International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I7P228

Volume 69 Issue 7, 208-214, July, 2021
©2021 Seventh Sense Research Group®

Deep Learning-Based Approach for Old Handwritten Music Symbol Recognition

Savitri Apparo Nawade[#1], Mallikarjun Hangarge[*2], Shivanand S Rumma[**2]

[#]Research Scholar, Department of P.G. Studies and Research in Computer Science, Gulbarga University, Kalaburagi, Karnataka, India

[+]Associate Professor & Head, Department of PG Studies and Research in Computer Science, Karnatak Arts, Science, and Commerce College, Bidar, Karnataka, India

[#]Chairman, Department of P.G. Studies And Research in Computer Science, Gulbarga University, Kalaburagi, Karnataka, India

[1]savitri.warad@gmail.com, [2]mhangarge@yahoo.co.in, [3]shivanand_sr@yahoo.co.in

***Abstract** - The advanced development in information and technology created a growing interest in optical music recognition for easy storage, access, and retrieval in digital form. By using OMR, we can transcribe music sheets into a machine-readable format. This facilitates the users to play, edit or compose the music. The handwritten music symbol recognition becomes more difficult as compared to print due to various issues such as a change in shape, distortion, etc. In this paper, the performance of deep learning-based method or old handwritten music symbol recognition was investigated by applying the MobileNetV2 architecture. In this stud, two approaches are presented. The first approach deals with the pure deep learning method, and in the second approach, the softmax layer is replaced with the traditional classifiers, namely-nearest neighbor classifier, support vector machine, and random forest classifier. Encouraging results were achieved on a publically available data set of old handwritten music symbols.*

Keywords — *Convolutional Neural Networks, Handwritten Music Symbol Recognition, Deep Learning, Support Vector Machine, K-Nearest Neighbour Classifier, Random Forest Classifier.*

I. INTRODUCTION

Automatic document image processing facilitates the understanding of the document content in layout, text, graphics, and other components. In the context of graphical symbols, music symbol and their transcription to the midi format are growing areas of interest since the last few decades. A lot of work has been reported in the past focusing on offline and online recognition of music. Recognition of printed music symbols and online symbols got enough success, and some commercial products have also come into the market.

Whereas handwritten music symbols recognition still has room for research due to various complexities such as variations in shape, writing styles, noise, and degradation caused by aging for high accuracy recognition. Recently the focus of research is shifted from traditional methods to deep learning-based trends.

In this study, the performance of the deep learning-based method for old handwritten music symbol recognition was investigated. Convolutional Neural Networks were applied in this approach while utilizing the simple and lightweight mobileNetV2 architecture. The experiments are carried out to show the efficacy of combining CNN and traditional classifiers such as SVM, Random Forest, and KNN.

The remainder of the paper is presented as: In the first section, we have introduced the problem. Related work was discussed in section 2. The proposed method is discussed in Section 3. Section 4 is dedicated to Experiments and Results. Finally, concluded in section 5.

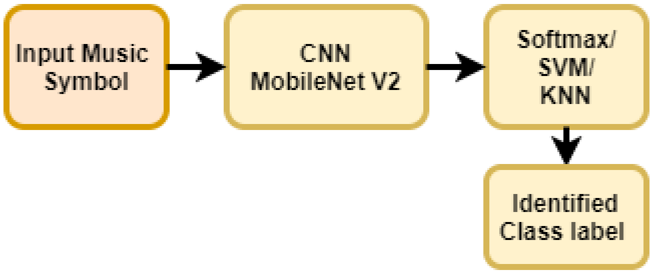

Figure 1.Schematic presentation of the proposed method

II. RELATED WORK

In [1] music symbol recognition method is presented for printed music sheet recognition; basic morphological operations were used to locate the symbol. Template matching was performed to identify similar symbols based on correlation. Graph-based method [2] called line adjacency graph (LAG) model is presented in for recognition of handwritten music symbols Dynamic Time Warping based method presented in [3] for handwritten music symbol recognition. Authors in [4] developed a method for handwritten music recognition; they used input image pixels as input to CNN for feature extraction, whereas for classification, they have applied BLSTM. In [5], use neural networks, support vector machine, KNN, and Hidden Markov Models for the recognition of music scores. Graph notations-based method is presented in [6] for recognition of music sheets with grammar and language rules.

