



OPEN

## Automated recognition of the cricket batting backlift technique in video footage using deep learning architectures

Tevin Moodley<sup>1</sup>, Dustin van der Haar<sup>1✉</sup> & Habib Noorbhai<sup>2</sup>

There have been limited studies demonstrating the validation of batting techniques in cricket using machine learning. This study demonstrates how the batting backlift technique in cricket can be automatically recognised in video footage and compares the performance of popular deep learning architectures, namely, AlexNet, Inception V3, Inception Resnet V2, and Xception. A dataset is created containing the lateral and straight backlift classes and assessed according to standard machine learning metrics. The architectures had similar performance with one false positive in the lateral class and a precision score of 100%, along with a recall score of 95%, and an f1-score of 98% for each architecture, respectively. The AlexNet architecture performed the worst out of the four architectures as it incorrectly classified four images that were supposed to be in the straight class. The architecture that is best suited for the problem domain is the Xception architecture with a loss of 0.03 and 98.2.5% accuracy, thus demonstrating its capability in differentiating between lateral and straight backlifts. This study provides a way forward in the automatic recognition of player patterns and motion capture, making it less challenging for sports scientists, biomechanists and video analysts working in the field.

Cricket batting has evolved considerably in recent years, with an added emphasis on the shorter formats of the game<sup>1</sup>. There has been a growing need to maximise performance and success at the highest level, as well as to understand particular playing patterns through sophisticated analysis and machine learning<sup>2</sup>.

The increased use of technology, combined with science and medicine, has been labelled as a game-changer within the sporting domain, with an emphasis on analysis<sup>3</sup>. The technological growth has seen significant breakthroughs within sports video content analysis, particularly through the advances in artificial intelligence, deep learning and multimedia technologies. Sports video analysis is domain dependant with unique challenges, which identifies several areas of research that require further investigation<sup>4,5</sup>. Several studies have exploited the use of technology where one particular study attempts to measure the vertical jump performance reliably using an iPhone application called *My Jump*<sup>6</sup>. Several other mobile applications have been developed to assist in player performance and provide feedback from a biomechanics perspective<sup>7-9</sup>. In the cricketing domain, mobile applications have been developed to analyse team performance, player injury, and match prediction<sup>3,10-12</sup>. While these applications have made several improvements within the cricketing domain, there is a lack of research dedicated toward the enhancement and improvement of cricket batting<sup>13</sup>.

The cricket batting technique is intricate that involves a series of complex gestures needed to perform a stroke, one of these gestures performed by the batsman is referred to as the batting backlift technique (BBT)<sup>14</sup>. Previous research has indicated that the BBT can be seen as a contributing factor to successful batsmanship<sup>13,15,16</sup>. There are two backlifts investigated in this study, namely the *lateral* batting backlift technique (LBBT), and the *straight* batting backlift technique (SBBT). The LBBT is a technique present where the toe and face of the bat are lifted laterally in the direction of second slip. The SBBT is represented whenever the toe and face of the bat are pointed toward the stumps and ground<sup>13</sup>.

**Related works.** Conventionally, many video analysts would record footage using video cameras and link these to their parent analysis software to identify player performances, patterns, as well as kinetic and kinematic

<sup>1</sup>Academy of Computer Science and Software Engineering, University of Johannesburg, Cnr University Road and Kingsway Avenue, Auckland Park Johannesburg, 2092 Gauteng, South Africa. <sup>2</sup>Biomedical Engineering and Healthcare Technology (BEAHT) Research Centre, Faculty of Health Sciences, University of Johannesburg, Cnr Siemert Beit Street, Doornfontein City Two, 2028 Gauteng, South Africa. ✉email: [dvanderhaar@uj.ac.za](mailto:dvanderhaar@uj.ac.za)

analyses. However, at times, such processes can be a tedious task. In the age of the fourth industrial revolution where automation and deep learning can be utilised to enhance real-time analysis and various identifications of match play, this provided impetus for further studies to document such validations.

There have been limited studies conducted demonstrating the validation of batting techniques in cricket either through mobile applications, platforms, machine learning or artificial intelligence. There has been an increase in applying computer vision techniques within the context of cricket. One example<sup>17</sup> propose a cricket stroke recognition model, which demonstrates how various cricket strokes such as block cut, drive and glance in cricket can be automatically recognised in video footage using different traditional and deep learning architectures.

