Does superior cognitive performance in football players predict future performance?

Research Project 2
36EC

Author : Nutsa Nanuashvili

Student number: 11699450
Master of Brain and Cognitive Sciences
University of Amsterdam

Supervisor and assessor : Dhr. Dr. I.G. Sligte
Examiner: Dhr. Dr. Y. Pinto

Faculty of Social and Behavioural Sciences
University of Amsterdam

UNIVERSITY OF AMSTERDAM
Institute for Interdisciplinary Studies

July 2020

Contents

Abstract

To successfully navigate a rapidly and dynamically changing game environment, top-level football players need to have superior cognitive abilities, in addition to exceptional motor and perceptual skills. The aim of this study was to examine the predictive value of a specific set of cognitive abilities, the so-called executive functions, in distinguishing between pro-level and elite-level football performance. Using the NeurOlympics test battery (BrainsFirst, 2016), the following executive functions were measured over a period spanning approximately 6 years (2014-2020): working memory, spatiotemporal anticipation, control, and attention. A logistic ridge regression algorithm was used to predict the current football performance level (elite or professional) of 229 adult football players based on the cognitive data that had been collected 0 to 6 years ago. The model could accurately classify based on the historical cognitive data whether a football player would play at an adult elite or adult professional level years later. Elite-level cognitive model scores, calculated as class probabilities, correlated moderately with an objective football success parameter, namely current market value. In addition, we trained models on data including only the players who were adults (above 18 years old) and those who were youth players during the cognitive test administration. Both models showed comparable predictive ability. However, the overall executive function profiles that were predictive of the current performance level differed when using adult or youth cognitive data. Future research should disentangle which combination of executive functions can best predict future football performance when cognition is tested at specific ages. This study shows that executive functions can effectively be used as additional criteria (besides physical, tactical, and motivation aspects) to predict future football performance, both in youth players and in adult players. This paves the way towards a novel approach of talent identification and development in football.

Introduction

Football players perform under high pressure in a rapidly and dynamically changing game environment. To put their physical abilities to use in an efficient way, they need to rely on their cognitive functions, such as working memory, attention, cognitive flexibility, inhibitory control and metacognition (Vestberg, Gustafson, Maurex, Ingvar, & Petrovic, 2012). These cognitive functions collectively are referred to (as part of) the executive functions (EFs) (Diamond, 2013). EFs are believed to enable goal-directed behaviour and mental flexibility that is needed in a highly demanding and unpredictable setting where acting automatically would lead to mistakes. Elite football players have been shown to outperform non-elites and the general population on test batteries measuring these EFs (Huijgen et al., 2015; Verburgh, Scherder, Van Lange, & Oosterlaan, 2014; Vestberg, Reinebo, Maurex, Ingvar, & Petrovic, 2017). Huijgen and colleagues (2015) showed that even after controlling for weekly training duration, young elite players showed significantly better inhibitory control and cognitive flexibility compared to a sub-elite group (Huijgen et al., 2015). Interestingly, Verburgh et al. (2014) could also dissociate highly talented and amateur players with 89% accuracy when EFs were used as predictors in a logistic regression model (Verburgh, Scherder, et al., 2014). Furthermore, a recent meta-analysis compared core executive functions in high- versus low-performance athletes and demonstrated significantly superior performance in the former group (Scharfen & Memmert, 2019a). Thus, apart from the exceptional motor and perceptual skills, top-level football players seem to possess superior executive functions frequently referred to as ‘game intelligence’ (Williams & Reilly, 2000).

Identifying the specific executive functions that are essential for successful football performance opens up possibilities for a novel approach to football talent scouting. The main aim of the scouting practice is the early identification and investment in players who have the potential to achieve elite-level performance in the future (Williams & Reilly, 2000). Recently, it has been proposed that taking players' cognitive abilities into account could provide additional insight into the prospects of a young player's performance. This will enable coaches and clubs to make more objective decisions about talent scouting (Beavan, Spielmann, & Mayer, 2019; Huijgen et al., 2015; Sakamoto, Takeuchi, Ihara, Ligao, & Suzukawa, 2018; Scharfen & Memmert, 2019a). However, before going down this avenue, scientific research still needs to identify the specific executive functions that distinguish adult elite (world-class) from adult expert (professional) football players.

