International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Musical Instruments Classification using Pre-Trained Model

S. Prabavathy[1], V. Rathikarani[2], P. Dhanalakshmi[3]

[1]Research Scholar, Department of Computer and Information Science,
[2]Assistant Professor, Department of Computer Science and Engineering,
[3]Professor, Department of Computer Science and Engineering,
Annamalai University, Chidambaram, Tamilnadu, India.

Abstract - *Musical instruments classification using machine is a very challenging task. Huge manual process required to classify the musical instruments. In this paper, the musical instruments classifies from a several acoustic features that include AlexNet. SVM and kNN are two modeling techniques used for classification. The proposed work compares the performance of SVM with kNN. Identifying the musical instruments and computing its accuracy is performed with the help of SVM and kNN, using AlexNet with SVM highest accuracy rate of 99% yields in classifying musical instruments. The proposed system tested sixteen musical instruments from four musical instrument families to find out the accuracy level using SVM and kNN.*

Keywords: AlexNet, Feature Extraction, Musical Instruments Classification, K-Nearest Neighborhood (kNN), Support Vector Machine (SVM).

1 INTRODUCTION

A musical instrument essentially converts energy supplied by the player into sound waves with characteristics that to a large extent are controllable by the player. The basic characteristics of the tones are pitch, loudness, duration and timbre. One of the basic functions of a musical instrument is to produce tones of the desired pitch [1]. Research on the automatic classification of musical instrument sounds has focused, for a long time, on classifying isolated notes from different instruments. This is an approach that has a very important trade-off: gain simplicity and tractability, as there is no need to first separate the sounds from a mixture, but we lose contextual and time-dependent cues that can be exploited as relevant features when classifying musical sounds in complex mixtures [2]. Deep convolutional neural network models may take days or even weeks to train on huge datasets. A way to shortcut this process is to reuse the model weights from pre-trained models that were developed for computer vision benchmark datasets such as ImageNet image recognition tasks [3]. AlexNet successfully demonstrated the capability of the convolutional neural network model in the domain, and kindled a fire that resulted in many more improvements and innovations, many demonstrated on the same ILSVRC task in subsequent years [3]. Christian Szegedy, et al. from Google achieved top results for object detection with their GoogleNet model that made use of the inception module and architecture. This is a very simple and powerful architectural unit that allows the model to learn not only parallel filters of the same size, but parallel filters of different sizes, allowing learning at multiple scales. Support Vector Machine have become a well established tool within machine learning. They work well in practice and have now been used across a wide range of applications from recognizing hand-written digits, to face identification, text categorization, bioinformatics, and database marketing [4]. Conceptually, they have many advantages justifying their popularity. When kNN is used for calculation, the output can be calculated as the class with the highest frequency from the k-most similar instances. Each instance in essence votes for their class and the class with the most votes is taken as the prediction. Class probabilities can be calculated as the normalized frequency of samples that belong to each class in the set of k most similar instances for a new data instance [5]. Musical instruments have their characteristic shapes and sizes, and gain insight into their possibilities and limitations. The interest here is in studying the varieties of sound which can be produced by musical instruments. The function of all musical instruments is to make waves. Instead of a variation in the height of the water surface, the wave generated by a musical instrument is an invisible variation in the pressure of the air. This type of wave carries the sound of an instrument to the ear of the listener; we call it a sound wave [6]. In this work, SVM and kNN are the models used to classify the musical instruments, AlexNet is used for extracting the features.

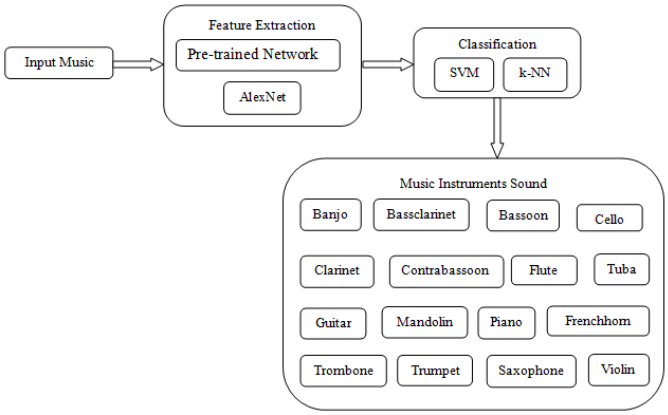

Fig -1 Block diagram of the proposed system

Sixteen musical instruments such as cello, banjo, guitar, violin and mandolin from string, bassoon, contrabassoon, clarinet, bass clarinet, saxophone and flute from woodwind, trombone, frenchhorn, tuba and trumpet from brass and piano from keyboard instrument family with 1284 music samples were collected from the online musical instrument database. The block diagram of the proposed work shows in fig. 1.