A study conducted by<sup>18</sup> propose a model that uses a deep CNN to recognise cricket batting shots. The proposed method is based on 2D convolution followed by a recurrent neural network for processing sequence of video frames and a 3D convolution network for capturing spatial and temporal features simultaneously<sup>18</sup>. The dataset used comprised of 800 batting shot clips consisting of the drive, pull, hook, cut, sweep and flick strokes. The model is able to recognise the different strokes being performed with 90% accuracy. The high model accuracy noted the implications of modern deep learning in applications for detecting various cricket activities and for decision making.

A cricket shot detection model is presented by<sup>19</sup>, which proposes a novel scheme to recognise various cricketing strokes. The proposed model uses a deep convolutional neural network that relies on saliency and optical flow to highlight static and dynamic cues. The study proposes an entirely new dataset consisting of 429 video clips of the different types of drive strokes performed by a batsman<sup>19</sup>. The proposed framework achieves an accuracy of 97.69% for a left-handed batsman and 93.18% for a right-handed batsman. The authors noted that future work would focus on extending the incorporation of native features for every defined stroke.

In each of the studies highlighted, there has been an attempt to recognise different cricketing strokes using deep learning methods on entirely different datasets. These studies highlight the success of deep learning within the problem domain and further motivate the use of deep learning within the cricketing context. While deep learning has successfully recognised different cricket strokes, there are limited studies and no datasets that address the cricketing backlift, which is a key aspect of the batting technique. By applying deep learning methods to the backlift recognition task, this research is able to achieve automated recognition of the cricket batting backlift technique in video footage, which is novel within the cricketing environment. Additionally, the dataset created in this study can be used upon request, which will benchmark future works.

The contributions of the article can be noted as follows; producing a model to achieve an end to end backlift recognition, the use of transfer learning and how it improves performance in the context even in the presence of little data, the Xception architecture is the best performing architecture for the backlift recognition task, which in itself is a largely unexplored area, and finally, the dataset produced the firsts of its kind, which will allow future works to be implemented using a baseline dataset.

The outline of this study begins with section “**Methods**”, which unpacks the terminology, the dataset used and the different types of architectures implemented in this study. Section “**Results**” discusses the results obtained in the study, which is further unpacked in section “**Discussion**”. Finally, the contributions and future works are outlined in section “**Conclusion**”.

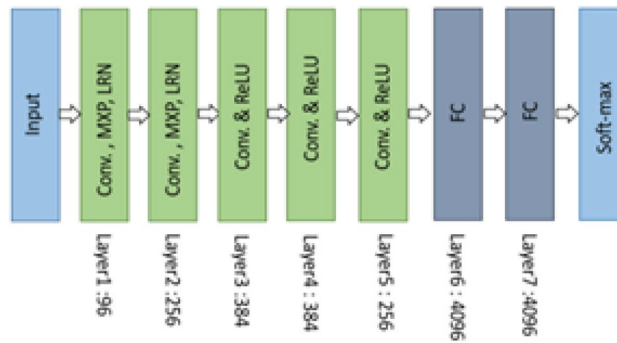
## Methods

Traditionally, types of backlift (lateral or straight) are categorised manually using an expert and analysing specific pose and alignment attributes. In this study, we attempt to automate this process by following deep learning-based methods instead of using a hand-crafted feature engineered based approach. It uses representational learning that implicitly constructs its own features to differentiate between the types of backlift by training on known cases of each backlift class.

We start by first creating a dataset of images that contains two types of backlifts, along with annotations specifying which type of backlift each image depicts. We then select, create and train deep learning architectures using the created dataset and benchmark against common machine learning metrics to determine its efficacy.

**Dataset creation.** There are specific requirements that must be considered during the construction of the dataset. The process began through a comprehensive YouTube search of First-Class International Cricket Test Match highlights, where the match's environment has fewer variations to consider. The investigation was to select various batsmen who performed the lateral and straight backlifts, ten batsmen that exhibited the straight backlift and ten batsmen that demonstrated the lateral backlift. The batsman selected for the straight backlift; Babar Azam, Themba Bavuma, Rahul Dravid, JP Duminy, Dean Elgar, Mahela Jayawardene, Ajinkya Rahane, Joe Root, Rory Burns, Ben Stokes, and David Warner. The batsmen selected for the lateral backlift; AB de Villiers, Hashim Amla, Quintin de Kock, Faf du Plessis, Kevin Pieterse, Kumar Sangakkara, Brian Lara, Ricky Ponting, Steve Smith, and Virat Kohli. Using the Labelbox editor, each object within the scene is labelled, allowing for easier isolation and extraction of the batsman in each frame. The frame used for constructing the dataset was when the bowler is about to release the ball towards the batsman. The frame is identified as the ideal time period for the position of the batsman at the instant of delivery<sup>20</sup>. Using an 80:20 data split, the training class had 160 images, and the testing class had 40 images, resulting in a total of 200 images, which will serve as a baseline to draw comparisons of the proposed architectures. The image aspect ratio chosen through testing and validation is 128 × 128, which is chosen to avoid distorting the original image.