The existing scientific literature suffers from several limitations that currently make it impossible to assess the predictive ability of executive functions in future football performance level. Different executive functions have been shown to have different development rates from early childhood throughout adolescence (Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Luciana, Conklin, Hooper, & Yarger, 2005; Verburgh, Scherder, et al., 2014). This makes conclusions that scientific studies draw on the relationship between EFs and football performance, using small samples of preadolescent and adolescent players, problematic (Beavan et al., 2019; Huijgen et al., 2015; Sakamoto et al., 2018). The main problem with this approach is that during testing the athletes' brain is still maturing and the measured EFs are not yet at their peak of development. Additionally, the majority of these studies only employ a cross-sectional approach and there is no follow-up record of the young athlete's future performance (Beavan et al., 2019; Scharfen & Memmert, 2019b, 2019a; Verburgh, Scherder, et al., 2014; Voss, Kramer, Basak, Prakash, & Roberts, 2010), so the predictive validity is unknown. Furthermore, to investigate the necessary EFs for elite-level performance, it is crucial to have a clear definition of 'elite' and 'expert' athletes. In the context of football, elite players are world-class players competing in the first teams of football clubs ranked as the world's top one hundred (https://footballdatabase.com/ranking/world). On the other hand, expert athletes are those who play football professionally in lower-ranking teams. Previous studies only compared EFs of professionals to amateur players (Verburgh, Scherder, et al., 2014) or players from youth elite academies to non-elite peers (Beavan et al., 2019; Vestberg et al., 2017). So, the question still remains which specific EFs can distinguish adult elite players (world-class players) from adult professional (expert) players (for review see, Scharfen & Memmert, 2019). Therefore, the results from previous work cannot be used to infer the relation of executive functions essential for elite athletes and future success in football. Overcoming these obstacles in scientific research would pave the way for youth sports clubs to implement EFs, along with other criteria, into talent identification programs.

The goals of the current study are to overcome the limitations of the earlier work and to create reliable metrics to further facilitate cognition-based identification of future elite players. In the first step, we will use a machine learning algorithm to classify adult football players into elite or professional based on cognitive data collected several years before. This will allow us to create an 'elite football brain' model that is indicative of the importance of EFs for elite-level football performance. Next, as an exploratory analysis, we will train the classifier only on adult data and apply this elite football brain model to the EF data from youth players. This will allow us to see whether using only mature EFs in the model can be used to predict future football level of youth players. Finally, we will create "youth elite football brain" and "adult elite football brain" models by training classifiers using youth and adult EFs separately. This will show whether current football level can be accurately predicted based on EF data collected during youth and adulthood as well as it will allow us to compare different EFs important for adult and youth elite-level players. Thus, we aim to create a model of the elite football brain that can be used as an additional criterion to predict the future performance of football players.

To assess essential executive functions for elite-level performance, the Dutch company BrainsFirst has developed a set of online games called NeurOlympics (see demo https://neurolympics.nl/campaign/index-default.html?c=43). Each game is created to test four executive functions and their subcomponents: working memory, anticipation, attention, and cognitive control. Executive functions of adult elite and adult professional players have been tested with NeurOlympics game. Additionally, using the same NeurOlympics games, EF data of youth players were collected twice per year, 0 - 6 years ago. Based on the data of presently adult players, who were either adults or youth during the test, an "elite football brain" model will be created by using EFs as predictor variables and the current football success - elite and professional as an outcome. We hypothesize that elite football brain model will be able to accurately predict football level based only on players EFs. Furthermore, youth players classified as "elites" will be more successful 0-6 years after having taken the EFs test. To assess whether players with an "elite football brain" also are more successful in the future, their historical elite football brain model scores will be correlated to a measure of their current football success, namely their current market value.

We expect that adult and youth players with an elite football brain have higher market value 0-6 years after test administration than players with a professional football brain. It is important to note that besides cognitive abilities, there is an interplay of multiple factors that determine the future performance of young players. To be successful in selecting young players with the best potential, one has to consider physical, psychological, and physiological profiles (Abade et al., 2014; Murr, Raabe, & Höner, 2018; Williams & Reilly, 2000) as well as technical and tactical capabilities (for review see, Sarmento, Anguera, Pereira, & Araújo, 2018). Considering this, we expect to see only a modest correlation (r = .2-.3) between market value and elite football brain model scores. We argue that although the executive functions can provide only a piece of the bigger puzzle, taking them into account will facilitate in a more holistic approach to talent identification.

