International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 11, No.1 (Mar-2022)

https://dx.doi.org/10.12785/ijcds/110196

Sounds Recognition in the Battlefield Using Convolutional Neural Network

Khalid M.O. Nahar[1], Fedaa Al-Omari[2], Nouh Alhindawi[3] and Mustafa Banikhalaf[1]

[1]*Department of Computer Sciences, Yarmouk University, Irbid, Jordan*
[2]*Directorate of Education of the District of Qasbah Irbid, Irbid Governorate, Ministry of Education, Jordan*
[3]*Department of Computer Science and Software Engineering, Faculty of Sciences and Information Technology, Jadara University, Irbid, Jordan*

Received 13 Sep. 2021, Revised 21 Jan. 2022, Accepted 20 Feb. 2022, Published 31 Mar. 2022

Abstract: Predicting enemy movements on the battlefield, especially when military raids occur is one of the important factors in battle winning. The enemies may be far away or hidden, but sounds are heard. Based on sounds that are outcomes from hidden enemies and by identifying the type of sound, a lot of information could be gained in further physical processing. The approximate location, distance, and the sound direction could be predicted. Moreover, establishing a sensitive model that relies on distinguishing military sounds will assist soldiers in alerting their military troops or camps for a near or faraway danger. Therefore, in this research, we build a Convolutional Neural Network (CNN) model for sound recognition in the battlefield. The mel frequency cepstral coefficients (MFCCs) features is used in this research to distinguish five types of sound; soldiers marching sound, plane sound, refiles sound, military vehicle sound, and missile launchers sound. The results showed that the CNN model accomplished the mission with an accuracy of 95.3% on testing data, while it showed 93.6% of accuracy on the outlet or unseen data. As a novel attempt and idea, the results were so promising.

Keywords: Convolutional Neural Network (CNN), Deep Learning, Sounds Recognition, MFCC, Sound Classification, Battlefield Sounds

1. Introduction

The problem of the classification of Military sounds did not receive much attention from researchers in the past years. Up to date, and up to our knowledge, signal processing and machine learning techniques have not been applied to this problem. Some small attempts were carried by using wavelet filter banks [1], [2] and most recently Deep Neural Networks (DNN) [3], [4]. After deep investigation, we found that CNN as a deep learner is very useful to tackle the problem under study, which is mainly sound classification and identification [5], [6]. CNN can capture the energy patterns through time and frequency when applied on spectrogram [7], [8]. The capturing of the energy pattern important to differentiate between various similar noisy sound sources, such as engines and jackhammers [5]. Moreover, using CNN kernels (filters) with a small receptive field, the CNN model will successfully learn and later identify Spectro temporal patterns that represent different sound classes despite the masked part of the sound (in time/frequency) by other disturbing sources (noise).

The firm learning of a deep neural network, with high model capacity, are subject to the availability of huge amounts of training data. The availability of training data is very important to learn the non-linear function do that it yields high classification accuracy on unseen data in Artificial Intelligence (AI) domain [9], [10], [11], and machine learning fields [12], [13], [14], [15]. A potential clarification for the restricted investigation of CNNs and the trouble to enhance easier models is the overall shortage of labeled datasets designed for sound classification [5].

In this paper, we present a deep Convolutional Neural Network (CNN) for automatic identification and classification of the Military sounds. The CNN model capable of distinguishing five types of military sound; Soldiers Marching Sound, Drone or Plane Sound, Refiles Sound, Military Vehicle Sound, and Missile Launchers Sound.

This paper describes a distinguished idea in using sound features and its recognition approaches for military application. The research idea in this paper is competitive, novel, and could be improved to be used in many other fields. The paper investigates the dependency on sound features for determine the source and type of sound especially in blinded sites.

The rest of the paper gives an overview of the related

E-mail address: khalids@yu.edu.jo, fedaaomari75@gmail.com, hindawi@jadara.edu.jo

work in Section 2, the methodology and the model architecture are introduced in Section 3. The model evaluation criteria are introduced in Section 4, followed by the experimental results in Section 5. Finally, the conclusion and future work in Section 6.

2. Related Work

In [4], the authors used CNN in classifying short audio clips of environmental sounds on a dataset of 2,000 short (5 seconds) environmental recording involving 50 similarly adjusted classes of sounds in 5 significant gatherings (creatures, regular soundscapes and water sounds, human non-discourse sounds, inside/homegrown sounds, and outside/metropolitan noises).

In [16], an attempt to identify bird species in a new method for the audio classification based on CNN. They used an audio dataset, which contains more than 33,000 registrations of 999 different types. Authors in [16], gained an average accuracy of 67% when predicting the main types per unit, and scored 56% when using background types as additional prediction targets.

In [17], the researchers applied a deep learning task on a robotic designed for heart sound verification. The target was to recognize anomalies in the sounds of the heart. They describe an automated cardiac sound classification algorithm that combines the use of time-frequency thermal map representations with CNN. In this research, a total of 4,430 records from 1,072 subjects, resulting in 30 hours of sound recordings for the human heart. A total dataset of 1,277 heart recordings for 308 people was isolated to be used as parked test data to assess challenge submissions. The overall score of 84% was achieved using a single convolutional neural network.