**Model for implementation.** The models proposed for this research paper is the AlexNet, Inception V3, Inception Resnet V2, and Xception architecture<sup>21</sup>. As of late, deep learning has gathered tremendous success in various domains<sup>22,23</sup>. The AlexNet architecture, which is relatively older, belongs to the deep Convolutional



**Figure 1.** The AlexNet architecture depicting the different layers used for classification<sup>22</sup>.

Neural Network (CNN) structure proposed by Krizhevsky and subsequently won the ImageNet object recognition challenge in 2012<sup>24</sup>.

AlexNet managed to achieve recognition accuracy that was better than most traditional machine learning approaches of the time<sup>22</sup>. The significant breakthrough in machine learning and computer vision-related tasks has seen the AlexNet architecture widely used across various domains<sup>25</sup>. Figure 1, depicts the structure of the AlexNet architecture. Local Response Normalisation (LRN) is performed at the first layer using 96 receptive filter<sup>22</sup>, where LRN is responsible for connecting the layers using an organised spatial pattern. Max pooling allows for dimensionality reduction in which assumptions are subsequently derived concerning the features contained within the sub-regions and is performed using  $3 \times 3$  filters with a size of 2. Each layer has a number of kernels with a specific size. The second ( $5 \times 5 \times 48$ ) layer has 256, the third ( $3 \times 3 \times 256$ ) and fourth ( $3 \times 3 \times 192$ ) have 384, and the fifth ( $3 \times 3 \times 192$ ) has 256 kernels<sup>23</sup>. Layer six and seven is the fully-connected layers made up of 4096 neurons each. Finally, the softmax fully-connected layer represents the number of labelled classes.

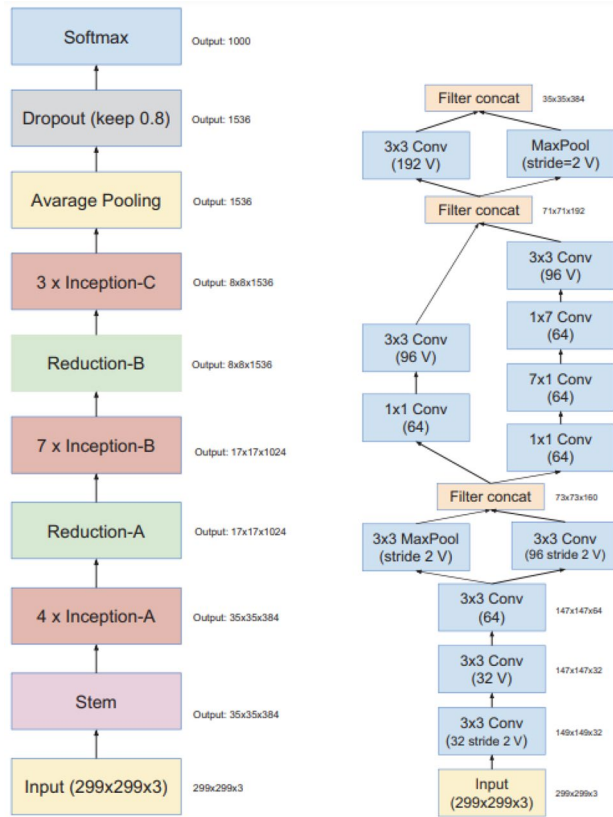
In 2014, Szegedy et al., introduced a newer network known as GoogleLeNet, otherwise known as Inception V1. The Inception V1 architecture was then refined over the years and subsequently referred to as Inception V2, Inception V3, and Inception Resnet<sup>26</sup>. The Inception V3 model was proposed that consisted of 48 layers. Unlike the AlexNet architecture, the Inception V3 architecture performs some of its calculations simultaneously<sup>27</sup>. Szegedy et al. describe the Inception V3 architecture<sup>28</sup>. The traditional  $7 \times 7$  convolution is factorised into three  $3 \times 3$  convolutions. The inception part of the network has three traditional modules at the  $35 \times 35$  convolution with 288 filters each, which is reduced to a  $17 \times 17$  grid with 7968 filters<sup>28</sup>. The inception modules are then followed by five instances of factorised inception modules, which is reduced to an  $8 \times 8 \times 1280$  grid using the grid reduction technique.

Using the Inception architecture and residual connections, the Inception Resnet V2 architecture was formed<sup>29</sup>. Inception Resnet V2 is a convolutional neural network that implements concatenation in each multi-branch architecture. Residual models are well known for training very deep architectures<sup>29</sup>. Figure 2 represents the overall schema and the detailed composition of the Inception Resnet V2 architecture, where the inception blocks can be seen<sup>30</sup>. Using the filter expansion layer and residual modules after each inception block, the dimensionality of the filter bank is scaled up to compensate for the dimensionality reduction induced by the inception block<sup>30</sup>.

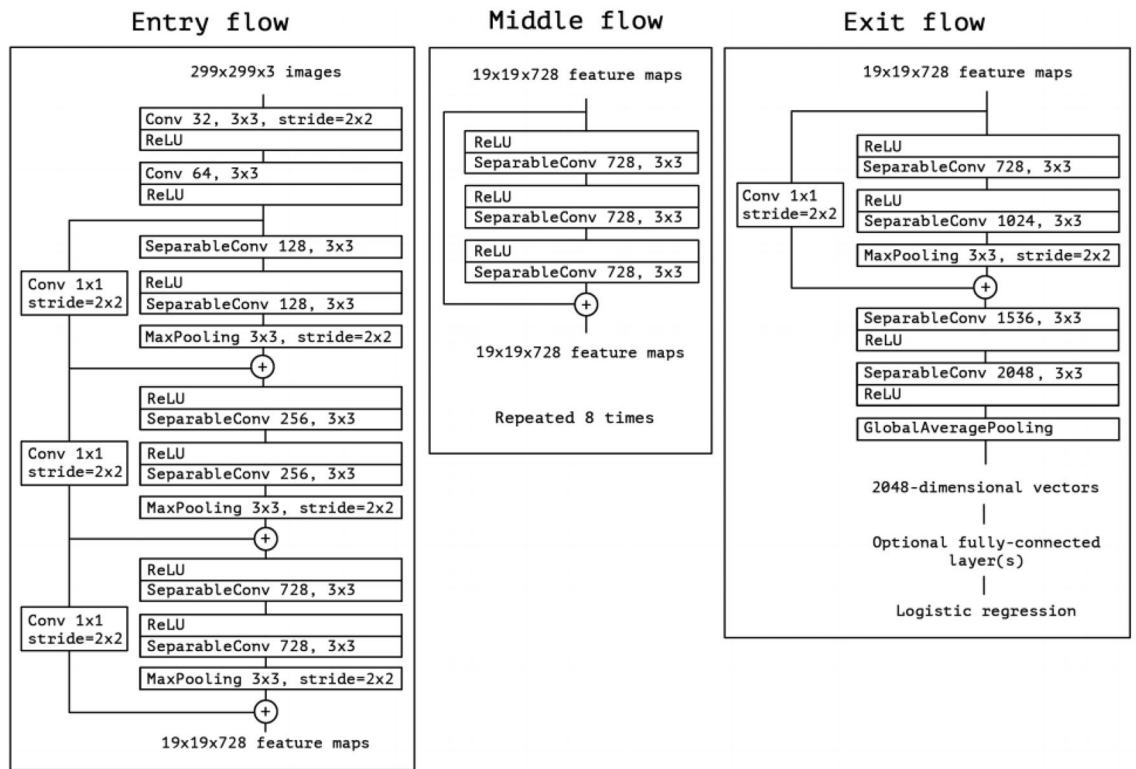
According to Chollet, the architecture dubbed as Xception is said to slightly outperform the Inception V3 architecture on the ImageNet dataset<sup>26</sup>. The Xception architecture is made up of 36 convolution layers and is able to decouple the mapping between cross-channel correlations and spatial correlations. Each layer is structured into 14 modules, where each module has a linear residual connection around it<sup>26</sup>. Figure 3 represents the Xception architecture that undergoes an entry flow, middle flow, and exit flow.

Transfer learning allows for the transfer of knowledge from one task to another, which has been used in highly specialised disciplines, where the availability of large scale quality data proves challenging<sup>31</sup>. Generally, a neural network requires a large dataset to train from scratch, unfortunately, these large datasets are not always available, which is why transfer learning is beneficial in this study, where the dataset is noted for being smaller in size. Transfer learning draws a starting point in which a good machine learning model can be built with comparatively little training data since the model is already pre-trained. Pre-trained models are seen as the efficient optimisation procedure, which also supports the improvement of classification problems<sup>31</sup>. In order to justify the architecture of choice, Table 1 is added that compares the top-1 and top-5 accuracy of the chosen architectures in the study. Using the top-1 and top-5 accuracy scores, the Inception V3, Inception Resnet V2, and Xception architectures have been compared on the ImageNet dataset, highlighting a performance comparison for each architecture. Top-1 accuracy refers to the conventional accuracy, where the model's result must match the expected output exactly. Top-5 specifies any of the top 5 probabilities which must match the expected output. Table 1 illustrates a comparison of the various architectures on the ImageNet dataset. As seen in Table 1, the Xception architecture yields the best performance when applied to ImageNet with a top-1 accuracy of 80.3%, which serves as further evidence as to why the Xception architecture is the most suitable architecture for this study, further motivations will be unpacked in the results section.

While transfer learning offers many advantages related to time, computational complexity, accuracy, and smaller datasets, there are some limitations to its use<sup>32</sup>. One of the major challenges related to transfer learning is to produce positive transfer between tasks while avoiding negative transfer, which is a decrease in performance.



**Figure 2.** The left side represents the overall schema for the pure Inception ResNet V2, where the right side illustrates the detailed composition of the stem<sup>30</sup>.



**Figure 3.** A figure representing the Xception architecture<sup>26</sup>.

Network	Top-1 Accuracy	Top-5 Accuracy
Alexnet <sup>22</sup>	0.6330	0.8460
Inception V3 <sup>28</sup>	0.790	0.945
Inception Resnet V2 <sup>30</sup>	0.779	0.937
Xception <sup>26</sup>	0.803	0.953

**Table 1.** The Inception V3, Inception Resnet V2, and Xception architectures pre-trained and bench-marked on the ImageNet dataset to illustrate each architectures performance on a generalised dataset and to justify the selection of architectures in the study.

Architecture	Class	Lateral	Straight	Accuracy (%)	Loss (%)
AlexNet	Lateral	19	1	82.45	34.17
	Straight	4	16		
Inception V3	Lateral	19	1	93.75	0.13
	Straight	0	20		
Inception Resnet V2	Lateral	19	1	96.1	0.12
	Straight	0	20		
Xception	Lateral	19	1	98.2	0.03
	Straight	0	20		

**Table 2.** The confusion matrix representing each of the architectures across ten runs, where the false and true positive predictions for each class are represented.

Negative transfer occurs as there is no specific standard that defines the manner in which tasks are related from both domain and algorithmic perspectives, which makes it challenging to find solutions<sup>32</sup>. A more well-studied limitation surrounding transfer learning is overfitting that is apparent when a new model learns details and noises from training data that negatively impacts its output<sup>32</sup>. Fortunately, these limitations were not noted in this study. Overfitting was not encountered, and therefore, there was no need to reduce the network's capacity, apply regularisation or add dropout layers.

During implementation, the following parameters were altered to ensure the Inception V3, Inception Resnet V2, and Xception architectures were fairly compared against the AlexNet architecture. Each architecture made use of transfer learning. For each architecture, the fully connected layer at the top of the network was set to *False*, which was done as the input shape of the images were changed to  $128 \times 128$  pixels. A global average pooling layer was added as a substitute fully connected layer. A dense output layer with activation softmax was added to match the number of classes in the study. The image weights were set to none. Again this is to ensure that the architectures are compared fairly to AlexNet. Finally, the input tensor is set to none, and the input size is set to  $128 \times 128$  pixels. No other parameters were altered.

Data augmentation techniques were applied to each architecture: shear range of 0.2, image rescaling of 1./255, horizontal flip set to true, and a zoom range of 0.2. Various parameters were selected through testing and validation; the activation parameter used is the rectified linear unit (ReLU), which manages to converge faster and more reliably in this study<sup>33</sup>. The softmax activation function is used at the output layer in order to predict a multinomial probability distribution. Using the Adam optimiser, the model manages the sparse gradients of noisy data. Using the various architectures defined in this study offers a novel approach to the problem domain, where metrics can determine the success of applying deep learning methods within the cricketing field. The results section will further unpack the findings of the proposed model and form a comparison on the different architectures to identify different areas of focus in future works. The source code may be found on the github repository, and the dataset may be used upon request.

**Data analysis.** Each architecture's performance is evaluated using the accuracy, confusion matrix, precision, recall, and f1-score metrics. To understand each metric the True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) values are defined. TP occurs when the proposed model manages to correctly predict positive observations, and TN occurs when the model correctly predicts negative observations. FP is apparent when the model incorrectly predicts positive observations, and the FN occurs when the model incorrectly predicts false observations. The model accuracy will formulate a comprehensive understanding of the architecture's performance<sup>34</sup>. In order to testify to the effectiveness and efficiency of the results, multiple runs are carried out for each architecture. The results obtained in Table 2 is a result of running the data against each architecture ten times and computing the averages for each metric. Completing multiple runs ensures that the data is behaving correctly and the the results obtained are consistent verifiable across the respective architectures used in this study.



Architecture	Class	Precision (%)	Recall (%)	F1-Score (%)
AlexNet	Lateral	83	95	88
	Straight	94	80	86
Inception V3	Lateral	100	95	97
	Straight	95	100	98
Inception Resnet V2	Lateral	100	95	97
	Straight	95	100	98
Xception	Lateral	100	95	97
	Straight	95	100	98

**Table 3.** The average precision, recall, and f1-scores across ten runs for the lateral and straight backlifts for each architecture.

We can expect the model to successfully draw distinctions between the respective classes from the proposed architecture, thereby automatically recognising different backlifts in video footage, supported in the results section.

## Results

The metrics illustrated in Table 2 represents the confusion matrix for each architecture. The number of TP, FP, TN, and FN are highlighted. The AlexNet architecture struggles to correctly predict for the straight class, as it incorrectly predicts four images as a lateral backlift, where in fact, the images represent a straight backlift. The remaining architectures, Inception V3, Inception Resnet V2, and Xception, all have the same amount of misclassifications, each incorrectly predicting a single image as a straight backlift that represents a lateral backlift. The architectures accuracy and loss scores are illustrated in Table 2. The noteworthy observation is the loss score of the Xception architecture of 0.03%, which will be discussed in section 4.

The metrics are shown in Table 3, which illustrates each architecture's precision, recall, and f1-score for the respective classes. For each class (with a support of 20), the number of false positives is computed as shown in Table 2. Similarly to Table 2 the Inception V3, Inception Resnet V2, and Xception architectures all exhibit the same scores, the lateral class had a precision score of 100%, a recall score of 95%, and an f1-score of 97%. For the straight class, the three architectures had precision scores of 95%, recall scores of 100%, and f1-scores of 98%. The AlexNet architecture scores differed. The lateral class had a precision score of 83%, a recall score of 95%, and an f1-score of 88%. Finally, the straight class had a precision score of 94%, a recall score of 80%, and an f1-score of 86% as seen in Table 3.

## Discussion

This research paper looks to recognise different cricket backlifts using the AlexNet, Inception V3, Inception Resnet V2, and Xception architectures, which is successfully achieved as demonstrated by the results. The misclassifications highlighted in Table 2 for the Inception V3, Inception Resnet V2, and Xception architectures are inspected to determine if the misprediction is the same image across the board. The image shown in Fig. 4 is an example of an image that was misclassified for all three architectures, thus suggesting the image represents a high number of features that stem from a straight backlift. Further analysis of the respective image also highlights the difficulty to recognise the positioning of the bat. Due to the bat being one of the key determining factors of distinguishing between a lateral and straight backlift, it could be the reason for the misclassification. Reasons as to why the architectures make these false-positive predictions may be due to the bat being blurred, making it hard to determine the angle at which the bat is faced. However, the batsman involved is Quinton de Kock, who is known to have a lateral backlift.

The Inception V3, Inception Resnet V2 and Xception architectures have similar results making it difficult to determine which architecture would be best suited for the problem domain. At the same time, additional data in the future may provide a more concise and deeper knowledge. Each architecture within this study has been trailed and tested in a generalised context as seen in Table 1. These architectures have proven to be successful in various domains, contributing to advancements in computer vision. Traditionally the Xception architecture is said to yield better performance, which is largely due to its Top-1 accuracy<sup>26</sup>. The Top-1 accuracy measures the proportion of examples for which the predicted label matches the single target. The loss score of 0.05% for the Xception architecture further highlights the architectures ability to predict on a single image. It yielded the lowest lost score from all architectures coupled with the past performance Top-1 accuracy scores<sup>26</sup>, and it can be concluded that the Xception architecture is best suited for the problem domain. The performance of the Xception architecture further highlights the added benefit of transfer learning, the architecture both in a generalised context, as seen on the ImageNet dataset, and specialised context, as seen in this study, has outperformed the Inception V2 and Inception Resnet V2 architectures, thus highlighting its efficiency within the problem domain.

One study by<sup>18</sup> identified and categorised various cricket batting shots from various videos. The approach and methods undertaken were based on deep convolutional neural networks. The first approach uses 2D CNN's with recurrent neural networks for processing video footage. The second approach implements a 3D CNN to capture spatial and temporal features simultaneously. Using a dataset with approximately 800 batting shot clips, the proposed model achieved a 90% accuracy<sup>18</sup>. The high accuracy obtained in the study indicates the high



**Figure 4.** The image that is mispredicted by the Inception V3, Inception Resnet V2, and Xception architectures as a straight backlift, where the image represents a lateral backlift<sup>35</sup>.

implications of modern artificial intelligence and deep learning in applications for detecting various cricket activities and decision-making purposes. Similarly, in our study, the high accuracy was revealed to be 98.2%, further validating their argument.