Methods

Predictor cognitive variables

Players were tested with the NeurOlympics games (see demo https://neurolympics.nl/campaign/index-default.html?c=43) twice per year, from 2014 to 2020. Overall, the games took about 45 minutes with small variation per participant. The NeurOlympics games consists of four games and each one is measuring different EFs: working memory, anticipation, control, and attention, in this order.

The first game is called "Collect" and measures working memory. It was created based on the work of McNab et al. (2015) and others (Gold et al., 2006; Johnson et al., 2013; KyllingsbæK & Bundesen, 2009; McNab & Dolan, 2014; McNab et al., 2015; Xie et al., 2017). The game has a very high test-retest reliability (r = .90). The main task on this game is to remember the position or the orientation of the displayed stimuli (Figure 1A). Distractors are embedded among the target stimuli either during the encoding (encoding distractors), or during the retention phases of working memory (delayed distractors). The four aspects of working memory tested by this game are capacity, identity, composure (or resilience), and filtering. Capacity measures the total number of stimuli remembered where grouping can be used; identity: detail-specific memory for the total number of correctly remembered orientations of stimuli; composure/resilience: the total number of remembered stimuli when distractors are present during the retention interval; filtering: the ability to ignore irrelevant information (distractors) while encoding the memory items.

The second game, called “Activate”, is designed to test spatiotemporal anticipation skills and the ability to maintain performance level under stress (Figure 1B). The game was custom-made by BrainsFirst and has a test-retest reliability of *r* = .80. The subject needs to accurately anticipate the trajectories of stimuli, using spatial and temporal information. The task is to anticipate the direction and speed of the falling stimuli and hit them with ammo that is released from four cannons (Figure 1B). The sooner (top of the screen) the stimuli get hit, the more points will be awarded. The performance on the game is measured as the time spent until challenge (in minutes; reaching 4000 points), after which the pace of the game accelerates. The less time spent, the better the performance. The second component tested on the game is stress resilience: the number of minutes spent in challenge mode, when the pace of the game picks up. The more time spent in challenge mode, the better the coping ability under pressure.

The third game called “Connect”, measures cognitive control, important for quickly adjusting behavior according to task demands (Isoda & Hikosaka, 2007; Neubert, Mars, Buch, Olivier, & Rushworth, 2010). The task is to rapidly indicate whether the left or right stimulus is congruent with the middle one (Figure 1C). The color of the middle stimuli alternates between red and blue and the reference stimuli appear either left or right of the middle stimulus. After several repetition trials, the side of the target stimulus or the color of target stimulus changes, and these changes require inhibition of the automatic behavior and mental flexibility. The four assessed components are speed: reaction time on no switch, repetition trials; automatic: accuracy on repetition trials; inhibition: accuracy on side-switch trials compared to repetition trials; flexibility: reaction time when the colors change. The game has a good test-retest reliability (*r* = .83).

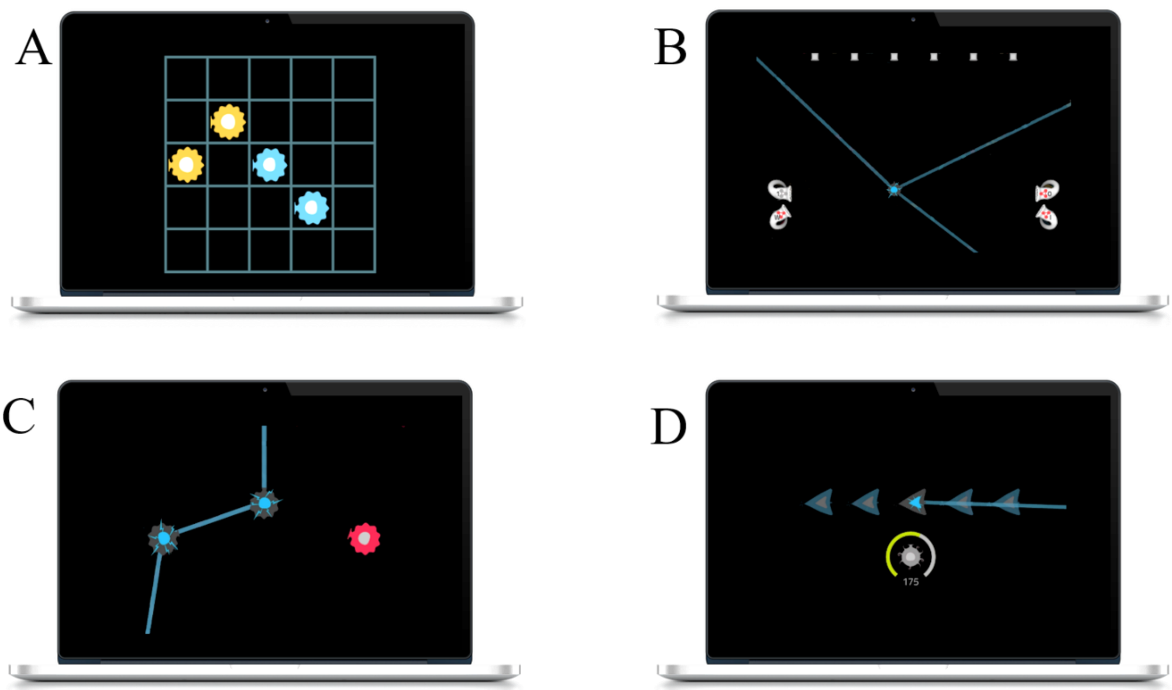


Figure 1. Visual illustration of NeurOlympics games: **(A)** Collect tests working memory **(B)** Activate tests anticipation **(C)** Connect tests control **(D)** Synchronize tests attention

The fourth game, called “Synchronize”, tests attention and is based on Attention Network Task (ANT) (Fan et al., 2009; Fan, McCandliss, Fossella, Flombaum, & Posner, 2005; Fan, McCandliss, Sommer, Raz, & Posner, 2002). In this game, the target stimulus appears in the middle of the screen and is flanked by similar stimuli. The flanking stimuli either point in the same direction of the target (congruent trial) or in the opposite direction (incongruent trial). The task for a player is to react according to the direction of the middle (target) stimulus. In some trials, the onset of the stimuli is preceded by cues, located either in the middle, to alert the participant’s attention, or at the spatial location where the target stimulus could appear. These latter spatial cues either correctly indicate the upcoming stimulus location or incorrectly – the wrong spatial location. Based on accuracy and reaction time, six components of attention are measured: speed, performance, concentration, move, disengage, and guide. First is the speed of the subject’s decision on the game (reaction time); performance: accuracy of the performance; concentration: ability to increase reaction time relative to accuracy when the central cue is present; move: an ability to increase reaction time relative to accuracy when the correct spatial cue is present; disengage: an ability to maintain accuracy (decrease ratio of reaction time and accuracy) when the false special cue is present. Finally, guide accounts for the capacity of prioritization and self-reflection of the performance and is measured as the difference in the ratio of reaction time and accuracy between the congruent and the incongruent trials. The game has a good test-retest reliability (r = .80).

For every condition of each game, the resulted raw scores are saved as separate variables, 53 in total. From these raw scores 16 above mentioned EF components are calculated. Overall, it comprises of 69 cognitive variables in the data.

Participants and Data Preparation

The initial dataset included data of 623 male football players. Because the majority of players were tested more than once (min. 1 and max. 7), each testing session was included in the data as a separate data point (observation). Thus, the initial dataset contained 1479 separate observations (mean current age 19.38, SD = 3.79). The participants comprised of a mixture of players who were either 18 years old, or younger at the age of testing (n = 474), and players who were adults at the age of testing (n = 149). Tested adult participants (mean age at test, 22.40, SD = 3.76) played in one of the Dutch premier league clubs namely, AZ Alkmaar, PSV Eindhoven, Vitesse Arnhem or Willem II. Whereas youth participants (mean age at test 15.02, SD = 1.85), were enrolled in youth academy of one of the above-mentioned clubs. The participants were tested twice per season over 5 years, starting from 2014.