The CNN-based approach is presented in [7] for the recognition of handwritten music symbols. VGG architecture was used as a feature extractor, and later traditional classifiers such KNN, Random Forest, and SVM are also evaluated. [8] The authors presented an approach for isolated handwritten music symbol recognition based on hybrid features extracted using discrete wavelet transform, Radon Transform, and Statistical Filters with KNN classifiers. [9] The authors developed an algorithm based on a combination of Radon and Discrete Wavelet Transform-based features and a KNN classifier with a tenfold cross-validation technique for recognition of isolated handwritten music symbols. The optical Music Object Recognition technique is given in [10] based on the deep learning method. The model takes an image of the music sheet as input and provides the symbol and notes categories. In [19], authors presented a method based on texture analysis named daisy descriptors, but because of its high dimension space, they have applied the feature selection method and achieved encouraging results. More details about music symbol recognition can be found in [21][22].

From the above paragraph, it can be seen that optical music symbol recognition is carried out on various modalities such as online, printed, and handwritten with various tools and techniques. In this paper, the lightweight deep learning-based technique for old handwritten music symbol recognition is presented and compared with previously published research.

III. PROPOSED METHOD

Old handwritten music symbol recognition is a challenging problem due to various issues in the shape and quality of the symbols due to aging. For effective recognition of these symbols, a method based on Convolutional Neural Network (CNN) is presented. The method utilizes MobileNetV2 architecture; with this, exhaustive experiments were performed to test its efficacy. For better understanding, the schematic diagram of the proposed approach is given in figure 1. The summary of our method is given below:

Step 1. Read the input of isolated music symbols with the size of 124x124x3

Step 2. Train the Convolutional Neural Network using MobileNetv2 architecture by using the principle of transfer learning.

Step 3. Replace the softmax layer with KNN, SVM, and RF and train with deep features.

Step 4. Compute

Step 5. Choose the optimal model and end the process.

***A. MobileNet*:** MobileNet was introduced in[15,16] based on depthwise separable convolutions. In MobileNet single filter is applied to each input channel, later point-wise convolutions combine the outputs depth-wise by applying 1x1 convolutions. The architecture applies separately the both filtering a combining layer. The basic mobile net applies depth-wise convolutions with the size of 3x3. MobileNet takes input and projects it into a higher dimension into a tensor with a low dimension. The MobileNetV2 comprises the 2D convolution layers, bottleneck layers, 1D convolution layers, Relu6, average pooling layer to form a network architecture. Setting of hyperparameters is given as follows: batch_size = 32, img_height = 128,img_width = 128,seed = 123, epochs = 200. The details about architecture are given in Table 1 and Fig. 1.

Table 1 : MobileNetV2 architecture[16]

Input	Operator	t	c	n	s
224x224x3	2D Convolution	-	32	1	2
112x112x32	Bottleneck	1	16	1	1
112x112x16	Bottleneck	6	24	2	1
56x56x24	Bottleneck	6	32	3	2
28x28x32	Bottleneck	6	64	4	2
14x14x64	Bottleneck	6	96	3	2
14x14x96	Bottleneck	6	160	3	1
7x7x160	Bottleneck	6	320	1	2
7x7x320	2D Convolution (1x1)	-	1280	1	1
7x7x1280	Average Pooling (7x7)	-	-	1	1
1x1x1280	2D Convolution (1x1)	-	k	-	-

***B. k- Nearest Neighbor Algorithm*:** KNN classifier is the most commonly used classification technique in pattern recognition problems due to its simplicity [14]. It is a supervised algorithm works on principle of majority voting and nearest neighbor search. Being a non-parametric technique, this algorithm does not make any assumption about the underlying data. First it stores the training data provided with labels of categories and when given unknown data point, it is going to be classified in to the category which is most similar to the new data.