A previous study by<sup>13</sup> analysed the batting backlift technique using hand crafted features in the Open Computer Vision (CV) library, Android, and JavaScript. The system comprised of three main components; frontal view interface, lateral view interface, and a back-end system. The system was able to detect the type of backlift presented by the batsman by analysing and tracking the positional placement of batsman and the bat. The system was novel and provided a means to gather real-time data, which could be used for analysis. The improvements that deep learning has made within current research suggests that applications such as the one presented by<sup>13</sup> can be improved drastically and requires further investigation.

Cust et al.<sup>5</sup> postulated that “future work should look to adopt, adapt and expand on current models associated with a specific sports movement to work towards flexible models for mainstream analysis implementation. Investigation of deep learning methods in comparison to conventional machine learning algorithms will be of particular interest to evaluate if the trend of superior performances is beneficial for sport-specific movement recognition.” The approach demonstrated in this paper attempted to fill this void in which automation and automated recognition (coupled with deep learning methods) are key for player performance analysis in real-time.

## Conclusion

This research article was able to achieve an end to end backlift recognition, highlight the advantages of transfer learning, and identify the Xception architecture as the best performing architecture for the backlift recognition task, which is a largely unexplored area, and create a new dataset within the domain. The AlexNet architecture was noted for its significant breakthrough in machine learning and computer vision-related tasks. Since then, newer architectures have been introduced, such as Inception V3, Inception Resnet V3, Xception, and more. The objective of this research study was to make distinctions between the different types of backlift techniques in cricket. Having discussed the results proposed by each architecture, the aim of this approach has been achieved. With the Xception architecture performing optimally, it would be useful further to investigate other biomechanical networks in future sports science research. This study also provides a way forward in the automatic recognition of player patterns and motion capture, making it less challenging for sports scientists, biomechanists, and video analysts working in the field. Future improvements can be made by investigating more finer grade movements by looking at segmenting specific objects using semantic segmentation methods<sup>36,37</sup> and factoring more temporal aspects such as finer grain gestures<sup>38</sup>. Further investigation is required to determine how a model should be trained to maximise the benefit for both coaches and athletes for all sports. Furthermore, future work should evaluate the generalisation ability of similar models in a match situation and analyse players of varied demographics, including age, gender, skill level and format type. The models could provide better correlations to batting backlift technique as well as cricket batting performance.