In [18], the authors applied Deep Neural Networks (DNN), CNN, and Recurrent Neural Network (RNN) to address the problem of cough detection. They evaluated the performance of the two networks and compared them to CNN in identifying cough sounds. They produced an average of 40 cough sounds, which yields a total of 627 cough examples. CNN showed higher specificity by 92.7% whereas the RNN attained higher sensitivity of 87.7%.

The researchers in [19], [1] introduced a novel heart sound classification model based on machine learning. In this model and specifically, in the classification stage; each feature vector was classified into "normal", "abnormal". The model used a robust feature representation generated by a wavelet-based CNN, and Support Vector Machine (SVM) of each test case recording [7], [20]. In addition, their model combined physiological and spectral features to represent the characteristics of the entire test recording. Accuracy results achieved were 82% of sensitivity, and a specificity of 82%, 85%, and 78%, respectively. The test was carried on a hidden challenge testing set.

A musical instrument classification was carried by [21] using CNN. The proposed classifier outperformed the baseline result from traditional handcrafted features and classifiers.

The study in [22] presented two main contributions: the first one is a CNN architecture for environmental sound classification, and the second one is the sound audio augmentation to get more audio files and overcome the problem of huge sound dataset. Actually, they explored the influence of different augmentations on CNN performance before applying it. These data augmentation process enhanced the model of environmental sound classification. The dataset used in this research was the Urban-Sound8K dataset [23]. The dataset contains 8,732 sound files with a duration of 4s taken from the field, the achieved accuracy based on this dataset was 74%.

TABLE I. Summary of Related Work

Ref.	Subject	Algo.	Dataset	Accur.
[4]	Classifying short environmental sounds	CNNs	2,000 short recordings	NA
[16]	Audio classification of bird species	CNNs	33,000 recordings of 999 types	68%
[17]	Deep learning for heart sounds anomalies	CNN	4,430 records of 1,072 groups	84%
[18]	Two different DNN for cough detection	CNNs and RNN	Total of 627 cough examples	CNN= 92.7% RNN= 87.7%.
[19]	Heart sound classification	CNNs and SVM	NA	Fscore= 81.2%
[21]	Music classification method	CNNs	NA	Exceeds baseline
[23]	Sound classification with audio augmentation	CNNs	Short 8,732 Audio files	NA
[24]	Multiple sound classification and detection	SVM	NA	96.7%
[25]	Classifying flying motors sounds	CNNs	745 voices of 9 major groups	87%

In [24], the researchers proposed a novel machine learning

(ML) model for detection and classification of amateur drones (ADr) sounds noisy environment. They used Support Vector Machine (SVM) with accuracy around 96.7% accuracy for ADr detection.

In [25], the flying motors were tackled based on CNN for classification. The researchers in [25] classify the syllables of the sounds of flying motors consisting of 745 voices groped into nine major groups. The accuracy archived was 87%. Table I summarizes our literature.

3. Research Methodology

In this paper, we tackled the problem of identifying military sounds that may be found on the battlefield. As an initial attempt, we focused on five types of famous military sounds like; soldiers marching sound, plane sound, refiles sound, military vehicle sound, and missile launchers sound. The proposed methodology consists of the following steps:

1) Data Collection and Sounds Labeling.
2) Data Preprocessing.
3) Sound Features Extraction.
4) Training, Validation, and Testing.
5) Results and Evaluation.

Figure 1 is a graphical representation of the proposed methodology. After phase 5 an evaluation of the resulted and its satisfactory and possible improvement is to be judged so that we can tune the CNN parameters and back again to the training and validation step.

In the subsequent sections, we will explain in detail each phase that appears in Figure 1.

A. Dataset Collection and Sounds Labeling

A set of 754 recordings from different subjects and from different repositories were collected from the internet, and its sounds and videos repositories were the main sources like the sources in[1] [2] [3] [4] [5] [6] [7] [8]. The data collected contained sounds of rifles, planes, marching soldiers (regular and irregular marching), military vehicles, and missile launchers. Figure 2 is a screenshot of the folder that contains the collected data, every file was named based on its type then numbered.

Since the recorded sounds were gathered from various public storehouses, there were a few varieties in the recording conditions, testing rate, frequencies, and commotion levels related with each recording. These sounds must be pre-prepared and formally dressed before the training stage which is entirely clarified in the next subsection (3-B). In these sounds, the military sound occasions were marked as a source of perspective for regulated supervised learning and to quantify the exhibition of the proposed military sound classifier.