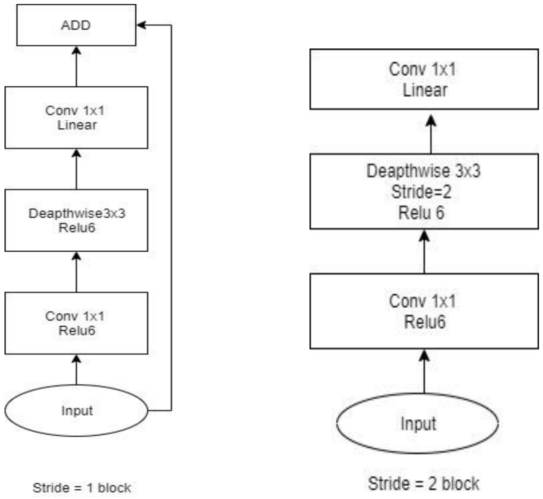

Figure 2. Convolution blocks for MobileNetV2

The similarity will be computed using suitable distance measure such Euclidean distance. In this case, there are seven categories represented by various old handwritten music symbols and used value of K=1.

Let M and N be the training and testing samples respectively denoting the feature vectors of music symbols. The Euclidean distance between M and N is defined on its components Mi& Ni is given below:

$$d(M, N) = \sqrt{\sum_{i=1}^{n}(M_i - N_i)^2} \text{ ……. ……….. Eq.[1]}$$

3.3 Support Vector Machine (SVM): SVM is one of the most popular supervised learning algorithms used in machine learning for classification problem. Fundamentally, this algorithm belongs to family of supervised learning algorithms and developed by Vapnik [17] based on statistical learning theory. SVM tries classifying the data by transforming it to hyperplane which provides the maximum margin for class separation.

For the given n feature vectors denoted as xi, a hyperplane :
$g(x) = w^T.x - b$ …….…………………….Eq. [2]
separates each feature vector into two class:
$y_i(w^T.x_i - b) \geq 1$……..…………………….Eq.[3]
for maximizing the margin of hyperplane. In this work, a simple linear SVM is employed, to investigate the separability of the deep features.

***C. Random Forest Classifier*:** Random Forest (RF) is meta learning algorithm [12] used for classification as well as regression task. It is also called a meta estimator which fits a number of decision tree classifiers on subsamples, later it uses the averaging criteria for good accuracy and to avoid the over fitting. Random forest is composed of number decision trees and the procedure of decision trees can be realized from steps given below [11]:

Step1. Let n be the training samples from dataset and number of features in feature vector will be denoted by Xi .
Step2. To build the training set each time for the tree n time replacement is made from all n samples.
Step3. The features will be fi <<xi for the decision at node of tree and each tree built in such way that it will be at its largest extent.
Step4. Each tree provides a result for particular class the classifier chooses average result for majority for the classification.
Step5. The label will be assigned based on the voting given by all trees.

IV. EXPERIMENTS AND RESULTS

A. Dataset: To evaluate our approach based on convolutional neural networks, publicly available Alicia Fornes dataset of old handwritten music symbols from [18] was used. This dataset comprises total 4098 music symbols out of which 2128 are clefs and 1970 are accidentals. These symbols are classified into seven classes namely Accidental double sharp(AS), Accidental flat(AF), Accidental Natural(AN), Accidental sharp(AS), Clef alto(CA), Clef bass(CB) and clef Treble(CT). Some samples from dataset are shown below:

