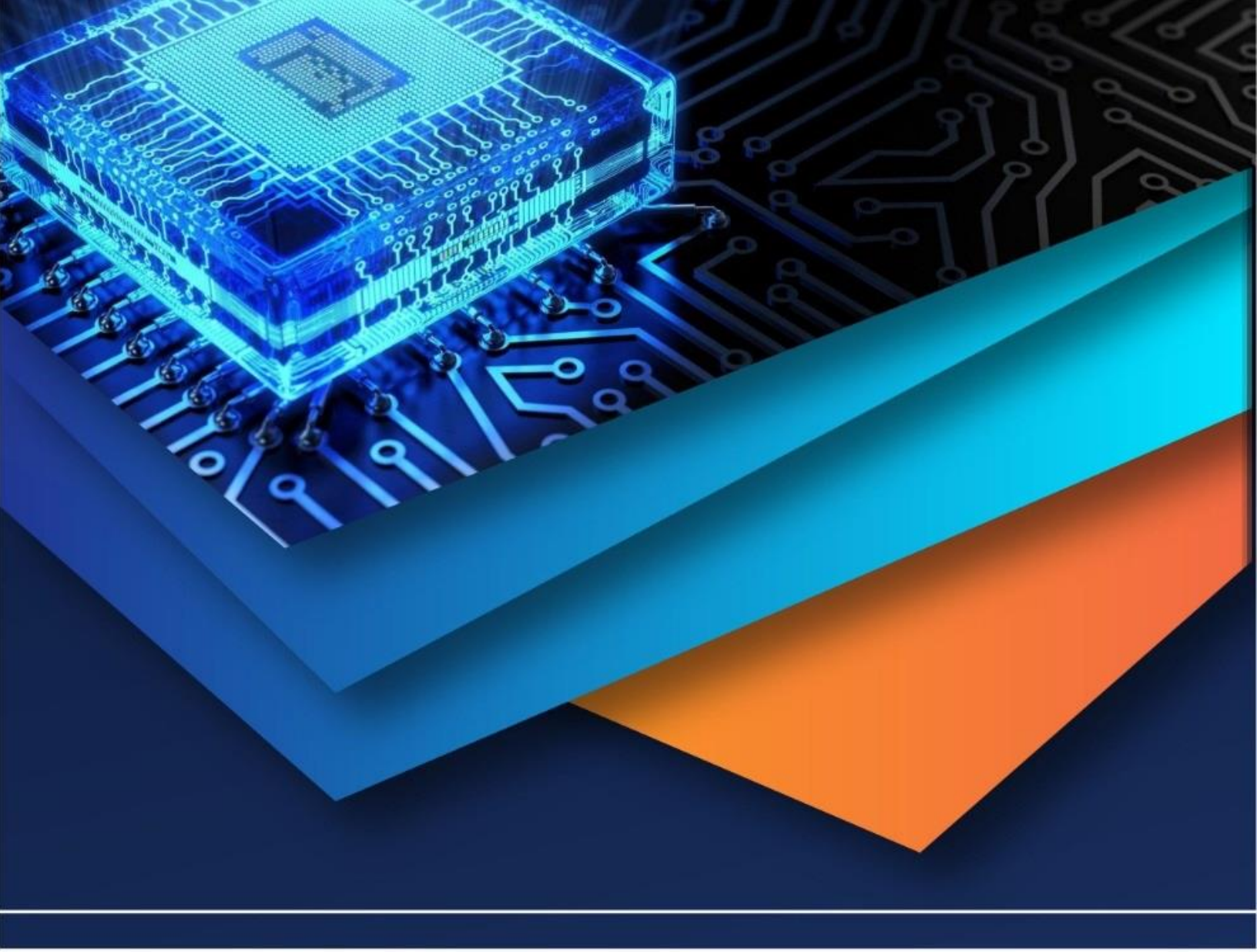
The goal of this paper was to evaluate whether recurrent Recurrent Neural Networks can be successfully applied to environmental sound classification tasks, especially considering the limited nature of datasets available in this field. It seems that they are indeed a viable solution to this problem. The results show that a recurrent model(accuracy 40%) outperforms common approaches based on feed forward multi-layer perceptron model(accuracy 20%) and achieves a similar level as other feature learning methods. Although, taking into consideration much longer training times, the result cannot be considered groundbreaking but it shows that Recurrent Neural Networks can be effectively applied in environmental sound classification tasks even with limited datasets and simple data augmentation. What is more, it is quite likely that a considerable increase in the size of the available dataset would vastly improve the performance of trained models, as the gap in computational performance is still profound causing considerable amount of time for training. One of the possible questions open for future inquiry is whether Recurrent Neural Networks could be effectively used with other less complex models, as they seem to concentrate on sequential aspects of sound events.