2. LITERATURE REVIEW

In [7] musical instrument sounds classified automatic, the features extracted by MFCC to model the tones and probabilistic neural networks used as classifier. Direct and hierarchical classification structures were compared, the direct instrument classification, did perform better than the hierarchical approach. In [8] the accuracy of a model trained using MFCC and GFCC, by comparing these two descriptors, MFCC perform better than GFCC but the difference is far from being relevant. By training the SVM model using the sig files musical instruments classified with good results. In [9] the modified Alexnet architecture used to classify diabetic retinopathy images; Softmax and Rectified Linear Activation Unit (ReLU) layers are used to obtain a high level of accuracy. The performance of the proposed algorithm has been validated using Messidor database, modified Alexnet architecture is used to categorize fundus images. In [10] the combination of neural network model and SVM was applied to identify musical instruments by giving one note from sets of orchestral musical sounds, good results achieved by using SVM model. In [11] hybrid GMM and SVM approach is used to identify the speaker. The proposed classification system is formed by an ensemble of SVM binary classifiers which emerge from the GMMs of the baseline system using the Fisher kernel method. In [12] cluster-based tree algorithm is used to accelerate kNN classification without any presuppositions about the metric form and properties of a dissimilarity measure. kNN performs classification much faster than the condensation-based tree algorithm and gives high performance. In [13] to improve the efficiency of the k-NN classification by incorporating two novel ideas, the reduction of the template size and preprocessing method to preclude participation of a large portion of prototype patterns which are unlikely to match the test pattern. This work notably speeds up the classification without compromising accuracy.

3. FEATURE EXTRACTION

3.1 AlexNet

Alexnet consists of eight layers in which the first five is the convolutional layer and the remaining three is the fully connected layers. AlexNet uses different activation functions called the ReLU after every convolution layer, it also has new processing types called the dropout after fully connected layers 1 and 2. The input features reduced to 1024 before sending it to the fully connected layers by the arrangement of convolution layer and the pooling layer.

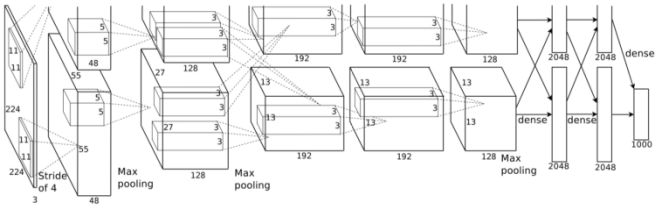

Fig -2 AlexNet architecture

In [14] fruit recognition is done by using CNN, AlexNet and GoogleNet with high accuracy with different running time and the recognition of nine different types of fruit in this proposed work.

4. CLASSIFIERS

4.1 Support vector machine (SVM)

SVM is used for pattern classification and for nonlinear regression. It constructs a linear model to estimate the decision function using non-linear class boundary which based on the support vectors. If the data are linearly divided, it trains the linear machines for a best hyperplane that split the data without mistake and into the highest distance between the hyperplane and the nearby training points, that are neighboring to the finest separating hyperplane are called as support vectors.

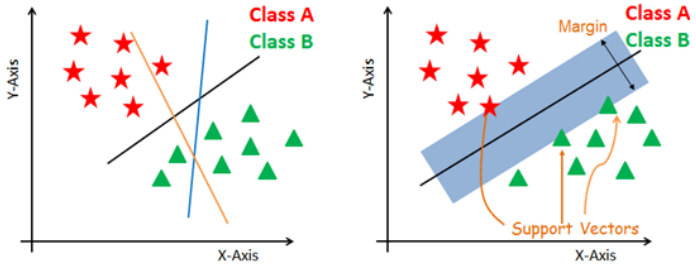

Fig -4.1 Process of SVM

Fig. 4 shows the process of the SVM. It maps the input data into a high dimensional feature space during some nonlinear mapping selected a priori.

4.2 K-Nearest Neighbor (kNN)

kNN algorithm is very effective and very simple. When this cassifier is used, the output is calculated like the class with the maximum frequency from the k-most related instances. Each instance in the essence votes for the classes and the classes with the majority votes is taken for prediction.

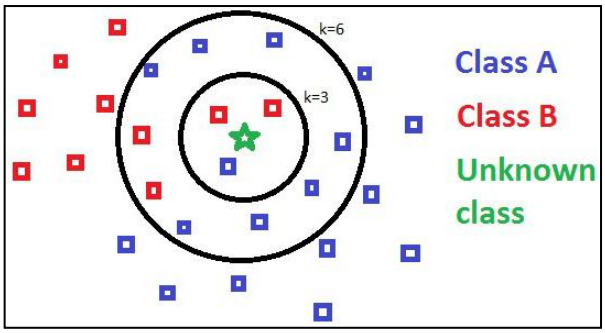

Fig -4.2 K-Nearest Neighbor

Class probabilities is calculated as the normalize rate of samples that belongs to each class in the set of k most similar instances for a fresh data instance [5]. Fig. 5 shows the process of K-Nearest Neighbor (kNN).

