Cricket Shot Classification using Convolutional Neural Network

A DISSERTATION

submitted in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY

by

Md Ezaz Anwar	(IIT2018004)
Kalyan Kumar Dubey	(IIT2018005)
Ankit Kumar Das	(IIT2018020)
Abhishek Kumar Mishra	(IIT2018026)
Raushan Raj	(IIT2018031)

Under the supervision of

Prof. VRIJENDRA SINGH

Professor

DEPARTMENT OF INFORMATION TECHNOLOGY

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

ALLAHABAD - 211015, (UP) INDIA

DECLARATION

We hereby certify that this project report entitled "**Cricket Shot Classification using CNN-SoftMax and CNN-SVM Classifiers**" is submitted in partial fulfilment of the requirement of the Degree of Bachelor of Technology in Department of Information Technology, Indian Institute of Information Technology, Allahabad.

We have given due credit to the original authors/ sources through proper citation for all the words, ideas, diagrams, graphics, computer programs, experiments, results, websites, that are not my original contribution. We have used quotation marks to identify verbatim sentences and given credit to the original authors/sources.

Place: Allahabad, Prayagraj	Md. Ezaz Anwar (IIT2018004)
Date: 29-Nov-2021	Kalyan Kr. Dubey (IIT2018005)
	Ankit Kumar Das (IIT2018020)
	Abhishek Mishra (IIT2018026)
	Raushan Raj (IIT2018031)

CERTIFICATE

It is certified that the work contained in the project titled "**Cricket Shot Classification using CNN-SoftMax and CNN-SVM Classifiers**" has been carried out under my supervision and that this work has not been submitted elsewhere.

Signature of Supervisor

Date: 29-Nov-2021

Prof. Vrijendra Singh

Professor

ACKNOWLEDGEMENT

We would like to express our gratitude to the mercy of Almighty, that gave us enough strength to complete this project work. We are very thankful to our supervisor **Prof. Vrijendra Singh**, Professor, Department of Information technology, IIIT-Allahabad to guide us throughout the project work and always motivated us to work hard. The regular discussions and extensive comments helped us to improve the project and a direct impact can be seen on the project.

We are thankful to **Dr. Triloki Pant**, Assistant Professor, IIIT-Allahabad, for providing valuable guidance at the beginning of the project. We are also grateful to all the faculty members and research scholars who have been very supporting in providing resources to complete this project.

At the end, we wish to thank all our friends who have always been a great support to us and a source of motivation to work forward.

Md. Ezaz Anwar (IIT2018004)
Kalyan Kr. Dubey (IIT2018005)
Ankit Kumar Das (IIT2018020)
Abhishek Mishra (IIT2018026)
Raushan Raj (IIT2018031)

ABSTRACT

Artificial intelligence has emerged as the latest data processing powerhouse in this technological era. The use of machine learning and computer vision algorithms in data processing has become common since their introduction. Deep neural networks have yet to be applied to various tasks such as analysing sports data and studying the output of these models. So, for this project, we have proposed two models that we have named "Shot-Net Pro." One is a 19-layer convolutional neural network with a SoftMax output layer, while the other is a modified version of the first with L2-SVM (simulated by squared hinge loss) at the output layer. Cut Shot, Cover Drive, Straight Drive, Pull Shot, Scoop Shot, and Leg Glance are the six different types of cricket shots it classifies. Both of our proposed models achieved fairly high accuracy with low respective losses. Both the models performed better than the already available 'Shot-Net' model by a good margin.

Contents

1. Motivation

Every domain is becoming more automated as the number of Artificial Intelligence applications grows. Machine learning has allowed systems to learn the process on their own, reducing the need for human intervention. While AI has not been widely used in sports such as cricket and football, there are some places where it can be extremely beneficial. We are trying to develop a deep learning model for classification of cricket shots. The main purpose of performing such activities is to teach the system the shots the players are playing which can, in turn, help the system do commentary in the future too.

2. Problem Statements

(i) Develop a SoftMax based convolutional neural network for classification of cricket shots.
(ii) Develop a SVM based convolutional neural network for classification of cricket shots.
(iii) Compare the performance of both models using different metrics.

3. Introduction

Cricket is one of the most popular and exciting sport in many countries across the world especially in the Indian Subcontinent. Batting is a part of the game, which is the art of hitting the ball bowled by bowler in order to score runs. Batsmen have to accommodate themselves according to different conditions such as different nature of pitches, different altitudes and different weather. So, for a batsman to be successful he must, along with better batting skills, have quick reflexes, and be able to adapt to the conditions quickly. In recent years, there is a revolution of Technology in the field of Cricket also. With the growth in the application of Artificial Intelligence, Machine Learning has enabled the system to learn things on their own in order to reduce human work. So, we can apply AI techniques in Cricket, which can be used for coaching and visualisation purposes along with automated commentary in near future.

We have developed Deep Learning models to detect which kind of shot batsman has played. Our approach uses convolutional neural networks to classify different types of shots. Our model works on image dataset. We have used CNN for feature extraction (dimensionality reduction). So, we have proposed a SoftMax based CNN and a SVM based CNN model. Our models contain five convolutional layers, five max-pooling layers, six dropout layers, two dense (fully connected) layers and one flatten layer. Later we have compared these two models based on their performance and accuracy. In this project, we have 4200 images, 700 images per shot and out of these 700 images, 600 images are augmented and 100 images are original. The augmentation techniques used were Rotation by $\pm$ 30°, Shearing, Addition of salt and pepper noise, translation and shading.

Our aim is to classify 6 types of cricket shots. We have used the Shot-Net dataset [11] for our project. The developed models will detect the type of cricket shots and apart from that it will also recognize the similarities and dis-similarities between different cricket shots. The 4th section of this report elaborates the literature review and about the related work that has been done in this field. Further, 5th section deals with the proposed methodology including the background study. 6th section specifies the proposed models and 8th and 9th sections elaborate the training and testing of models. The last two sections conclude our study and tell about future scope of our study.

4. Literature Review

In recent times, convolutional neural networks incorporated with fully connected layers have achieved state-of-art performance in image classification tasks. The CNN was first proposed by LeCun et al. in 1989. Saad Albavwi et al. [1] explained all the elements and issues related to CNN, their working and stated the parameters which affect CNN efficiency.

The loss functions also hold importance in machine learning. Mohammad Norouzi et al. [7] studied different output layer regularisation strategies and loss functions on image classification tasks. They observed meaningful differences in model predictions, accuracy, calibration, and out-of-distribution robustness for networks trained with different objectives. The study identified many similarities among networks trained with different objectives. Different losses and regularizers achieved broadly similar accuracies on CIFAR-10, CIFAR-100, and ImageNet. It was found that the choice of loss function affects representations in only the last few layers of the network, which suggests that there are inherent limitations to what can be achieved by manipulation of loss. However, it was also found that different objectives lead to substantially different penultimate layer representations. They found that class separation is an important factor that distinguishes these different penultimate layer representations, and show that it is inversely related to transferability of representations to other tasks.

There are some studies [4-6] which incorporated SVM in CNN. Hend Basly et al. [6] used a deep convolutional neural network that offers the possibility of having more powerful extracted features from sequence video frames. The feature vector is then given as input to SVM to assign the label (recognise the activity). MSR Daily activity dataset is used for training and evaluation. Yichuan Tang [4] demonstrated the advantage of replacing SoftMax output layer with linear SVM. The results show that by simply replacing SoftMax with linear SVM gives significant gains on popular datasets, MNIST, CIFAR-10, and the ICML 2013 Representation Learning Workshop's face expression recognition challenge. Abien Fred M. Agarap [5] also did the same thing as [4] i.e., incorporated SVM in place of SoftMax layer.

It's well known that if the ML model has access to more data, even if data quality is lower, as long as the useful features can be extracted from the dataset. Data augmentation can be used for this job. Jason Wang and Luis Perez [9] explored the effectiveness of data augmentation techniques. Traditional data augmentation techniques are quite successful. This study experimented with GANs to generate images of different styles. They also proposed a method to allow neural nets to learn the augmentations that best improve the classifier. The model was experimented on various datasets.

Applying computer vision techniques in analysis of cricket events is becoming increasingly popular. There are several studies on activity recognition in cricket. Md. Harun-Ur-Rashid et al. [10] proposed a CNN based classification method with Interception-V3 for detection and differentiating waist height no balls with fair balls and achieved a fairly good accuracy. Rohit Kumar et al. [2] built a dataset of different outcomes in cricket (run, dot, boundary, wicket). They used VGGNet for feature extraction which were passed to LSTM to learn the sequence for one video at a time. The results weren't much convincing though. Md Nafee Al Islam et al. [8] proposed a CNN model for identifying eighteen different bowlers from seven cricket playing nations based on their bowling action using transfer learning. They also created a dataset of

8100 images for training and evaluating the model. They added a few dense layers on the top of a pretrained VGG16 model trained on ImageNet dataset after removing the output layer. Md Ferdouse Ahmed Foysal et al. [11] proposed a thirteen-layer CNN for classification of different cricket shots. The network hit the accuracy of 80%. AZM E Chaudhary et al. [3] proposed a different kind of methodology based on motion vector for classifying cricket shots. They defined eight classes of angles to detect cricket shots. The method is focussed on motion vectors which help in measuring the angle of any precise cricket shot.

5. Proposed Methodology

A. DATASET

We will use the ShotNet [11] dataset to train and evaluate our model. It contains 4200 images with 6 classes of cricket shots, namely Straight drive, Pull shot, Cut shot, Cover drive, Leg glance and Scoop shot, with 700 images in each class. The authors [11] collected 100 images per class and later augmented it using five different techniques to expand the dataset to 700 images per class.

Figure 1: Some example images

The augmentation techniques used were Rotation by $\pm 30°$, Shearing, Addition of salt and pepper noise, translation and shading.

B. BACKGROUND OF CNN

Deep learning makes use of artificial neural networks. Neural networks work just like our brains. Convolutional Neural Networks (CNNs) are one of the most efficient deep learning networks. This is an artificial neural network also known as ANN feedforward. Information circulates across networks in a "feed-forward" network. CNN functions similarly to a biological visual cortex. CNN is one of the most common image classification models. CNN outperforms all other image classification algorithm in terms of classification accuracy. We don't have to pick features in CNN, but we do in other image classification algorithms. Different types of layers are used in CNN. A filter or moving centre runs through the picture in the convolutional layer. It usually occupies a specific part of a 2D matrix (image representation), applies point multiplication, and stores the result in another matrix [1].

Convolution is represented by following mathematical formula,
(the size of filter is $(2a+1) \times (2b+1)$)

$$h(x,y) = \sum_{i=-a}^{a} \sum_{j=-b}^{b} F(i,j) I(x+i, y+j) + b \qquad (1)$$

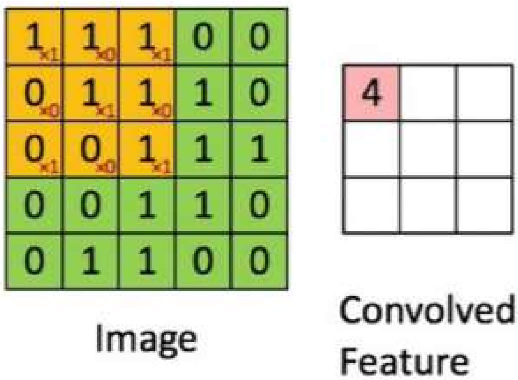

Figure 2: Convolution of an image with a filter

The dimension of the output matrix given by following equation where:
O – Output dimension
N – Input dimension
F – Window size
S – Stride
P – Padding

$$O = \lfloor \frac{N-F+2P}{S} \rfloor + 1 \quad (2)$$

The pooling layer is generally adjacent to the convolution layer. It was mainly used for memory reduction and quick calculation. Decreases the volume. Max pooling is one of the most used levels by CNN. Set up a kernel and find the maximum number from the array [1].

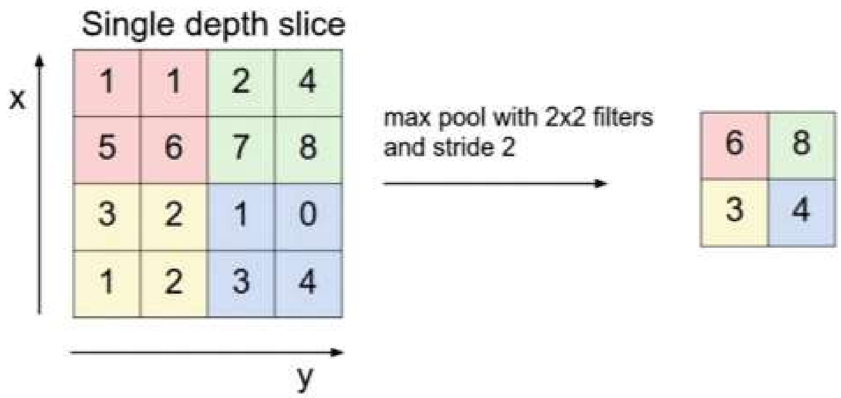

Figure 3: Max pooling with 2×2 filter with stride 2

A flatten layer converts 2D or 3D array input from previous layer converts it to 1D array. A fully connected layer is the one where all the outputs of previous layer is connected to every neuron of the layer. The output layer of the neural network shows the probability of the classes. It is calculated by the "SoftMax" function. The equation for calculating the probability is given below [1].

$$\sigma(\mathrm{X_i}) = \frac{\mathrm{e}^{\mathrm{X_i}}}{\sum_j e^{X_j}} \quad (3)$$

C. BACKGROUND OF SVM

i. Support Vector Machines

Given training data and its corresponding labels $(\boldsymbol{x_n}, y_n)$, $n = 1, \ldots, N$, $\boldsymbol{x_n} \epsilon \mathbb{R}^D$, $t_n \epsilon \{-1, +1\}$, SVMs learning consists of the following constrained optimization [4]:

$$\min_{w,\xi_n} \frac{1}{2} \boldsymbol{w^T w} + C \sum_{n=1}^{N} \xi_n \quad (4)$$

$$s.t. \boldsymbol{w^T x_n} t_n \geq 1 - \xi_n \ \forall n$$

$$\xi_n \geq 0 \ \forall n$$

ξ_n are slack variables which penalizes data points which violate the margin requirements. The corresponding unconstrained optimization problem is [4]:

$$\min_{w,} \frac{1}{2}\boldsymbol{w^T w} + C\sum_{n=1}^{N} \max\left(1 - \boldsymbol{w^T x_n} t_n, 0\right) \tag{5}$$

L1-SVM is not differentiable. So, we have the need of a differentiable SVM, which is L2-SVM. L2-SVM minimizes the squared hinge loss [4]:

$$\min_{w,} \frac{1}{2}\boldsymbol{w^T w} + C\sum_{n=1}^{N} \max(1 - \boldsymbol{w^T x_n} t_n, 0)^2 \tag{6}$$

To predict the class label of a test data $\boldsymbol{x}$:

$$\arg\max_{t}(\boldsymbol{w^T x})t \tag{7}$$

ii. Multiclass SVM

For k class problems, k linear SVMs will be trained independently, where the data from the other classes form the negative cases. The output of the k^{th} SVM is [4]

$$a_k(\boldsymbol{x}) = \boldsymbol{w^T x} \tag{8}$$

The predicted class is:

$$\arg\max_{k} a_k(\boldsymbol{x}) \tag{9}$$

SVMs simply try to find the maximum margin between data points of different classes, while SoftMax layer maximizes the likelihood or minimizes cross-entropy [4].

The feature extraction can be done using the convolutional layers which will then be used as the input to SVM. Transfer learning can also be used where some pretrained models can be used and SVM added after removing the output layer.

D. Loss Functions

1. Categorical cross entropy for CNN-SoftMax model

$$L_{CE} = -\sum_{i=1}^{n} t_i \log(p_i) \tag{10}$$

   where $t_i$is the truth label and p_i is the SoftMax probability for the i^{th} class.

2. Square hinge loss for CNN-SVM model

$$L_{SH} = \sum_{i=0}^{n} (\max(0,1 - y.y')^2) \tag{11}$$

   where ŷ the predicted value and y is either 1 or -1.

6. Timeline

25 Aug – 7 Sep: Read research papers on basics of CNN and use of CNN in the sports field.
8 Sep – 21 Sep: Read research papers on CNN involving SVM.
3 Oct – 21 Oct: Building SoftMax based model.
22 Oct – 11 Nov: Building SVM based model.
12 Nov – 21 Nov: Compare the models and write the final report.

7. Proposed Model

We came up with our own CNN model which is a slight modification of the model proposed by Md Ferdouse Ahmed Foysal et al.[11]. Our model has 19 layers. We named our model as Shotnet-Pro. Figure 4 shows the architecture of our model. The convolutional layers use ReLU as activation function. The penultimate dense layer also uses ReLU as activation function. The last layer (output layer) uses SoftMax as the activation function. The dropout layers have dropout parameter of 0.2.

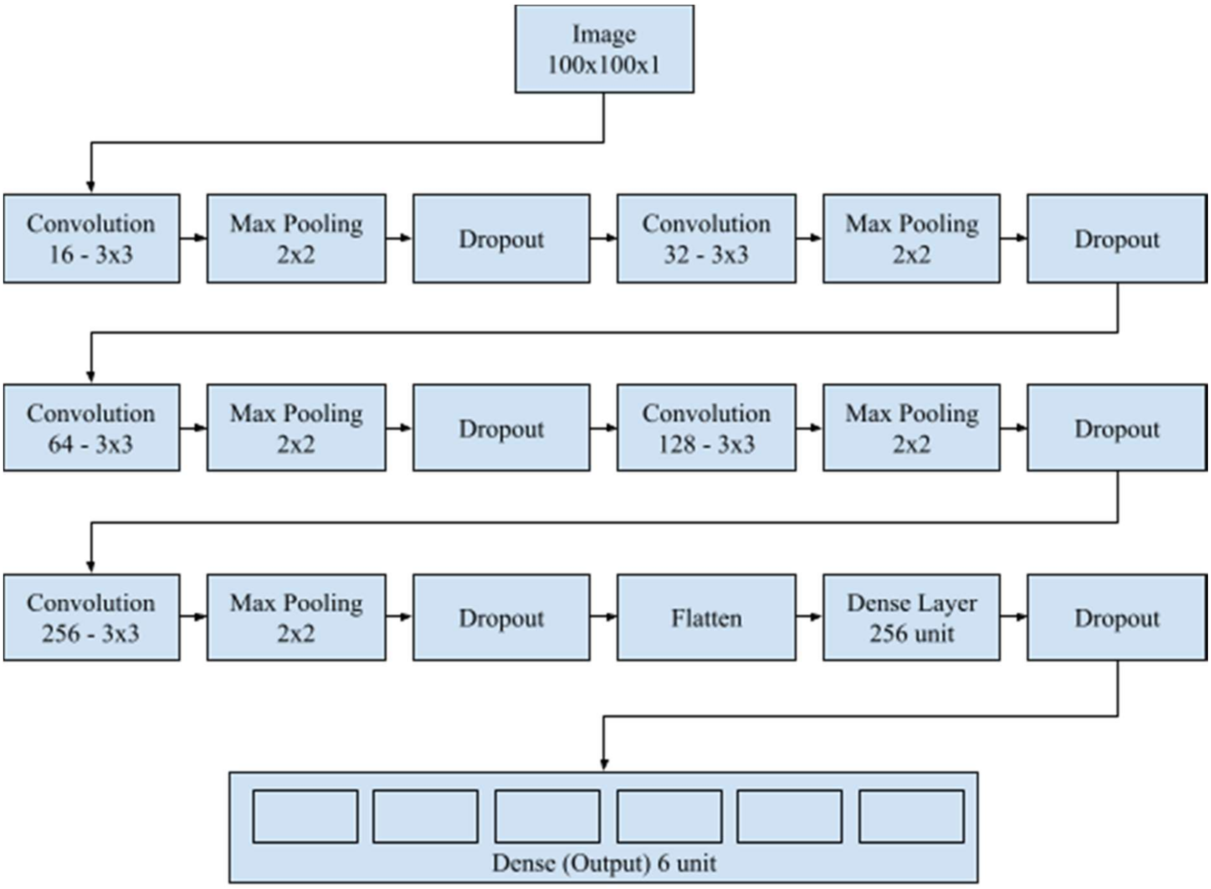

Figure 4: Architecture of our model

For the CNN-SVM model, the last layer is replaced by a L2-SVM (Squared-Hinge Loss). It is difficult to implement SVM with neural network, so we have used (to simulate SVM) 'linear' as activation function in last layer and replaced the cross-entropy loss with 'squared hinge loss'.

8. Training the Model

Adam optimizer is used to compile both the models. 80% of the dataset is used training and 20% for testing. The training dataset has 3360 images with 540 images per shot. The test dataset has 840 images with 140 images per shot. 5-Fold cross validation is used for evaluating the performance. The dataset is split into five equal parts, the model is trained on four parts, and validated on the remaining one. This is repeated five times with every part used as validation set exactly once. The network was trained for 40 epochs in every fold.