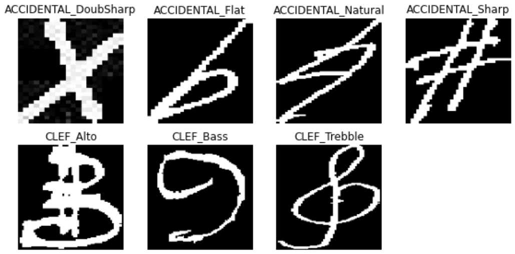

Figure 3. Some samples of music symbol from database

B. Evaluation Protocol: Aim of presented work is to evaluate the performance of CNN based deep features for recognition of old handwritten music symbols. To do this, the dataset is divided into two parts namely training and testing set with size of 3684 and 410 respectively. Further , precision (P), recall(R), F1 score and Accuracy were computed as quantitative measures to evaluate the performance of our method and the same are defined below:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F1 - Score = \frac{2 * (R * P)}{R + P}$$

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

Where TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative.

C. Results and Discussion: This work aimed to develop efficient algorithm to recognize old handwritten music symbols, to validate the performance of presented method sires of experiments with light weight CNN Model such MobileNetV2 were performed. Initially, pre-trained model with existing weights and setting, only last dense layer was modified as per the number of classes under consideration. Secondly, the model was trained from scratch with our dataset while keeping mobilenetv2 architecture as it is, and finally, softmax was replaced with traditional classifiers namely SVM, KNN and RF. Accuracy vs. loss graphically explained for the training and testing procedure in Figure2 and Figure3 for deeper understanding.

In table 2. the results are shown for recognition of old handwritten music symbols with transfer learning and training from scratch using mobileNetV2, from table 1 it can be noted that for all classes precision, recall and F1-measure are more than 97% which shows the significance of presented method. Whereas when model is trained from scratch gives enhanced results as compared to transfer learning.

From table 3 the performance of support vector machine for old handwritten music symbol recognition can be explored. Further experiments are also carried out in both scenarios such as transfer learning and training from scratch. SVM has given encouraging results with precision, recall and f-measure more than 96%. SVM performed slightly well when applied to the features extracted by training the CNN from scratch as compared to pre-trained CNN. Performance of KNN classifier was observed with deep features for handwritten music symbol recognition and presented in table 4. The value of K=1 and distance measure as Euclidean distance during the experiments. The same way as SVM classifier KNN have performed well, again model trained from scratch given high accuracy compared to pre-trained CNN.

Overall precision, recall and f-measure is noted as near to 100% with KNN. Later, the random forest classifier was evaluated with deep features for old handwritten music symbol recognition and noted superior performance as compared to KNN and SVM; the results are shown in table 4. Random Forest also performed smart when applied to CNN trained from scratch. Precision, recall and f1-measure for old handwritten music symbol recognition using Random Forest are noted more than 98% for both the scenarios.

Table 2. Old Handwritten Music Symbol Recognition Results using MobileNetV2 CNN based on Transfer Learning and Training from scratch

Class	CNN (Transfer Learning)			CNN (Trained from scratch)		
	PR	RE	F1	PR	RE	F1
AD	1.00	1.00	1.00	1.00	1.00	1.00
AF	0.95	1.00	0.98	1.00	1.00	1.00
AN	1.00	0.95	0.97	1.00	0.98	0.99
AS	1.00	1.00	1.00	1.00	1.00	1.00
CA	0.97	0.98	0.98	1.00	1.00	1.00
CB	1.00	1.00	1.00	1.00	1.00	1.00
CT	0.98	0.98	0.98	0.99	1.00	0.99