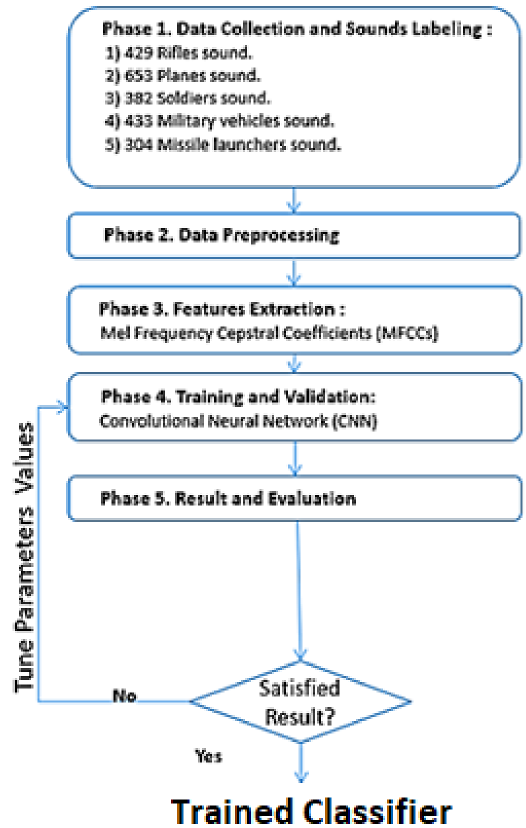

Figure 1. Methodology Graphical Representation

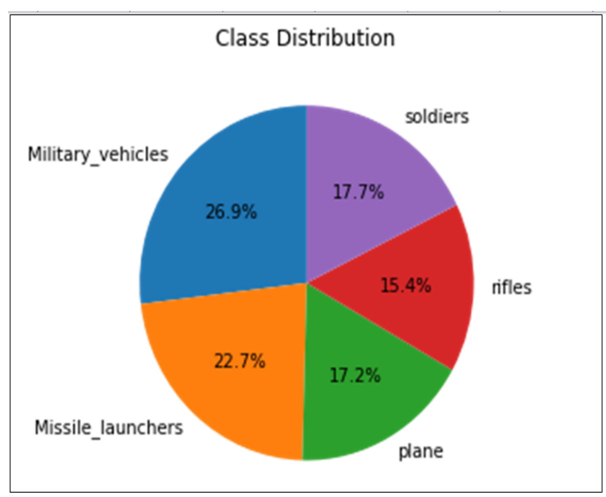

Figure 2. Distribution of recordings

[1] https://www.youtube.com/watch?v=ZJeOQ7BOkRI
[2] https://retired.sounddogs.com/results.asp?Type=1&categoryid=1005&subcategoryid=58
[3] https://www.youtube.com/watch?v=xX3MjZo4Ufg
[4] https://www.freesoundeffec ts.com/free-sounds/gun-10081/
[5] http://soundbible.com/tags-gun.html
[6] https://www.fesliyanstudios.com/royalty-free-sound-effectsdownload/gun-shooting-300
[7] https://www.youtube.com/watch?v=YDNVI-f4CQQ
[8] https://www.youtube.com/watch?v=VHoOg4twnO0

We labeled about 2,201 events from the recordings as military-sounds based on the five predetermined types that we mentioned earlier. The military-sound events presented in the recordings included sounds of Rifles, Drones or Fighter planes, Soldiers Marching (Regular and Irregular Marching), Military Vehicles, and Missile Launchers. Table II shows the final labeling distribution on the recording's events labels in the formulation of the dataset.

TABLE II. Distribution of label in the dataset

Labels	# of Recordings
Rifles	429
Planes	653
Planes	653
Soldiers marching	382
Military vehicles	433
Missile launchers	304
Total:	**2,201**

Figure 2 shows the percentage distribution of military sound recordings based on the sum of seconds for each class.

For evaluation proposes we split the recordings of each type into training and testing samples by 70% for training and leave 30% for testing and performance evaluation. Table III shows the split data samples.

TABLE III. Distribution of label in the dataset

Dataset	Training data Split	Testing data Split
Rifles	294	135
Planes	404	241
Planes	653	
Soldiers	202	188
Military Vehicles	249	184
Missile Launchers	210	94
Total:	**1,359**	**842**

B. Data Preprocessing

As we said before and due to the variations found in data that is to be used for training and testing, all the recorded sounds were first preprocessed to reduce varieties. By applying the following steps, we uniform the audio files without affecting its basic features:

1) **Resampling**: we resampled all the recordings to 16000 Hz.
2) **Normalization**: in this step, the recorded sounds amplitude was scaled to be in the range between -1 and 1 at the meantime preserving its distribution. The main statistical parameters of the recorded sounds remain the same, this is called (Normalization).
3) **Segmentation**: the audio signals were segmented into 200ms frames with an overlap ratio of 50% between successive segments.
4) **Silence Truncation**: it is valuable to eliminate silence pieces of the audios to guarantee that all further handling is performed distinctly on pieces of the signals containing the sound event. Nonetheless, this requires the silence evacuation way to be simpler than the further preparing stages. To accomplish this, in this paper, silence parts and noisy parts were identified and taken out based on the mean of deviation in the past examples with the assist of a specific threshold. Accordingly, further preparing was performed just when the normal energy of the frame was over this threshold as in [26].

C. Features Extraction

Many features can be extracted from the sound wave like Linear Predictive Coding (LPC), Spectral Centroid, Spectral Roll off (SCSR), Mel-Frequency Cepstral Coefficients (MFCCs)..., etc. However, in this study, we will use the MFCCs features which are mostly used in such problems.

MFCCs are one of the important audio features and majorly used whenever working on audio signals. MFCCs of signals is a small set of features (usually about 10–20) which concisely describe the overall shape of a spectral envelope[9]. Figure 3, is an example of a spectrum for MFCC taken from collected Audio.

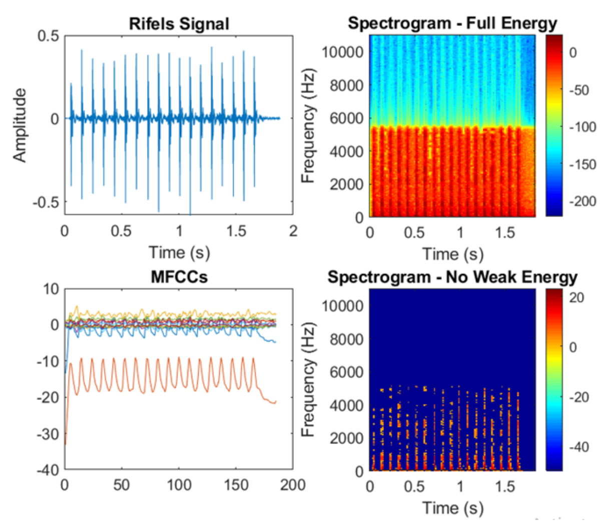

Figure 3. A sample of rifles Audio Signal Properties (Signal, MFCC's, Full Energy Spectrum, and Spectrum without Weak Energy)

Figure 4, is another sample of the same sound type but, different properties (Pitch, Amplitude, Phase, Frequency...etc) which mean different energy, amplitude, and frequencies even it is of the same type.

From Figure 3, and Figure 4, we notice the similarity in the MFCCs even the spectrum is different, while In

[9]https://towardsdatascience.com/extract-features-of-music-75a3f9bc265d

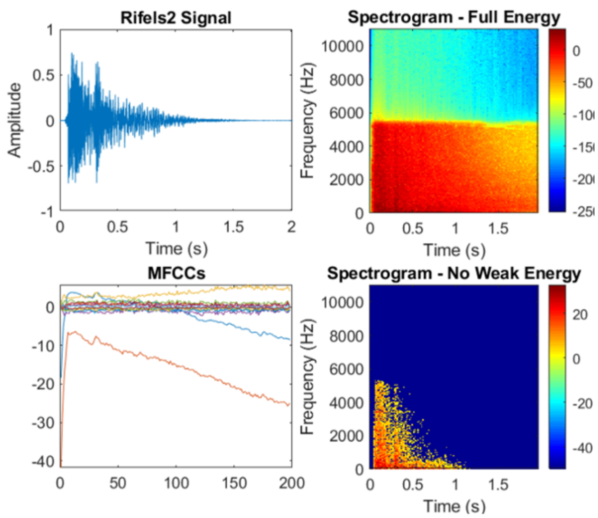

Figure 4. A sample of another type of rifles, the same audio properties were extracted

Figure 5, which represents a different sound we notice the divergence in the MFCCs compared to Refiles sound.

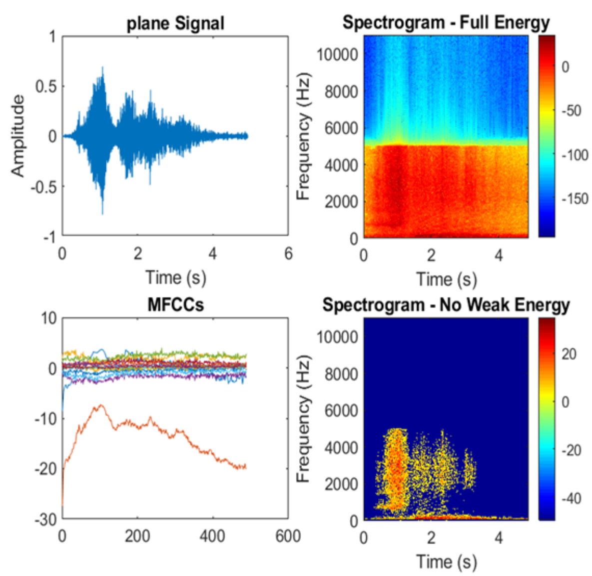

Figure 5. A sample of fighting plane Audio Signal Properties (Signal, MFCC's, Full Energy Spectrum, and Spectrum without Weak Energy)

In Figure 6, and Figure 7, represent both the missile launchers and military vehicles respectively. It is clear from the figures that the MFCC features, and spectrogram analysis is different especially in MFCC's which makes it possible for CNN to learn very well. Moreover, to generalize the CNN learner for these five military sounds, the collected audio corpus contains multiple audios for the same class from different resources.

Figure 3, Figure 4, Figure 5, Figure 6, and Figure 7, are samples taken from each class to show the differences in their properties.

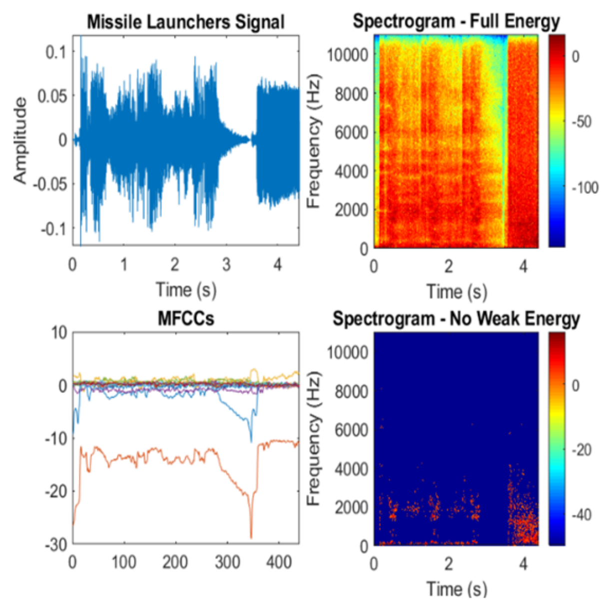

Figure 6. A sample of Military Vehicles Audio Signal Properties (Signal, MFCC's, Full Energy Spectrum, and Spectrum Without Weak Energy)

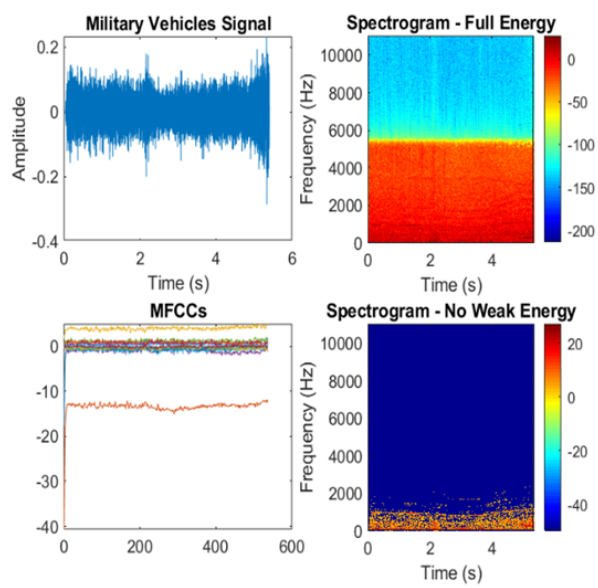

Figure 7. A sample of fighting plane Audio Signal Properties (Signal, MFCC's, Full Energy Spectrum, and Spectrum without Weak Energy)

D. Training and Testing

Deep Learning (DL) is a part of the Machine Learning family, where learning can be supervised, unsupervised, or semi-supervised. DL uses several nonlinear processing layers, so the model depth will be referred to as the number of layers that data passed through.

Since DL can be implemented through several techniques but, researchers recommended and trusted CNN for this kind of problems (speech recognition, and sound detection), which encourage us to use it.

E. CNN Model Architecture

The detailed architecture of CNN is shown in Figure 8. The activation function "RELU" is applied to each layer except for the last dense layer, the "softmax" activation function was applied. A sampling rate of 16 kHz, the optimizer "Adam" with learning rate (LR=0.001) and lose function "categorical_crossentropy".

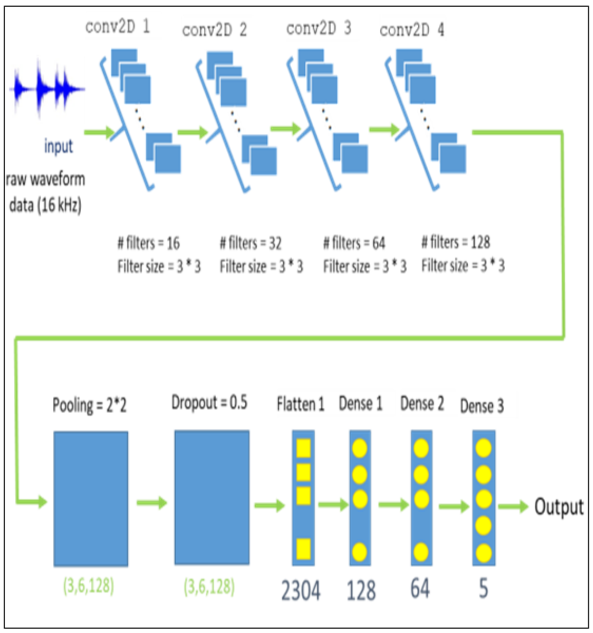

Figure 8. Detailed architecture of CNN

We apply four time-convolutional layers, in the first convolutional-2D layer, 16 filters with filter size 3×3 were applied. In the second convolutional-2D layer, 32 filters with filter size 3×3 were applied. Meanwhile, in the third convolutional-2D layer, 64 filters with filter size 3×3 were applied. In the fourth convolutional-2D layer, 128 filters with filter size 3×3 were applied. After that, a Max Pooling-2D with pool size 2×2 was applied. 50% of the dropout was set, then flatten was applied. Then, three dense layers are set with 128, 64, and 5 neurons, respectively.

Table IV contains all the parameters necessary for the deep learning model. These parameters were tuned to get the maximum accuracy.

F. Training and Validation

Figure 9 is a screenshot of the model training and validation for CNN model, it is for the last 3 epochs of size 40, and a batch size of 64.