VI. ACKNOWLEDGEMENT

We would like to thank Justin Salamon and Urban Sound team for providing details on the results obtained on the UrbanSound8K dataset for the purpose of academic research.

REFERENCES

[1] R. F. Lyon, “Machine hearing: An emerging field,” IEEE Signal Process. Mag., vol. 42, no. 5, pp. 1414–1416, Sep. 2010.
[2] “Detection and Classification of Acoustic Scenes and Events” by Dan Stowell, Dimitrios Giannoulis, Emmanouil Benetos.
[3] D. Giannoulis, E. Benetos, D. Stowell, M. Rossignol, M. Lagrange and M. D. Plumbley, “Detection and classification of acoustic scenes and events: An IEEE AASP challenge,” in Proc. IEEE Workshop Appl. Signal Process. Audio Acoust., Oct. 2013
[4] S. Chu, S. Narayanan, and C.-C. Kuo, “Environmental sound recognition with time-frequency audio features,” IEEE Trans. on Audio, Speech, and Language Processing, vol. 17, no. 6, pp. 1142–1158, Aug. 2009.
[5] V. Bisot, R. Serizel, S. Essid, and G. Richard, “Acoustic scene classification with matrix factorization for unsupervised feature learning,” in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, Mar. 2016, pp. 6445–6449.
[6]] E. Cakir, T. Heittola, H. Huttunen, and T. Virtanen, “Polyphonic sound event detection using multi label deep neural networks,” in 2015 International Joint Conference on Neural Networks (IJCNN), July 2015.
[7] K. J. Piczak, “Environmental sound classification with convolutional neural networks,” in 25th International Workshop on Machine Learning for Signal Processing (MLSP), Boston, MA, USA, Sep. 2015.
[8] D. Giannoulis, E. Benetos, D. Stowell, M. Rossignol, M. Lagrange, and M. D. Plumbley, “Detection and classification of acoustic scenes and events: An IEEE AASP challenge,” in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz.
[9] S. Sigtia, A. Stark, S. Krstulovic, and M. lumbley, “Automatic environmental sound recognition: Performance versus computational cost,” EEE/ACM Transactions on Audio, Speech, and Language Processing, vol. PP, no. 99, pp. 1–1, 2016.
[10] C. V. Cotton and D. P. W. Ellis, “Spectral vs. spectro-temporal features for acoustic event detection,” in IEEE Worksh. on Apps. of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, Oct. 2011, pp. 69–72.
[11] J. Salamon, C. Jacoby, and J. P. Bello, “A dataset and taxonomy for urban sound research,” in 22nd ACM International Conference on Multimedia (ACM-MM’14), Orlando, FL, USA, Nov. 2014.
[12] K. J. Piczak, “ESC: Dataset for environmental sound classification,” in 23rd ACM International Conference on Multimedia, Brisbane, Australia, Oct. 2015, pp. 1015–1018.
[13] G. Parascandolo, H. Huttunen, and T. Virtanen, “Recurrent neural networks for polyphonic sound event detection in real life recordings,” in International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, Mar. 2016, pp. 6440–6444.
[14] L. Bottou, “Large-scale machine learning with stochastic gradient descent,” in 19th International Conference on Computational Statistics (COMPSTAT), Paris, France, Aug. 2010, pp. 177–186.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY