Advances in Artificial Intelligence and Machine Learning; Research 2 (4) 557-566 Received 10-12-2022; Accepted 24-12-2022; Published 30-12-2022

Can a Face Tell Us Anything About an NBA Prospect? - A Deep Learning Approach

Andreas Gavros agavros@arch.auth.gr
School of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Central Macedonia 54638,
Greece

Foteini Gavrou foteinigavrou@gmail.com
Independent Researcher
Greece

Corresponding Author: Andreas Gavros

Abstract

Statistical analysis and modeling is becoming increasingly popular in professional sports organizations. Sophisticated methods and models of sports talent evaluation have been created for this purpose. In this research, we present a different perspective from the dominant tactic of statistical data analysis. We deploy Convolutional Neural Networks in an attempt to predict the career trajectory of newly drafted players from each draft class. We created a database consisting of about 1500 image data from players in every draft class since 1990. We then divided the players into five different quality classes based on their NBA career. Next, we trained popular image classification models in our data and conducted a series of tests in an attempt to create models that will provide reliable predictions of the rookie players' careers. The results of this study suggest that there is a potential correlation between facial characteristics and athletic talent, worth of further investigation.

Keywords: Convolutional neural networks, Deep learning, Transfer learning, Computer vision, NBA.

1. INTRODUCTION

The inspiration for this research was drawn from a series of articles published in 2014 [1]. NBA professional teams, most notably the Milwaukee Bucks, collaborated with facial coding experts in an effort to evaluate the talent of players who had declared for that year's draft class. So we made an attempt to create artificial intelligence models that would be able to mimic this human activity. Based on this logic,the goal of this approach was to analyze certain facial characteristics to make an assessment of whether a young player is capable of standing at the NBA level and how good a player will be.

Citation: A. Gavros and F. Gavrou. Can a Face Tell Us Anything About an NBA Prospect? - A Deep Learning Approach. Advances in Artificial Intelligence and Machine Learning 2022;2(4):38

Given the above, a legitimate question arises. Is there any truth or scientific basis to this analysis? Even further, what are the physiognomic characteristics that can provide information about aspects of a person's personality (if any) and more specifically their ability in a sport (basketball)? In this paper we will not attempt to answer this question. Instead we will attempt to create Convolutional Neural Networks (CNNs) that will perform this task and evaluate them. CNNs are capable of extracting far more information from an image, than a human observer [2]. Therefore, CNNs theoretically will be able to outperform a human in assessing talent based on visual stimuli. Obviously, we do not propose the methodology we present in this study as a deterministic model to assess a player's potential talent or career. The main aim of this paper is to explore one of the countless possibilities of CNNs and to provide an innovative approach to evaluate athletic talent.

2. RELATED WORK

Since 1960 that sports analytics emerged as a field of analysis in basketball [3], statistical analysis of professional sports has made rapid progress. The success of Oakland Athletics' use of statistical methods in their decision-making process [4], turned the attention of many sports organizations to the benefits of applied statistics [5]. In the field of basketball, which is the subject of the present work, much research has been conducted on the evaluation of particular game strategies (the value of shooting threes in a game for example) [6], and the evaluation of in-game decisions by coaches [7].

CNNs have been deployed, to process image input such as video and produce real-time data of a basketball game and measure its effects on both players and coaches [8]. Deep Learning (DL) can be deployed to analyze which offensive play has the best performance in a basketball game and then optimize these systems to make them even more efficient [9]. DL models are capable of finding patterns in the offensive movements of players and suggesting optimal distances between players, and providing optimal positions on court for the initiation of plays [10]. Analyses also, have been conducted to explore the characteristics of draft prospects that influence decision-makers of NBA teams front offices to make their decisions [11]. Of course, teams are more interested in assessing a player's ability and whether he will be able to successfully adapt to the demanding environment of the NBA. But here an important and obvious question arises: Can ML, CNN or DL algorithms outperform their human competitors in basketball talent evaluation? And if the answer to this question is yes, in what areas, in addition to those currently known, can these models be applied to provide even more critical information for NBA teams? In the latter question we will attempt to provide a new perspective by presenting this research.

3. MATERIALS AND METHODS

The basis of the hypothesis we try to explore in this study is whether we can make one rough approximation of a rookie player's career, based solely on his facial features. We will then analyse the methodology we followed to produce the relevant predictions.

3.1 Rationale of Study

The analysis was based exclusively on the data available from the 1990 season to the 2019 season, a period of 30 years, which is a relatively big amount of time regarding professional team sports. This choice was based on the following two factors: First, the availability of data. Image data of sufficient quality before 1990 is much more difficult to obtain. Second, the style of the game. Arguably, the game of basketball changed dramatically after the arrival of Michael Jordan [12]. Of course, for a few years after his selection in the 1984 draft [13], there were players who came from a much different era of the game. Therefore, we made the hypothesis that from 1990 onwards, the modern era of basketball began. We also decided to work exclusively on the rookies (i.e. the newcomers to the NBA league). Finding the next great talent is a problem that all teams face every year. Based on the interest and importance of the problem, we decided to base our analysis on rookie talent evaluation.

3.2 Method of Classifying Players Into Talent Categories

We divided the players into five classes based on their *Potential talent*. *Potential talent* is one of the key concepts of this study and refers to the expected career of a player and not on the career he eventually had or will have. Furthermore, *Potential talent* refers to the career that a certain player was expected to have and not in the total numbers that a player put up in his career. To make this logic clearer we will present some real-life examples:

- The career stats for a player like Derrick Rose show a player who had some great seasons, but then had an overall mediocre career. But to classify a player of this quality as mediocre would be mislabelling, so he was classified as excellent.
- Greg Oden was unable to have a meaningful NBA career due to injuries. However, he was an arguably great talent and it is certain that if healthy, he would have had a tremendous career. Therefore, his potential talent was also classified as excellent.

The assessment of potential talent was based more on the author's criterion, rather than on statistical data. We did not use a specific formula for evaluating a player's quality, nor did we attempt to create our own method. Inevitably this option will lead to data mislabeling, which will be more apparent in classes that report on lower-level talent. However, the aim of this study was not to propose a method of evaluating a player's career with a mathematical formula. Our goal was to propose a talent prediction method and that is what we focused on. The talent evaluation process for each player individually, was a time-consuming process, which can nevertheless be completed in a reasonable amount of time, as the size of the overall database allows it. The categories (classes) in which this classification was made, were the following:

- **Not-ready:** Players that did not have the talent required to play in the NBA level and did not manage to have an impact on this level of competition.
- **Lower level:** Players that were part of an NBA team roster, but they were limited-role players at best, with low playing time.

Figure 1: Example of the typical player image that was used in the database

Table 1: Distribution of quality classes in the database.

Quality Class	Not-ready	Lower level	Mediocre	Very Good	Excellent
Number of players	607	352	327	237	146

- **Mediocre:** Players that had some decent seasons and had a significant role in the teams they played, but were not considered among the stars of the league.
- **Very Good:** Players that had a good overall career but teams were not necessarily ”built” around them, however they were essential parts of their teams.
- **Excellent:** Players that had the talent to be an All-Star for one or multiple seasons, or were particularly good in a certain part of the game (rebounding for example).

The assessment of players was based on their whole career and not on particular seasons. The rational of this decision lies on the the various parameters that can affect a player’s performance in a single season. A player on a team that is uncompetitive for a season (therefore even a player of average talent, has the opportunity to increase his averages), a good season by a player that did not continue in subsequent years, are factors that are misleading in talent evaluation. Based on this reasoning we decided to disregard particular seasons of a player (either good or bad) and focus on the bigger picture of a player’s career.

3.3 Data Collection

The database was created on the logic that models should be trained exclusively on the athlete’s facial features, excluding body physical characteristics from the analysis. An example of a typical image that was used in training session of the models, is provided in FIGURE 1. The aim of the study is to explore the potential of a player as a rookie, so it was decided to use images of athletes as rookies. The images were accessed via web scraping. As a result, we were not able to have a consistency in the images we used in the database. In optimal conditions the database would consist exclusively of photos taken with the same procedure, have the same dimensions, the players would have the same pose, etc. This condition could not be met, as the players of the last thirty years are examined and as we go back in time the data becomes more and more inaccessible and the diversity in the way players are photographed increases. Therefore the database consists of highly differentiated images, a problem we tried to address through photo editing, to create a more ”homogenized” set of images.

Table 2: Distribution of quality classes in the dataset used to fit the pre-trained models.