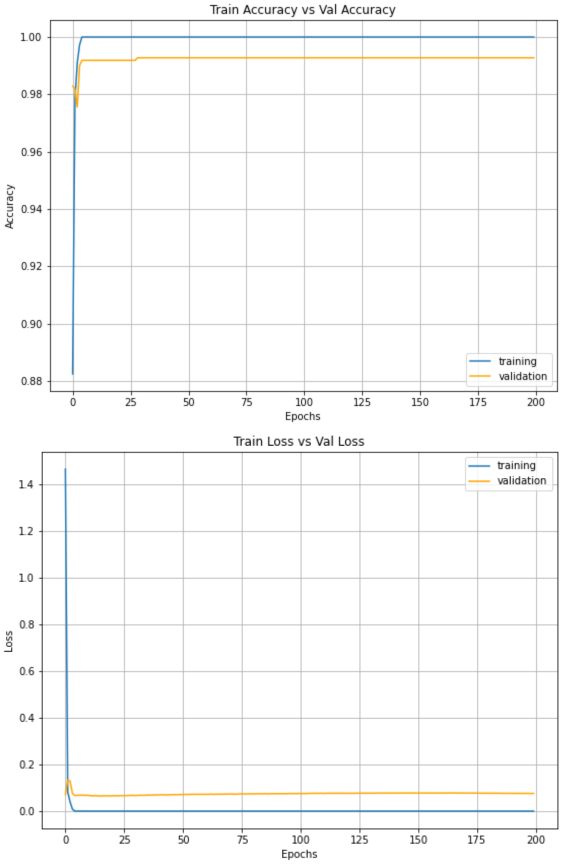

Figure 4. Plots of Accuracy & Loss on train and test set when MobileNetv2 applied for transfer learning.

Overall recognition accuracy using MobileNetv2 and deep features with SVM, KNN and, Random forest for both scenarios such as transfer learning and training from scratch is given in Table 5. When transfer learning was applied, CNN has given the accuracy of 98.53 %, CNN+SVM given the accuracy of 98.78%, CNN+KNN and CNN+SVM has given the accuracy of 98.78%, whereas CNN+RF performed superiorly. When the model trained from starch CNN has given the accuracy of 99.75% which highest accuracy as compared to all. SVM, KNN and, RF have given the accuracies 99.02%, 99.26% and, 99.51% respectively.

Table 3. Old Handwritten Music Symbol Recognition Results using MobileNetV2+SVM based on Transfer Learning and Training from scratch

Class	CNN+SVM (Transfer Learning)			CNN + SVM (trained from scratch)		
	PR	RE	F1	PR	RE	F1
AD	1.00	1.00	1.00	1.00	1.00	1.00
AF	0.98	1.00	0.99	0.95	1.00	0.98
AN	1.00	0.96	0.98	0.98	0.98	0.98
AS	1.00	1.00	1.00	1.00	1.00	1.00
CA	0.97	0.98	0.98	0.98	1.00	0.99
CB	1.00	1.00	1.00	1.00	1.00	1.00
CT	0.98	0.98	0.98	1.00	0.96	0.98

Table 3. Old Handwritten Music Symbol Recognition Results using MobileNetV2+ KNN based on Transfer Learning and Training from scratch

Class	CNN+KNN (Transfer Learning)			CNN + KNN (trained from scratch)		
	PR	RE	F1	PR	RE	F1
AD	1.00	1.00	1.00	1.00	1.00	1.00
AF	0.98	1.00	0.99	0.98	1.00	0.99
AN	1.00	0.96	0.98	1.00	0.98	0.99
AS	1.00	1.00	1.00	1.00	1.00	1.00
CA	0.97	0.98	0.98	0.98	1.00	0.99
CB	1.00	1.00	1.00	1.00	1.00	1.00
CT	0.98	0.98	0.98	0.99	0.98	0.98

Table 4. Old Handwritten Music Symbol Recognition Results using MobileNetV2+ Random Forest classifier based on Transfer Learning and Training from scratch.