9. Experimental Results

A. TRAINING PHASE

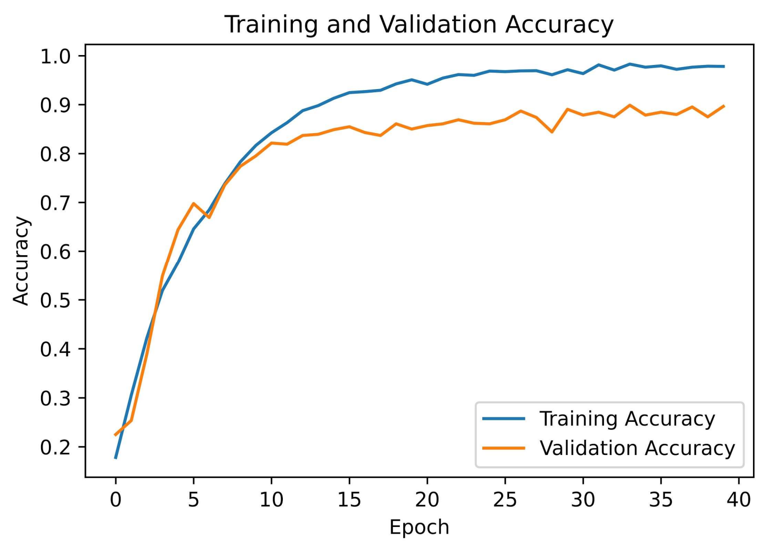

Figure 5: Training and validation accuracy of CNN-SoftMax model

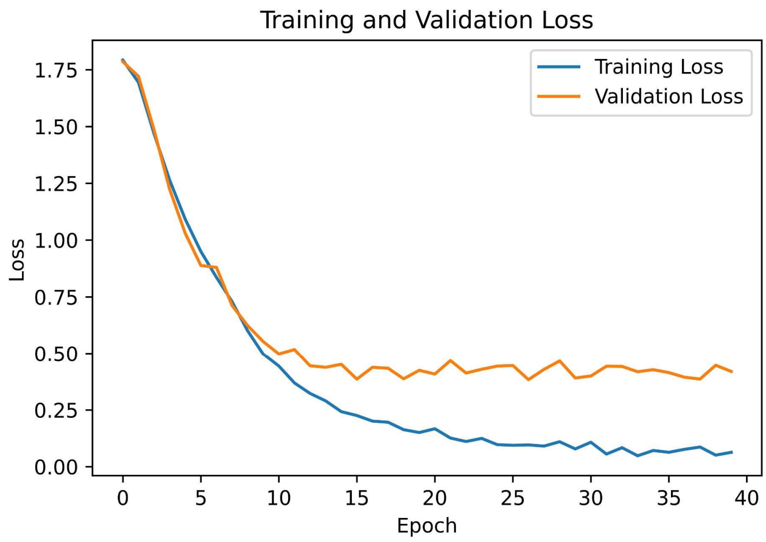

Figure 6: Training and validation loss plot of CNN-SoftMax model

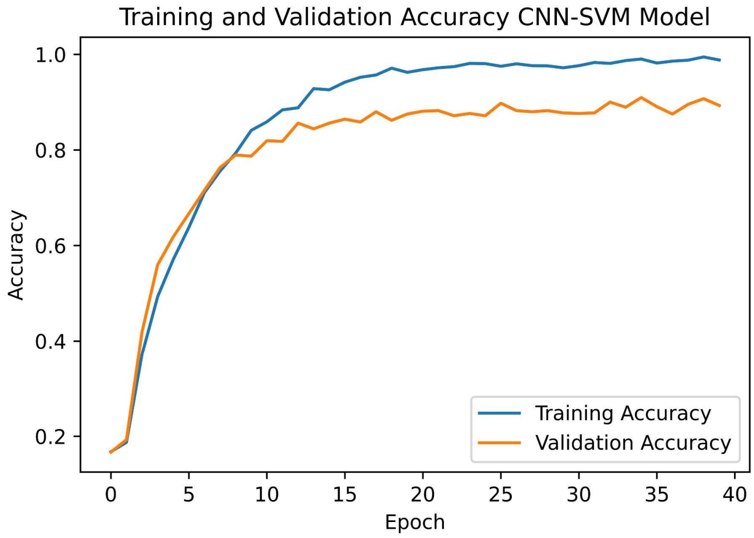

Figure 7: Training and validation accuracy of CNN-SVM model

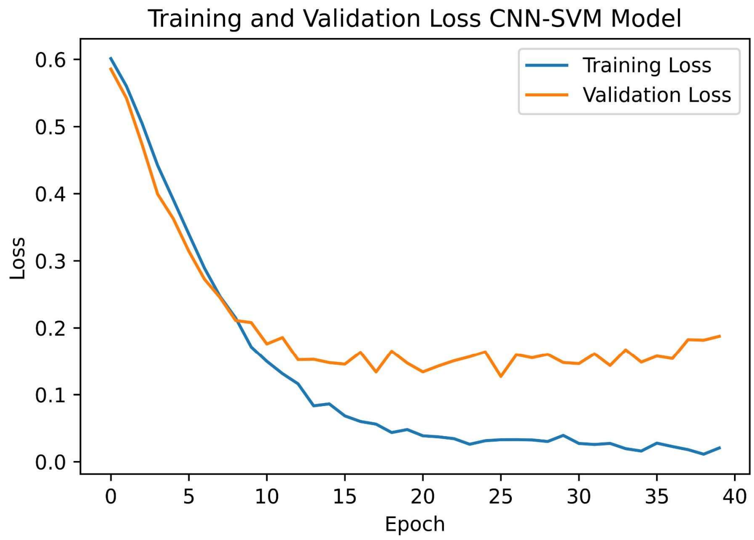

Figure 8: Training and validation loss of CNN-SVM model

It can be observed that CNN-SoftMax model converged with a good validation accuracy at low cross entropy loss. CNN-SVM model also converged with a fairly high accuracy and low squared hinge loss.

B. TEST PHASE

We evaluated the trained model on the following metrics: precision, F1-score and recall. The following classification reports were obtained for both the models along with the classification report of the Shot-Net model.

We can observe that both the models have achieved a fairly good performance with high precision, recall and F1-score. It can also be observed that both the models beat Shot-Net model in performance by a fair margin (80% accuracy in Shot-Net and around 90% in our models).