TABLE IV. The parameters used in the DL model

Model Parameters	Model Parameters Value
NUMBER DENSES	192
BATCH SIZE	64
NUMBER of LAYERS	10
DROPOUT	0.5
OPTIMIZER	Adam
NB EPOCHS	40
ACTIVATION FUNCTION	RELU
LOSS FUNCTION	Categorical-cross entropy

```
Epoch 37/40

76246/76246 [==========================] - 149s 2ms/step -
loss: 0.1408 - acc: 0.9482 - val_loss: 0.1342 - val_acc: 0.
9525

Epoch 00037: val_acc improved from 0.94917 to 0.95251, savi
ng model to models\conv.model

Epoch 38/40

76246/76246 [==========================] - 148s 2ms/step -
loss: 0.1370 - acc: 0.9504 - val_loss: 0.1416 - val_acc: 0.
9503

Epoch 00038: val_acc did not improve from 0.95251

Epoch 39/40

76246/76246 [==========================] - 140s 2ms/step -
loss:| 0.1345 - acc: 0.9500 - val_loss: 0.1431 - val_acc: 0.
9482

Epoch 00039: val_acc did not improve from 0.95251

Epoch 40/40

76246/76246 [==========================] - 139s 2ms/step -
loss: 0.1319 - acc: 0.9511 - val_loss: 0.1504 - val_acc: 0.
9469

Epoch 00040: val_acc did not improve from 0.95251
```

Figure 9. Screen Shot of Model Training and Validation

4. Model Evaluation

Evaluation of the model is an important step for any learning model. One of the most used metrics is the accuracy score metric [27]. The accuracy is a good metric to evaluate the model. Sometimes using the accuracy score metric alone is not enough [28], [29]. Therefore, some other metrics were used such as F1-score, Precision, and Recall. The next sections explain these measurements.

A. Precision

Precision was used to check the classifier's ability to return relevant instances only. Equation 1 represents this metric. Simply it is the number of correct positive results divided by the number of the positive results predicted by the algorithm.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (1)$$

B. Recall

Recall (also known as sensitivity) is used to know the classifier's ability to identify all relevant instances. The equation used to calculate it is Equation 2. It is the number of correct positive results divided by the number of all relevant samples.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (2)$$

C. F-Measure

F-Measure is used to combine Precision and Recall into one measurement tool. It uses the harmonic mean to combine them. Equation 3 is used to calculate this measure.

$$F1 = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (3)$$

D. Accuracy

Accuracy is the most popular used performance measure. It is known as the ratio of correctly predicted observation of the total observations. Accuracy would give us a good indicator and strong evaluation only when the dataset is balanced. Since our data is not balanced, we used the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) counts to measure the accuracy. Equation 4 represents the definition of accuracy and Equation 5 is the accuracy based on previous counts when the dataset is not balanced.

$$Accuracy = \frac{NumberofCorrectPredictions}{TotalNumberofPredictions} \quad (4)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

5. Results and Discussion

Referring to our experiments, we found that, the results showed that the CNN model gave very good results in solving the problem of military sound classification and identification. CNN achieved an accuracy of 95.6% with training, achieved an accuracy of 95.2% with validation, and achieved an accuracy of 95.3% with testing (unseen data). Table V shows more details about performance measures.

After several experiments using different parameters, the highest accuracy was achieved, it was about 95.3% using CNN with unseen data. We tried to tune the learner parameters but, no further improvement appeared.

TABLE V. Performance Measures

Performance measures	Testing Accuracy
Precision	94%
Recall	97%
F-Measure	95%
Accuracy	95.3%

One of the ways to review and visualize the performance of CNN as a DL model through the graphical plots. Figure 10 shows the training and validation accuracy for each epoch. From the plot and based on several experiments, we stopped training at epoch number 40 for the CNN model, because after these numbers the models start to overfit the data.

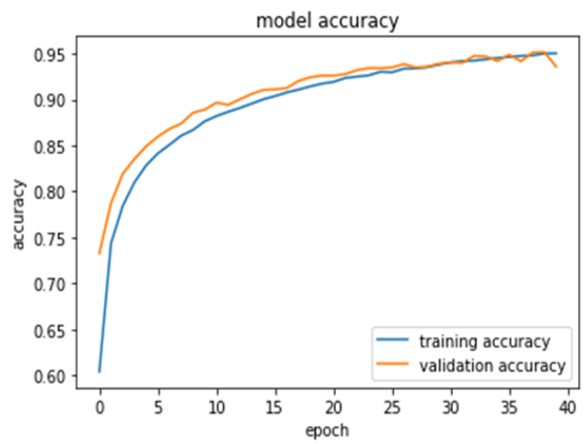

Figure 10. Training and Validation Accuracy

In Figure 11, the training and validation loss for each epoch is represented. Validation and testing loss keeps decreasing while the CNN model is running tell epoch number 40. It is clear also that after 40 epochs the validation loss starts to increase, which forces us to stop CNN training.

An important outcome was the confusion matrix, it is used to describe the performance of a classifier on test cases usually the test part of a dataset. As shown in Figure 12, the CNN model results in 47 wrong predictions and 958 right predictions. The wrong predictions are the sum of the cell's upper and lower primary diagonal, while the right prediction is the sum on the primary diagonal. as we mentioned earlier, the total of testing data was 1,001 recordings.

From the confusion matrix in Figure 12, most of the classifications were concentrated on the main diagonal, which means the classification achieved a high degree of accuracy.

For the sack of fact, a limitation on the research done in this paper should be mentioned that the real time application of research idea is not considered or discussed. The paper discussed the idea of predicting hidden enemy troops.

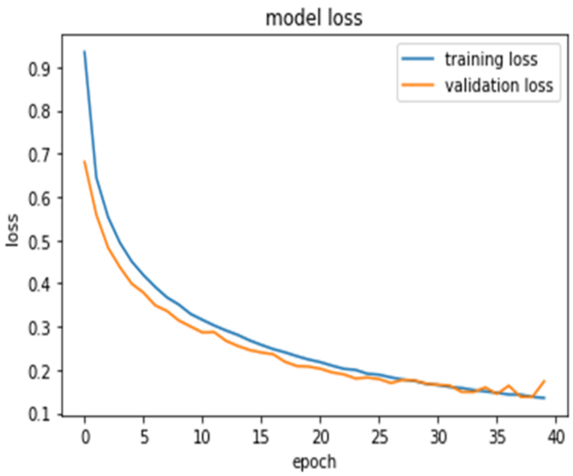

Figure 11. Training and Validation Loss

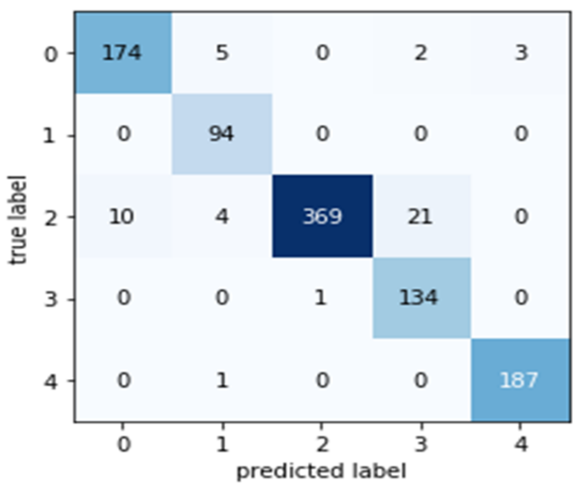

Figure 12. Confusion Matrix

6. Conclusion

The importance of such research appears in sensing hidden enemies on the battlefield, especially when military raids occur. The enemies may be far away or hidden but, sounds are heard. By identifying the type of sound, a lot of information could be gained in further physical processing such as the approximate location, distance, and sound direction could be predicted. Moreover, distinguishing sounds will assist soldiers in alerting their military troops.

The results of the CNN model showed excellent performance and accuracy of 95.3% with the testing data and 93.6% on unseen or outlet data. By looking at the results, the CNN model motivated us to extend the audio corpus such that it contains more types in an attempt to generalize the model for all battlefield sounds.

References

[1] J. Salamon and J. P. Bello, "Feature learning with deep scattering for urban sound analysis," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 724–728.

[2] J. T. Geiger and K. Helwani, "Improving event detection for audio surveillance using Gabor filterbank features," *2015 23rd European Signal Processing Conference, EUSIPCO 2015*, pp. 714–718, 2015.

[3] E. Cakir, T. Heittola, H. Huttunen, and T. Virtanen, "Polyphonic sound event detection using multi label deep neural networks," *Proceedings of the International Joint Conference on Neural Networks*, vol. 2015-September, 2015.

[4] K. J. Piczak, "2015 Ieee International Workshop on Machine Learning for Signal Processing Environmental Sound Classification With Convolutional Neural Networks," *IEEE international workshop on machine learning for signal processing, Boston, USA*, 2015.

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "A B7CEDGF HIB7PRQTSUDGQICWVYX HIB edCdSISIXvg5r ' CdQTw XvefCdS," *proc. OF THE IEEE*, 1998. [Online]. Available: http://ieeexplore.ieee.org/document/726791/full-text-section

[6] K. M. Nahar, B. Abul-Huda, A. Abu Naser, and R. M. Al-Khatib, "Twins and Similar Faces Recognition Using Geometric and Photometric Features with Transfer Learning," *International Journal of Computing and Digital System*, no. forthcoming, 2021.

[7] K. M. Nahar, R. M. Al-Khatib, M. A. Al-Shannaq, and M. M. Barhoush, "An efficient holy Quran recitation recognizer based on SVM learning model," *Jordanian Journal of Computers and Information Technology (JJCIT)*, vol. 6, no. 04, 2020.

[8] F. Alkhateeb, R. M. Al-Khatib, and I. A. Doush, "A survey for recent applications and variants of nature-inspired immune search algorithm," *International Journal of Computer Applications in Technology*, vol. 63, no. 4, pp. 354–370, 2020.

[9] R. M. Al-Khatib, T. Zerrouki, M. M. Abu Shquier, A. Balla, and A. Al-Khateeb, "A New Enhanced Arabic Light Stemmer for IR in Medical Documents," *Computers, Materials Continua*, vol. 68, no. 1, pp. 1255–1269, 2021.

[10] I. Abu Doush, M. A. Al-Betar, M. A. Awadallah, A. I. Hammouri, R. M. Al-Khatib, S. ElMustafa, and H. ALkhraisat, "Harmony search algorithm for patient admission scheduling problem," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 540–553, 2020.

[11] I. Abu Doush, I. Damaj, M. A. Al-Betar, M. A. Awadallah, R. M. Al-Khatib, A. E. Alchalabi, and A. L. Bolaji, "A survey on accessible context-aware systems," in *Technological Trends in Improved Mobility of the Visually Impaired*. Springer, 2020, pp. 29–63.

[12] K. Nahar, "Off-line arabic hand-writing recognition using artificial neural network with genetics algorithm," *International Arab Journal of Information Technology*, vol. 15, no. 4, pp. 701–707, 2018.

[13] H. Rashaideh, A. Sawaie, M. A. Al-Betar, L. M. Abualigah, M. M. Al-Laham, R. M. Al-Khatib, and M. Braik, "A Grey Wolf Optimizer for Text Document Clustering," *Journal of Intelligent Systems*, vol. 29, no. 1, 2020.

[14] K. M. Nahar, M. Al-shannaq, R. Alshorman, R. M. Al-Khatib, and M. A. Ot.tom, "Handicapped Wheelchair Movements Using Discrete Arabic Command Recognition," *Scientific Journal of King Faisal University (Basic and Applied Sciences)*, vol. 21, no. 1, pp. 171–184, 2020.

[15] R. M. Al-Khatib, M. A. Al-Betar, M. A. Awadallah, K. M. Nahar,

M. M. A. Shquier, A. M. Manasrah, and A. B. Doumi, "MGA-TSP: modernised genetic algorithm for the travelling salesman problem," *International Journal of Reasoning-based Intelligent Systems*, vol. 11, no. 3, pp. 215–226, 2019.

[16] E. Sprengel, M. Jaggi, Y. Kilcher, and T. Hofmann, "Audio based bird species identification using deep learning techniques," *CEUR Workshop Proceedings*, vol. 1609, pp. 547–559, 2016.

[17] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, "Recognizing abnormal heart sounds using deep learning," *CEUR Workshop Proceedings*, vol. 1891, pp. 13–19, 2017.

[18] J. Amoh and K. Odame, "Deep neural networks for identifying cough sounds," *IEEE transactions on biomedical circuits and systems*, vol. 10, no. 5, pp. 1003–1011, 2016.

[19] M. Tschannen, T. Kramer, G. Marti, M. Heinzmann, and T. Wiatowski, "Heart sound classification using deep structured features," *Computing in Cardiology*, vol. 43, pp. 565–568, 2016.

[20] K. M. Nahar, R. M. Al-Khatib, M. Al-Shannaq, M. Daradkeh, and R. Malkawi, "Direct Text Classifier for Thematic Arabic Discourse Documents," *International Arab Journal of Information Technology (IAJIT)*, vol. 17, no. 3, 2020.

[21] T. Park and T. Lee, "Musical instrument sound classification with deep convolutional neural network using feature fusion approach," 2015. [Online]. Available: http://arxiv.org/abs/1512.07370

[22] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal processing letters*, vol. 24, no. 3, pp. 279–283, 2017.

[23] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research," *MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia*, pp. 1041–1044, 2014.

[24] M. Z. Anwar, Z. Kaleem, and A. Jamalipour, "Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2526–2534, 2019.

[25] R. Girshick, "Fast R-CNN," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015 International Conference on Computer Vision, ICCV 2015, pp. 1440–1448, 2015.

[26] R. X. Adhi Pramono, S. Anas Imtiaz, and E. Rodriguez-Villegas, "Automatic Identification of Cough Events from Acoustic Signals," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 217–220, 2019.

[27] R. M. Al-Khatib, R. Abdullah, and N. Abdul Rashid, "A comparative taxonomy of parallel algorithms for RNA secondary structure prediction," *Evolutionary Bioinformatics*, vol. 2010, no. 6, pp. 27–45, 2010.

[28] K. M. Nahar, R. M. Al-Khatib, M. Barhoush, and A. A. Halin, "Mpf-leach: modified probability function for cluster head election in leach protocol," *International Journal of Computer Applications in Technology*, vol. 60, no. 3, pp. 267–280, 2019.

[29] R. M. Al-Khatib, N. Abdul Rashid, and R. Abdullah, "Thermodynamic heuristics with case-based reasoning: Combined insights for rna pseudoknot secondary structure," *Journal of Biomolecular Structure and Dynamics*, vol. 29, no. 1, pp. 1–26, 2011.

Khalid M.O. Nahar is an associate professor in the Department of Computer Sciences-Faculty of IT, Yarmouk University. He has a PhD in Computer Science and Engineering from King Fahd University of Petroleum and Minerals (KFUPM), KSA. His research interests include Speech Processing, NLP, AI, IoT, DL, and ML. Dr.Khalid was the IT-dean assistant for Quality Assurance in 2019. For now, Dr.Khalid is the chair of training and development department – Accreditation and Quality Assurance Center, Yarmouk University.

Fedaa M. Al-Omari is a computer science teacher in the directorate of education of the district of Qasbah Irbid, Jordan ministry of education. She had a lot of concerns in AI and ML applications. She implements some important applications and research ideas in MATLAB, C++, Python…etc.

Nouh Alhindawi is an associate Professor in Computer Science & Software Engineering Departments at Jadara University, Jordan since 2013. He obtained his Ph.D. in Computer Science-Software Engineering from Kent State University, USA in 2013 with Dr. Jonathan Maletic as his advisor. His research interests are in Software Engineering area, Information Retrieval (IR), and Latent Semantic Indexing (LSI); Using Information Retrieval Approaches for Improving Software Comprehension. Now he is the director of Information Technology and Electronic Transformation Directorate at Ministry Of Higher Education And Scientific Research – Jordan.

Mustafa Banikhalaf received the Ph.D. degree from the School of Computing, Edinburgh Napier University, in 2013. He is currently an Associate Professor in the Computer Science Department, Yarmouk uiversity.He has been working on the area of group communications, mobile and vehicular adhoc networks, Internet of Things (IoTs) and high-performance networks.