Quality Class	Not-ready	Lower level	Mediocre	Very Good	Excellent
Number of players	300	300	300	230	140

3.4 Database

The dataset consists of 1669 rookies. Distribution of classes is presented in TABLE 1. There is uneven distribution of data, as ”Not-ready” class makes up almost 1/3 of the database and ”Excellent” class is considerably smaller. This result reflects reality, as most of the players in each draft class fail to stand up at the NBA level, let alone be good prospects. This distribution of data will have an impact on the training of DL models, a problem discussed later. The size of the resulting database is quite small compared to most projects CNNs are deployed. However as this is a real-life problem it is an expected challenge.

3.5 CNN Model Development

As this study deals with an image-based problem, we chose to deploy *Transfer Learning*, which is one of the most popular options in image classification problems. The most well-known models have been trained for ImageNet Challenge, which is a quite complex and demanding task. They have the ability to find patterns in image data that are related with emotions, diseases, etc [14,15]. The pre-trained models we trained on our database were *ResNet-50*, *Xception*, *Inception-V3* and *VGG-16*. *VGG-16* model stacks together multiple convolutional, followed by a max-pooling layer. A 1x1 filter is set as last convolution layer in some layers, to induce non-linearity in the model, thus making it able to be deployed in more complex problems [16]. *ResNet-50* is a very deep CNN model consisting of 50 layers. The central idea of this model, is that it is built around the concept of *skip-connections*. Skip-connections are deployed to connect the input directly to the output of the layer, while skipping a number of connections [17]. *Inception-V3* is making use of various filters of varied size in order to prevent overfitting (the phenomenon of a model overly adapting to the train data). Its main difference with other models, is its ”wider” network architecture, consisting of multiple convolutional layers, of no great depth. *V3* in particular requires less computational resources, in comparison to *V1* and *V2* so it was considered as the the optimal choice out of the Inception ”family” of models [18]. The last model we used was *Xception*. It is wider known as an extreme version of Inception model. This results from the depthwise convolution it deploys, that is also followed by a point-wise convolution, which is not applied in all layers in order to be more efficient in terms of computational cost [19].

The small size of the resulting database allowed us to perform several tests for each pre-trained model (TABLE 2). For each model, more than 50 tests of different network architectures were tested. Successful architectures were further explored, with more tests conducted (reaching to more than 100 for each pre-trained model). Tests included the use of different cost/activation functions, number of training epochs, learning rate, optimizers, number of dense layers and the number of neurons in each dense layer. Also, tests were made on the number of hidden layers of the pre-trained models that would be trained on the data.

Finally, to deal with the problem of unequal distribution of data in the classes it was chosen to truncate some classes to create more evenly distributed data. It was chosen to delete inputs from the classes that had more data in favour of the top talent classes. The resulting final dataset is presented in TABLE 3. We made this decision in order to avoid training the models on a particular class, that would probably had an effect on the quality of predictions, putting us in the risk of producing CNN models that would be overly trained in players of lesser quality. Our goal is on the first level to produce reliable predictions for all classes, but on the second and equally important level to predict which players will have the biggest impact in the league. Therefore, our strategy was mainly driven around producing models that are more capable of producing reliable predictions for above average talent.

4. RESULTS

4.1 Results Evaluation Method

We used a 10-fold cross-validation process to evaluate the performance of CNN models. The database was split into 10 equally sized folds. Nine folds were used to train the model and the remaining one fold (10% of the database) was used as the unknown sample, where the model was called to make predictions. This process is eventually repeated for 10 times, until all the folds are used for evaluation. The metric we used to evaluate the CNN models was Accuracy, which is the most common metric in image classification projects.

As we are interested in producing predictions towards evaluating high-talented player prospects, we are interested in a metric that would evaluate the quality of predictions for each class. This metric is *Precision*, but we chose to rename it to be more understandable by people with no deep knowledge of DL. This metric is more oriented towards evaluating class-specific predictions: If a model manages to achieve a decent accuracy rate overall but has very low performance in *Very Good* or *Excellent* classes, it has practically very low value for an NBA organization. A model that does not achieve a high accuracy rate compared to other models, but manages to have higher rates in classes that refer to more talented players, then that model clearly has more value for this specific study. The name of the renamed metric was *Class-Prediction Quality* and is defined as follows:

$$\text{Class-Prediction Quality (\%)} = \frac{\text{Successful Predictions for a distinct class}}{\text{Predictions for a distinct class}} \times 100 \tag{1}$$

This metric takes more into account the predictions that a model produces. If a model produces 30 predictions for *Very Good* class and manages to be accurate in 10 occasions, it should be considered as a well-performing model. Of course, this metric can be biased if the total number of predictions is significantly low, so we are taking into account the total number of predictions for the concerned class, as well.

Table 3: Accuracy (%) of the best performing model of each group of pre-trained models in the unknown sample of images

Pre-trained model type	VGG-16	Inception-V3	ResNet-50	Xception
Accuracy rate (%)	22.6	23.62	24.49	26.77

4.2 Evaluation of Model Performance

The accuracy rates in the unknown sample of images were significantly low, however this was an expected result. Having high accuracy would be an alarming indicator that something went wrong in the training process or in the database configuration. The performance (accuracy rates) of the most successful models ranged between 20-25%.

1. **VGG-16:** *VGG-16* models did not provide high precision rates in the unknown sample of images. Although, some of the models managed to achieve accuracy rates of 24%, this was one of the low-end performing models. The performance of *VGG-16* models in the higher talent classes was rather low, with no network architecture achieving to overcome the threshold of 10% in *Very Good* and *Excellent* classes. In the lower level of talent classes, the models performed considerably better, with a model surpassing the mark of 25% accuracy for *Not-ready* and *Lower level* classes.
2. **ResNet-50:** The models of this group provided the worst performance. The *Class-Prediction Quality* rates in two highest talent classes were almost zero in all the models tested. As the results for this group of models were not encouraging, we chose to stop testing on more network architectures in order to focus on other models. Finding talent is quite difficult even for human experts, so the fact that the models fail in this area may prove that the models simulate human behaviour and they have an equally difficult time finding high-level athletic talent.
3. **Inception-V3:** *Inception-V3* models managed to achieve decent *Class-Prediction Quality* rates in all classes. The best performing model from this group, managed to display and accuracy rate of over 20% in four classes (*Not-ready, Lower level, Mediocre, Excellent*) and over 18% in *Very Good* class. The best performing models had a rather low number of neurons in dense layers ($<$100) and all these models deployed *Softmax* activation function.
4. **Xception:** The models of this group recorded the best performance overall. A small number of models had *Class-Prediction Quality* rates even above 25% in all classes. This performance was achieved with a number of 100-500 neurons in dense layers. Most successful network architectures used 3 dense layers. In all these layers the *Dropout* technique was applied, which had a beneficial effect on the performance. Specifically, it increased performance by 3-4% compared to models of identical network architecture in which *Dropout* was not implemented.

4.3 Best Performing Model

The best performing model will be presented in detail. This model achieved an overall accuracy rate of 26.77 %. The model performance by class is presented in TABLE 4. The selection of this model

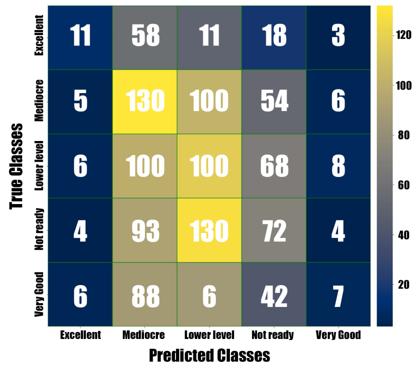

Figure 2: Confusion matrix of the predictions made by the best performing model

Table 4: Accuracy rate of best performing model, in each class.

	Not-ready	Lower level	Mediocre	Very Good	Excellent
Prediction Quality (%)	28.35	24.33	28.03	25	34.38

as the best, was based in its performance in the classes of higher talent. As shown in the confusion matrix (FIGURE 2) the model does not produce a large number of predictions for the *Very Good* and *Excellent* classes. Regarding the *Excellent* class, the *Class-Prediction Quality* is quite high, approaching 35%. The number of total predictions produced (32) indicates that the performance is not due to statistical error. The same applies to the *Very Good* class where the *Class-Prediction Quality* is not at the same level (25%), but it is still a quite remarkable performance based on a reliable number of predictions (28). The number of predictions for these two classes is significantly lower compared to the other classes, a feature that may be desirable in sports industry. Teams may favor a model that does not produce a large number of predictions, thus increasing the number of players that will be misjudged. Finally, we deployed this model to generate predictions for this season's rookies, and specifically for players from the second round of the draft, where such a model would have more potential for use. The model ranked only one player in the top two talent classes, namely the Excellent class: EJ Liddell who was drafted at number 41 by the New Orleans Pelicans. This choice (only one player) seems logical based on the fact that players selected lower in the draft have a lower chance of developing into league stars.