5. PROPOSED WORK

In this proposed work, each and every music samples are converted into spectrogram images with 610 x 450. The spectrogram images are resized into 224 x 224 and the stride 4 x 4 given as input for AlexNet. The input layer contains three feature map, the first is the convolution layer contains the activation function with 96 feature map and the size of the image is 55x55 where the filter size is 11x11 and the stride is 4 x 4. In the first max pooling the feature map remains as 96, the size of the image is 27 x 27, filter size is 3 x 3 and the stride changes to 2 x 2. In the second convolution layer the 256 feature map has been obtained. The image size is 23 x 23, the filter size is 5 x 5 and stride is 1 x 1. In the second pooling the feature map remains unchanged. The size of the feature map is 11 x 11, the filter size reduces to 3 x 3, and the stride is 2 x 2. In the third convolution layer the feature map is increased to 384, the image size is 9 x 9, the filter size is 3 x 3 and stride is 1 x 1. In the fourth convolution layer the feature remains unchanged; the image size is 7 x 7, filter size 3 x 3 with the stride of 1 x 1. The fifth convolution layer the feature map is 256, the image size is 5 x 5 and filter size is 3 x 3 with the stride of 1 x 1. The third max pooling the feature map remains unchanged, the size of 2 x 2, the feature size is 3 x 3, the stride 2 x 2.

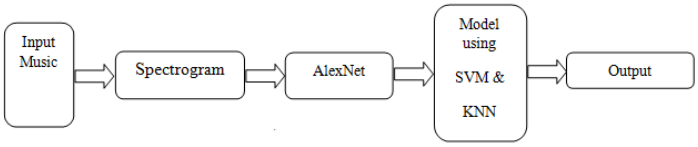

Fig -5 Block Diagram for Musical Instruments Classification

6. EXPERIMENTAL RESULTS

6.1 Dataset

The musical instruments sound data taken from the RWC database, musicbrainz.org, MINIM-UK musical instrument database, IRMAS: a dataset for instrument recognition in musical audio signals and NSynth dataset range from 1 sec. to 2 min. duration. 1284 music samples were collected from sixteen different musical instruments from four different families such as string, woodwind, keyboard and brass. 80% of musical samples were trained and 20% of data used for testing to classify the musical instruments.

6.2 Classification using AlexNet with SVM

In the proposed work the features were extracted by MFCC and the classifier SVM used to classify the musical instruments. The proposed system classifies the musical instrument with an accuracy of 99%. Fig. 6.1 shows musical instruments classification using MFCC with SVM.

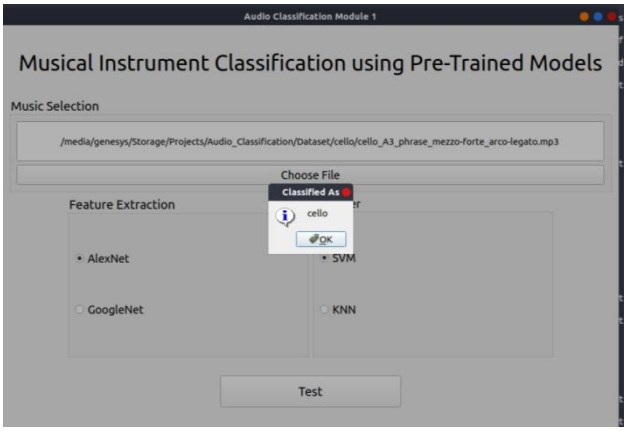

Fig -6.1 Musical Instruments Classification using AlexNet with SVM

6.3 Classification using AlexNet with kNN

In the proposed work the features were extracted by MFCC and the classifier kNN used to classify the musical instruments. The proposed system classifies the musical instrument with an accuracy of 98%. Fig. 6.2 shows musical instruments classification using MFCC with kNN.

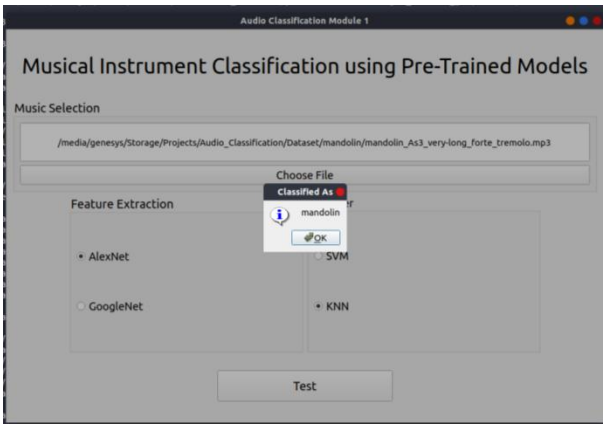

Fig -6.2 Musical Instruments Classification using AlexNet with kNN

7. PERFORMANCE MEASURES

The 16 musical instruments with different numbers of samples are considered for the proposed system. Table 1 shows the overall performance of precision, recall, F-Score and accuracy of musical instruments. The accuracy of each instrument calculated using the confusion matrix. The musical samples are trained and tested successfully; precision, F-Score, recall and accuracy of musical instruments are defined as

$$\textbf{Precision} = TP / (TP + FP)$$
$$\textbf{Recall} = TP / (TP + FN)$$
$$\textbf{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Table -1 Precision, F-Score, Recall, and Accuracy of Musical Instruments