Class	CNN+KNN (Transfer Learning)			CNN + KNN (trained from scratch)		
	PR	RE	F1	PR	RE	F1
AD	1.00	1.00	1.00	1.00	1.00	1.00
AF	0.98	1.00	0.99	1.00	1.00	1.00
AN	1.00	0.96	0.98	1.00	0.98	0.99
AS	1.00	1.00	1.00	1.00	1.00	1.00
CA	0.98	0.98	0.98	0.98	1.00	0.99
CB	1.00	1.00	1.00	1.00	1.00	1.00
CT	0.98	0.99	0.98	0.99	0.99	0.99

Table 5. Overall recognition accuracy in % of Old handwritten music symbols given by our method

Classifier	Transfer Learning	Trained from scratch
CNN(MobileNetV2)	98.53	99.75
CNN+SVM	98.78	99.02
CNN+KNN	98.78	99.26
CNN+RF	99.02	99.51

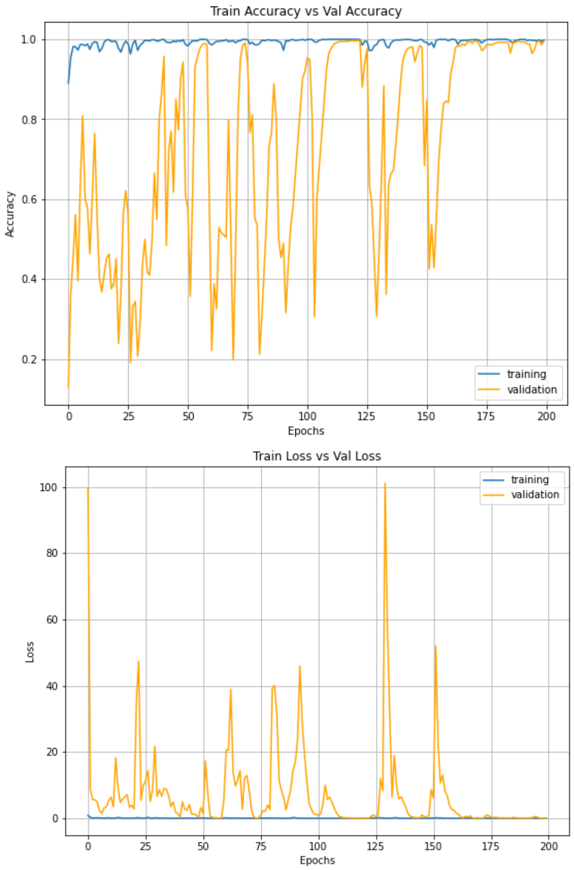

Figure 5: Plots of Accuracy & Loss on train and test set when MobileNetv2 trained from scratch

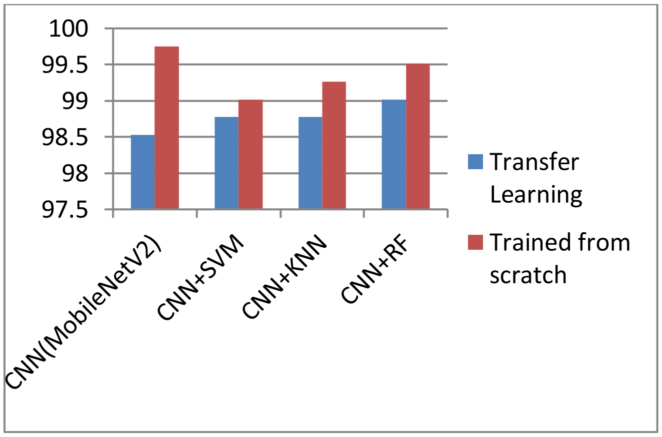

Figure 6: Graph showing comparison of various classification methods and CNN training scenarios for handwritten music symbol recognition.