5. CONCLUSION

The performance of the models suggests that there is a potential correlation between players' talent and the characteristics of their foresight. This statement however, raises more questions than it answers. In the event that a correlation does exist, what are the characteristics that models see as associated with a person's predisposition to perform well in a particular sport? The present research is not able to provide a reliable answer to these questions. Future studies may be able to provide more data on the problem we faced here. The results of this research are quite encouraging, as in the higher talent classes the accuracy rates are high, subject to proportions of course. A first question is obvious: could these results have been better? As the images of the players used were in many cases of low resolution, a first idea is to create a database of higher resolution images. Then as there is a large heterogeneity in the images of the players, a possible improvement of the accuracy rates could be based on a dataset consisting of images of a certain posing-style. Finally, we would like to dwell once again on the nature of this work. This research would be optimally seen as an exercise on the unexplored potential of CNNs. To appear confident that the results of this paper represent a certain conclusion would be unrealistic. However, we do not consider it unlikely that in the future NBA teams will photograph rookie players as part of their evaluation during their interview process. The methodology developed in this paper may be part of the talent analysis tools for NBA teams in the future.

6. ACKNOWLEDGMENT

Foteini Gavrou provided the image data through web-scraping and formatted the player images. Andreas Gavros formed the database that was used to train the models. He developed and evaluated the CNN models. Lastly, he performed the literature review.

Most of the work dedicated to conducting this research was done under quarantine conditions for COVID-19 in the spring of 2020. Therefore, the authors of this research would like to dedicate this work to the friends and families of the victims of this disease, who experienced this unprecedented situation in the worst possible way.

References

[1] https://www.nytimes.com/2014/12/26/sports/nba-bucks-looking-for-an-edge-hire-expert-in-face-time.html

[2] Sarvamangala DR, Raghavendra K. Convolutional Neural Networks in Medical Image Understanding: A Survey. Evol Intell. 2022;15:1-22.

[3] Hughes M, Franks I. Notational Analysis of Sport. J Sports Sci Med. 2004;3:1.

[4] Lewis M. Moneyball: The Art of Winning an Unfair Game Paperback. W.W. Norton and company. 2004.

[5] https://www.dell.com/en-us/perspectives/tech-takes-the-court-nba-teams-turn-to-data-and-analytics-for-a-competitive-edge/

[6] Caporale T, Collier TC. To Three or Not to Three?. J Lab Res. 2015;36:1-8.

[7] Demenius J, Kreivyte R. The Benefits of Advanced Data Analytics in Basketball: Approach of Managers and Coaches of Lithuanian Basketball League Teams. Balt. J., Sport, Health Sci. 2017;1:8–13.

[8] Yao P. Real-Time Analysis of Basketball Sports Data Based on Deep Learning. Complexity. 2021;2021:1-11.

[9] Javadpour L, Blakeslee J, Khazaeli M, Schroeder P. Optimizing the Best Play in Basketball Using Deep Learning. J Sports Anal. 2022;8:1-7.

[10] https://dspace.mit.edu/bitstream/handle/1721.1/119728/1078687450-MIT.pdf?sequence=1&isAllowed=y

[11] Berri DJ, Brook SL, Fenn AJ. From College to the Pros: Predicting the NBA Amateur Player Draft. J Prod Anal. 2011;35:25-35.

[12] Lane E, chapter 7. Mortgaging Michael Jordans Reputation. In: Ogden DC, Rosen JN, Lule J, Fox RF, editors. Fame to Infamy: Race, Sport, and the Fall From Grace. MS: University Press of Mississippi. 2010.

[13] Mack M. Michael Jordan: A Biography. J. Sport Hist. 2010;1:178-179.

[14] Mostavi M, Chiu YC, Huang Y, Chen Y. Convolutional Neural Network Models for Cancer Type Prediction Based on Gene Expression. BMC Medical Genomics. 2020;13:1-3.

[15] Ouyang X, Zhou P, Li CH, Liu L. Sentiment Analysis Using Convolutional Neural Network. In: IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence.2015;2015:2359-2364.

[16] https://arxiv.org/abs/1409.1556

[17] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog. 2015:770-778.

[18] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. Clin Orthop Relat. 2015:2818-26.

[19] F. Chollet. ”Xception: Deep Learning With Depthwise Separable Convolutions,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017;1800-1807.