Features	Classifiers	Precision (%)	Recall (%)	F-Score (%)	Accuracy (%)
ALEXNET	SVM	94.7	94.39	94.54475	99.37
	kNN	88.38	86.19	87.27126	98.16

Chart 1 shows the overall performance accuracy comparison of the proposed work.

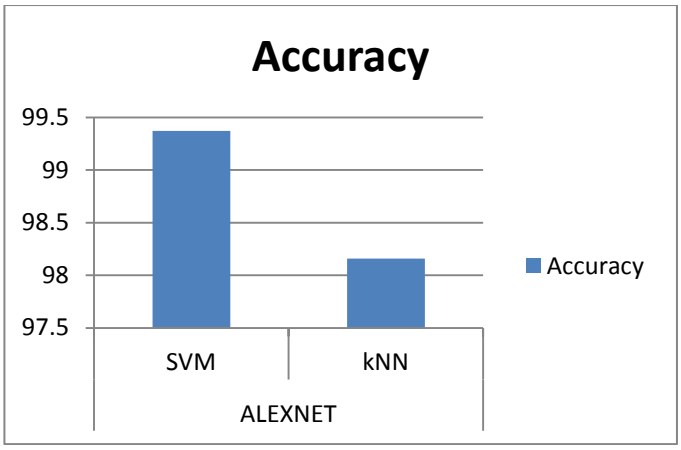

Chart 1 Accuracy comparison of proposed work

8. CONCLUSION

In this paper, automatic classification of musical instrument system has been proposed using SVM and kNN where AlexNet as features to classify the musical instruments. From the analysis, SVM shows an accuracy of 99.37% for AlexNet and kNN shows an accuracy of 98.16% using AlexNet. In this proposed work the performance of AlexNet gives good results. In future the classification of musical instrument system will develop using the other models.

9. REFERENCES

[1] Musical Sound, Instruments, and Equipment, By Panos Photinos
[2] Signal Processing Methods for Music Transcription edited by Anssi Klapuri, Manuel Davy
[3] Deep Learning for Computer Vision: Image Classification, Object Detection ..., By Jason Brownlee
[4] Learning with Support Vector Machines, By Colin Campbell, Yiming Ying
[5] Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch, By Jason Brownlee
[6] Musical Instruments: History, Technology, and Performance of Instruments of Western Music, By Professor of Musical Acoustics Donald Murray Campbell, Murray Campbell, Clive A. Greated, Arnold Myers, Senior Lecturer in Music and Director Arnold Myers.
[7] MUSTAFA SARIMOLLAOGLU, COSKUN BAYRAK, Musical Instrument Classification Using Neural Networks, Proceedings of the 5th WSEAS International Conference on Signal Processing, Istanbul, Turkey, May 27-29, 2006 (pp151-154)
[8] Daniele Scarano, Automatic Classification of Musical Instrument Samples, MASTER THESIS UPF / 2016

[9] T. Shanthi, R.S. Sabeenian, Modified Alexnet architecture for classification of diabetic retinopathy images, Computers and Electrical Engineering 76 (2019) 56–64

[10] Zhiwen Zhang(MSE), Hanze Tu(CCRMA), Yuan Li(CCRMA), Neural Network for Music Instrument Identification

[11] Shai Fine, JiH Navra'til, Ramesh A. Gopinath, A HYBRID GMM/SVM APPROACH TO SPEAKER IDENTIFICATION, 0-7803-7041 -410 1/$10.00 02001 IEEE

[12] Bin Zhang and S. N. Srihari, "Fast k-nearest neighbor classification using cluster-based trees," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 4, pp. 525-528, April 2004.

[13] Yingquan Wu, Krassimir Ianakiev, Venu Govindaraju, Improved k-nearest neighbor classification, Y. Wu et al. / Pattern Recognition 35 (2002) 2311–2318

[14] Azida Muhammad, Nur, Amelina Ab Nasir, Zaidah Ibrahim, and Nurbaity Sabri. 2018. “Evaluation of CNN, Alexnet and GoogleNet for Fruit Recognition.” Indonesian Journal of Electrical Engineering and Computer Science 12(2):468.