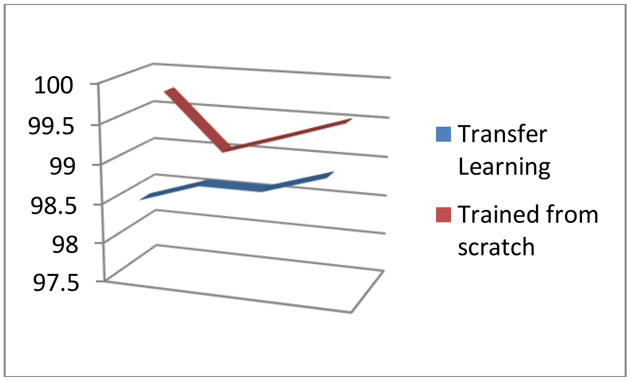

Figure 7. The line graph showing difference between the accuracies for handwritten music symbol recognition using transfer learning and training from scratch.

From the figure 4, one can observe that all classifiers have performed well with slight increase in accuracy for music symbol recognition. From figure 5 it can be noted that the CNN with mobilenetv2 architecture performed well as compared to the pre-trained model. Hence, from these experimental results, it can be learned that training from scratch though time-consuming but provides good results, and replacement of softmax layer with a traditional classifier such as Random Forest can enhance the results slightly.

Comparison of proposed method with previously published results in table 6 on the same dataset. In[3] authors presented projection profile-based features with dynamic time warping they achieved 95.81% of accuracy whereas our method has given 4% enhance results. Chain code histograms [20] given the accuracy of 98.05% whereas they suffer from noise and dependent on direction for computation. In[19] authors got 99.48 % of accuracy but it requires feature selection as an extra step due to high dimensionality of daisy descriptors. Out method given highest accuracy i.e.99.56% with CNN (mobileNetv2)+ Random Forest. In addition to this, our method does not require any sophisticated preprocessing or feature selection. It can be easily extended to printed as well as handwritten symbols with little modification such as hyperparameter tuning during the training procedure.

Table 6. Comparison with previous work

Authors	Method	Accuracy
Fornés, A et al. [3]	Projection Profile with Dynamic Time Warping	95.81%
Sukapla Chanda et al.[20]	Chain code Histogram with Modified Quadratic Classifier	98.05%
Malkar S et al. [19]	Daisy Descriptors with Grey Wolf Optimization	99.49%
Proposed Method	MobileNetV2 with Random Forest	**99.56%**

V. CONCLUSION

In this paper, the performance of MobileNetv2 for recognition of old handwritten music symbols was investigated. The in depth study of pre-trained as well as training from scratch was observed and noted that training from scratch gives the higher accuracy as compared to pre-train model. In addition to this, it is also observed that the performance of deep features with traditional classifiers for handwritten music symbol recognition was superior and understood that, performance of recognition is enhanced when Random forest is considered instead of softmax. Our method has outperformed as compared to existing methods on the same dataset and given accuracy of 99.58% .

In future, comparative study will be perfoemed with the various CNN models[23,24] for recognition of handwritten music symbols.

ACKNOWLEDGMENT

Authors thank to management of KRE Society for extending their support an dproviding necessary infrastructure to carryout this researchwork.

REFERENCES

[1] Ooi, Joyce Boon Ee, and Alan WC Tan., Music symbol recognition., (2011), 1-4.

[2] Na, I.S., Kim, S.H. Music symbol recognition by a LAG-based combination model. Multimedia Tools Applications 76(2017), 25563–25579.

[3] A. Fornés, J. Lladós, G. Sanchez., Old Handwritten Musical Symbol Classification by a Dynamic Time Warping Based Method, in Graphics Recognition: Recent Advances and New Opportunities, Lecture Notes in Computer Science, (Eds. Liu, W. and Lladós, J. and Ogier, J.M.) 5046(2008), 51-60, Springer-Verlag Berlin, Heidelberg.

[4] Baró, Arnau, Pau Riba, Jorge Calvo-Zaragoza, and Alicia Fornés., From optical music recognition to handwritten music recognition: A baseline." Pattern Recognition Letters 123(2019), 1-8.