Table 1: Classification report of CNN-SoftMax Model

Shot	Precision	Recall	F1-Score
Cut Shot	0.8943	0.9071	0.9007
Cover Drive	0.9136	0.9071	0.9103
Straight Drive	0.8851	0.9357	0.9097
Pull Shot	0.9357	0.9357	0.9357
Leg Glance	0.9624	0.9142	0.9377
Scoop Shot	0.9347	0.9214	0.9280
Average	0.9210	0.9202	0.9203

Table 2: Classification report of CNN-SVM Model

Shot	Precision	Recall	F1-Score
Cut Shot	0.7975	0.9285	0.8580
Cover Drive	0.8881	0.9071	0.8975
Straight Drive	0.9612	0.8857	0.9219
Pull Shot	0.9275	0.9142	0.9208
Leg Glance	0.9603	0.8642	0.9097
Scoop Shot	0.9078	0.9142	0.9110
Average	0.9070	0.9023	0.9032

Table 3: Classification report of Shot-Net Model [11]

Shot	Precision	Recall	F1-Score
Cut Shot	0.69	0.78	0.72
Cover Drive	0.74	0.78	0.76
Straight Drive	0.78	0.83	0.81
Pull Shot	0.89	0.77	0.83
Leg Glance	0.79	0.88	0.83
Scoop Shot	0.88	0.72	0.79
Average	0.80	0.79	0.79

The confusion matrices are also used to show the performance of our model. It is also a good metric to asses class wise performance and as a whole. The following figures show the confusion matrices, normalised and non-normalised for both the models.

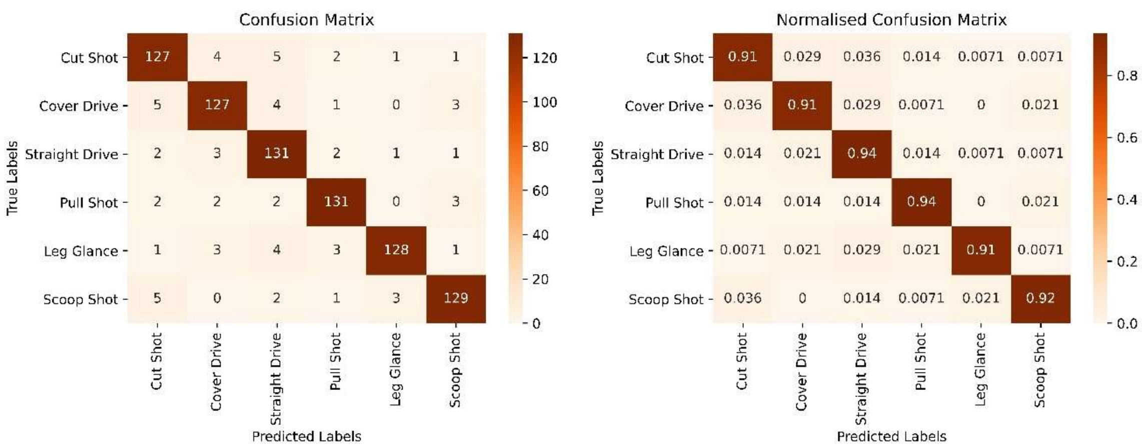

Figure 9: CNN-SoftMax Model (a) confusion matrix (L) (b) normalised confusion matrix (R)

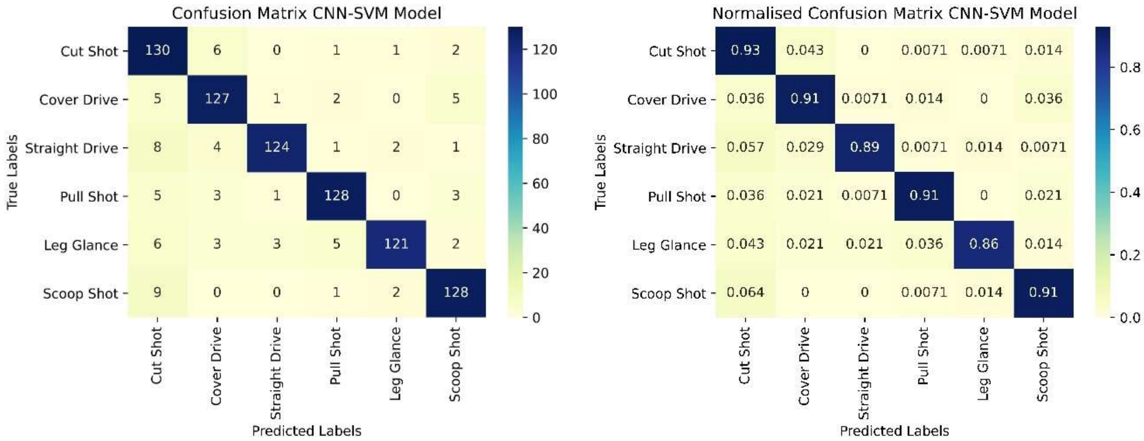

Figure 10: CNN-SVM Model (a) confusion matrix (L) (b) normalised confusion matrix (R)

C. K-FOLD CROSS VALIDATION RESULTS

k-fold cross validation is a rigorous method to evaluate the performance of the model. The training data is partitioned into *k* parts. Out of *k* parts, one part is used as validation data and remaining *k*-1 parts are used for training. The cross-validation process is then repeated *k* times, with each of *k* parts are used exactly once as validation data. Table 5 shows the results of *k*-fold cross validation of both the models.