Received: 26 August 2021; Accepted: 11 January 2022

Published online: 03 February 2022

## References

- Noorbhai, M. & Noakes, T. Advances in cricket in the 21st century: Science, performance and technology. *Afri. J. Phys. Health Educ. Recreat. Dance* **21**(4.2), 1310–1320 (2015).
- Yang, W. Analysis of sports image detection technology based on machine learning. *EURASIP J. Image Video Process.* **2019**(1), 1–8 (2019).
- Soomro, N., Sanders, R. & Soomro, M. Cricket injury prediction and surveillance by mobile application technology on smart-phones. *J. Sci. Med. Sport* **19**(6), 2–32 (2015).
- Xu, C., Cheng, J., Zhang, Y., Zhang, Y. & Lu, H. Sports video analysis: Semantics extraction, editorial content creation and adaptation. *J. Multimed.* **4**(2), 2009 (2009).
- Cust, E. E., Sweeting, A. J., Ball, K. & Robertson, S. Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance. *J. Sports Sci.* **37**(5), 568–600 (2019).
- Balsalobre-Fernández, C., Glaister, M. & Lockey, R. A. The validity and reliability of an iPhone app for measuring vertical jump performance. *J. Sports Sci.* **33**(15), 1574–1579 (2015).
- Dartfish, Dartfish, <http://www.dartfish.com>, Accessed: 2021-2-3 1999.
- Noraxon, <http://www.noraxon.com>, Accessed: 2021-2-3 (Apr. 2017).
- Quintic Consultancy Ltd, Quintic sports biomechanics video analysis software and sports biomechanics consultancy, <http://www.quintic.com>, Accessed: 2021-2-3.
- Cricket-21, <http://www.cricket-21.com>, Accessed: 2021-2-3 (2011).
- eagleeyedv.com, <http://www.eagleeyedv.com>, Accessed: 2021-2-3 (2012).
- McGrath, J. W., Neville, J., Stewart, T. & Cronin, J. Cricket fast bowling detection in a training setting using an inertial measurement unit and machine learning. *J. Sports Sci.* **37**(11), 1220–1226 (2019).
- Noorbhai, H., Chhaya, M. M. A. & Noakes, T. The use of a smartphone based mobile application for analysing the batting backlift technique in cricket. *Cogent Med.* **3**(1), 1214338 (2016).
- Noorbhai, H. A systematic review of the batting backlift technique in cricket. *J. Hum. Kinet.* **75**, 207 (2020).
- Noorbhai, M. H. & Noakes, T. D. A descriptive analysis of batting backlift techniques in cricket: Does the practice of elite cricketers follow the theory?. *J. Sports Sci.* **34**(20), 1930–1940 (2016).
- Noorbhai, M. & Noakes, T. The lateral batting backlift technique: Is it a contributing factor to success for professional cricket players at the highest level?. *South Afr. J. Sports Med.* **31**(1), 1–9 (2019).
- Moodley, T., & van der Haar, D. Cricket stroke recognition using computer vision methods. In *Information Science and Applications*, 171–181 (Springer, 2020).
- Khan, M. Z., Hassan, M. A., Farooq, A. & Khan, M. U. G. Deep cnn based data-driven recognition of cricket batting shots. In *2018 International Conference on Applied and Engineering Mathematics (ICAEM)*, 67–71. IEEE (2018).
- Semwal, A., Mishra, D., Raj, V., Sharma, J., & Mittal, A. Cricket shot detection from videos. In *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–6 IEEE (2018).
- Noorbhai, M. H. The batting backlift technique in cricket, Doctoral dissertation, University of Cape Town (2017).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural. Inf. Process. Syst.* **25**, 1097–1105 (2012).
- Alom, M. Z. *et al.* The history began from alexnet: A comprehensive survey on deep learning approaches (2018) arXiv preprint [arXiv:1803.01164](https://arxiv.org/abs/1803.01164).
- Wencheng, C., Xiaopeng, G., Hong, S. & Limin, Z. Offline Chinese signature verification based on alexnet. In *International Conference on Advanced Hybrid Information Processing*, 33–37. (Springer, 2017).
- Lu, S., Lu, Z. & Zhang, Y.-D. Pathological brain detection based on alexnet and transfer learning. *J. Comput. Sci.* **30**, 41–47 (2019).
- Moodley, T., & van der Haar, D. Scene recognition using alexnet to recognize significant events within cricket game footage. In *International Conference on Computer Vision and Graphics*, 98–109 (Springer, 2020).
- Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258 (2017).
- Mednikov, Y., Nehemia, S., Zheng, B., Benzaquen, O., & Lederman, D. Transfer representation learning using inception-v3 for the detection of masses in mammography. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 2587–2590 (2018).
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826 (2016).
- Kamble, R. M. *et al.* Automated diabetic macular edema (dme) analysis using fine tuning with inception-resnet-v2 on oct images. In *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE, pp. 442–446 (2018).
- Szegedy, C. *et al.* Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence* (2017).
- Sarki, R., *et al.* Convolutional neural networks for mild diabetic retinopathy detection: An experimental study, bioRxiv 763136 (2019).
- Torrey, L., & Shavlik, J. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. IGI global, pp. 242–264 (2010).
- Ide, H., & Kurita, T. Improvement of learning for cnn with relu activation by sparse regularization. In *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 2684–2691 (2017).
- Buckland, M. & Gey, F. The relationship between recall and precision. *J. Am. Soc. Inf. Sci.* **45**(1), 12–19 (1994).
- cricket.com.au, Quick wrap: De kock keeps proteas on top. <https://www.youtube.com/watch?v=2XXUm0u89nY&t=47s>
- Jiang, D. *et al.* Semantic segmentation for multiscale target based on object recognition using the improved faster-rcnn model. *Futur. Gener. Comput. Syst.* **123**, 94–104 (2021).
- Huang, L. *et al.* Jointly network image processing: multi-task image semantic segmentation of indoor scene based on cnn. *IET Image Proc.* **14**(15), 3689–3697 (2020).
- Liao, S. *et al.* Occlusion gesture recognition based on improved ssd. *Concurr. Comput. Pract. Exp.* **33**(6), e6063 (2021).

## Author contributions

T.M. carried out the data collection and implementation of the experiment. D.V.D.H. designed the deep learning experiments and assessment parameters. H.B. provided domain expertise and guidance on the lateral and straight classification categories All authors contributed to the write up of the article--update author contribution from author folder XML file.

## Competing interests

The authors declare no competing interests.



### Additional information

**Correspondence** and requests for materials should be addressed to D.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022