[5] A. Rebelo , G. Capela , J.S. Cardoso ,Optical recognition of music symbols: a com- parative study, International Journal of Document Analysis and Recognition. 13(1) (2010), 19–31

[6] Pacha, Alexander, Jorge Calvo-Zaragoza, and Jan Hajic Jr., Learning Notation Graph Construction for Full-Pipeline Optical Music Recognition, In ISMIR, (2019) 75-82.

[7] Calvo-Zaragoza, Jorge, Antonio-Javier Gallego, and Antonio Pertusa., Recognition of handwritten music symbols with convolutional neural codes., In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 1(2017),691-696. IEEE.

[8] Nawade, Savitri Apparao, Mallikarjun Hangarge, Chitra Dhawale, Mamun Bin Ibne Reaz, Rajmohan Pardeshi, and Norhana Arsad., Old handwritten music symbol recognition using directional multi-resolution spatial features, In 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), (2018),1-4. IEEE.

[9] Nawade, S.A., Rumma, S., Pardeshi, R. and Hangarge, M., Old Handwritten Music Symbol Recognition Using Radon and Discrete

Wavelet Transform. In Advances in Artificial Intelligence and Data Engineering (2021),1165-1171. Springer, Singapore.

[10] Huang, Zhiqing, Xiang Jia, and Yifan Guo., State-of-the-art model for music object recognition with deep learning., Applied Sciences 9(13)(2019), 2645.

[11] Rashad, Marwa, and Noura A. Semary., Isolated printed Arabic character recognition using KNN and random forest tree classifiers., In International Conference on Advanced Machine Learning Technologies and Applications, (2014),11-17. Springer, Cham.

[12] Briman, L.: Random Forests. Machine Learning 45(1)(2001), 5–32.

[13] Saunders, Craig, Mark O. Stitson, Jason Weston, Leon Bottou, and A. Smola., Support vector machine-reference manual., (1998).

[14] Liao, Yihua, and V. Rao Vemuri., Use of k-nearest neighbor classifier for intrusion detection., Computers & security 21(5) (2002), 439-448.

[15] MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H, arXiv:1704.04861, (2017).

[16] MobileNetV2: Inverted Residuals and Linear Bottlenecks, Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. arXiv preprint. arXiv:1801.04381, (2018).

[17] Vapnik, Vladimir N., An overview of statistical learning theory, IEEE transactions on neural networks 10(5)(1999) 988-999.

[18] Isolated Old Handwritten Music Symbol Dataset,http://www.cvc.uab.es/people/afornes/datasets/datasets.html

[19] Malakar, S., Ghosh, M., Chaterjee, A. et al. Offline music symbol recognition using Daisy feature and quantum Grey wolf optimization based feature selection. Multimedia Tools Applications 79(2020),32011–32036.

[20] S. Chanda, D. Das, U. Pal and F. Kimura., Offline Hand-Written Musical Symbol Recognition, 2014 14th International Conference on Frontiers in Handwriting Recognition, (2014), 405-410.

[21] Jorge Calvo-Zaragoza, Jan Hajič Jr., and Alexander Pacha. 2020. Understanding Optical Music Recognition. ACM Comput. Surv. 53, 4, Article 77 (September 2020), 35 pages. DOI:https://doi.org/10.1145/3397499

[22] Novotný, J. and J. Pokorný., Introduction to Optical Music Recognition: Overview and Practical Challenges, DATESO (2015).

[23] S.Sunitha, Dr.S.S. Sujatha., Combined Feature Learning And CNN For Polyp Detection In Wireless Capsule Endoscopy Images" International Journal of Engineering Trends and Technology 69.6(2021):206-215.

[24] Sunil Pandey, Naresh Kumar Nagwani, Shrish Verma., Analysis and Design of High Performance Deep Learning Algorithm: Convolutional Neural Networks, International Journal of Engineering Trends and Technology 69.6(2021):216-224.