Table 4: 5-Fold cross validation results

	Accuracy	
Fold	**CNN-SoftMax Model**	**CNN-SVM Model**
1	0.8917	0.8833
2	0.8988	0.9000
3	0.8917	0.8929
4	0.9036	0.8976
5	0.8857	0.9143
Avg.	0.8943	0.8976

Both the models performed almost same in terms of accuracy with CNN-SVM Model performing just slightly better. But is the difference is not very significant. So, it is not easy to say which model is better. Both the models, on the other hand, performed hands down better than the Shot-Net model with accuracies hovering around 90% against an 80% accuracy for Shot-Net model.

10. Conclusion

In this study, we have provided two models for classification of cricket shots using CNN-SoftMax model and CNN-SVM model (Shot-Net Pro) and compared the results achieved from both of them. We used five convolution layers, five max pooling layers, six dropout layers, one flatten layer and two dense layers. The final achieved result we've found from both the two is quite promising. CNN-SVM performed slightly better than CNN-SoftMax. Both the models outperformed the Shot-Net model by a fair margin of 10%. We hope that these methods will be developed as real applications in the future for the well being of the game of cricket.

11. Future Work

With the use of our proposed models, we can classify different types of cricket strokes. We used convolutional neural networks to model our classifier. In future the network can be upgraded to create a better neural network for better accuracy. We can upgrade this network to perform image classification on RGB images. The model can also be modified to perform classification on video dataset so that commentary can be automated in future. We can perform classification based on 3D depth images using deep learning. We can use different types of algorithms choosing which one is efficient.

12. References

[1] Saad ALBAWI, Tareq Abed MOHAMMED, Saad AL-ZAWI "Understanding of a Convolutional Neural Network" International Conference on Engineering and Technology 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.

[2] Rohit Kumar, D.Santhadevi, Dr. B. Janet "Outcome Classification in Cricket Using Deep Learning" IEEE International Conference on Cloud Computing in Emerging Markets (CCEM) 2019, pp. 55-58, doi: 10.1109/CCEM48484.2019.00012.

[3] D Karmaker, AZM E Chowdhury, M S U Miah, M A Imran and M H Rahman "Cricket Shot Classification Using Motion Vector" Second International Conference on Computing Technology and Information Management (ICCTIM) 2015, pp. 125-129, doi: 10.1109/ICCTIM.2015.7224605.

[4] Yichuan Tang "Deep Learning using Linear Support Vector Machines", International Conference on Machine Learning 2013, pp. 1-6.

[5] Abien Fred M. Agarap "An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification" arXiv:1712.03541v2 Dec 2017, pp. 5–8, https://arxiv.org/abs/1712.03541.

[6] Hend Basly, Wael Ouarda, Fatma Ezahra Sayadi, Bouraoui Ouni, and Adel M. Alimi "CNN-SVM Learning Approach Based Human Activity Recognition" International Conference on Image and Signal Processing ICISP 2020, pp 271-281, https://doi.org/10.1007/978-3-030-51935-3_29.

[7] Simon Kornblith, Honglak Lee, Ting Chen, Mohammad Norouzi "What's in a loss function for image classification?" arXiv:2010.16402v1 Oct 2020, pp. 1-20. https://arxiv.org/abs/2010.16402v1

[8] Md Nafee Al Islam, Tanzil Bin Hassan, Siamul Karim Khan "A CNN-based approach to classify cricket bowlers based on their bowling actions" IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON) 2019, pp. 130-134, doi: 10.1109/SPICSCON48833.2019.9065090.

[9] Jason Wang, Luis Perez "The Effectiveness of Data Augmentation in Image Classification using Deep Learning" arXiv:1712.04621v1 Dec 2017, pp. 1-8. https://arxiv.org/abs/1712.04621

[10] Md. Harun-Ur-Rashid, Shekina Khatun, Mehe Zabin Trisha, Nafis Neehal, Md. Zahid Hasan "Crick-net: A Convolutional Neural Network based Classification Approach for Detecting Waist High No Balls in Cricket" arXiv:1805.05974v2 May 2018, pp. 1-10. https://arxiv.org/abs/1805.05974

[11] Md. Ferdouse Ahmed Foysal, Mohammad Shakirul Islam, Asif Karim, and Nafis Neehal "Shot-Net: A Convolutional Neural Network for Classifying Different Cricket Shots" International Conference on Recent Trends in Image Processing and Pattern Recognition RTIP2R 2018, pp. 111–120, doi: https://doi.org/10.1007/978-981-13-9181-1_10.