In addition to collecting the cognitive variables, we collected data about each player’s current career performance from https://www.transfermarkt.com/. The data included performance information from the latest season played by each player. The important variables for the analysis were current age, market value, current club, level of the league and national team of the player. Each subject was given a unique id, which we used to accurately link the performance information to their cognitive variables collected from NeurOlympics game. Based on the performance information, we sorted players and used the result as an outcome variable in the analysis (Figure 2). We categorized each player either into youth or adult (> 18 years old), and whether the players had a market value or not. Young players and adult players who did not have a market value were excluded for further analysis. From the adult players with market value, we sorted players as “elite”, who played either in world top 100 hundred clubs (https://footballdatabase.com/ranking/world/1), or in a U-23/A-team of top 15 countries (https://www.fifa.com/fifa-world-ranking/ranking-table/men/#all). Players who did not meet the “elite criteria” but had a market value or their league level was second tier, were categorized as "professional" (pro) players. The result was the dataset including adult players either belonging to elite, or professional football level.

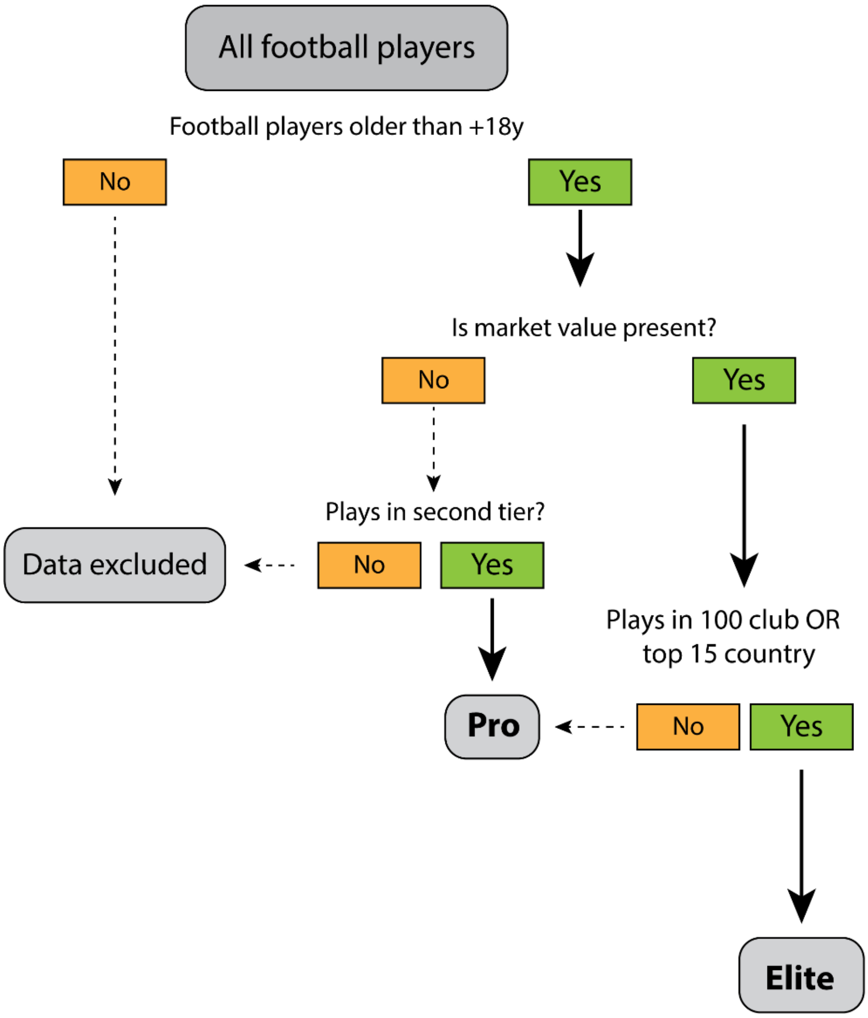


Figure 2. Decision-tree of sorting football players into elite and professional (pro)

Logistic ridge regression

Our main aim was to create the elite football brain model that would be able to generalize to new data of young football players and predict their future performance. We decided against a linear logistic regression method, because it has the tendency towards lowering model bias at the expense of higher variance. The high variance models produce an overfit to the existing data and thus lower the generalization of the predictions to new data. Because large number of highly correlated predictors were present in our analysis, we decided to use a logistic ridge regression model instead. We used the current level of the adult football players (adult elite vs. adult professional) as a response variable (discrete outcome) and their EF components measured with NeurOlympics game as predictors (continuous variables). The "regularized" regression technique introduces some bias in the model, but it effectively reduces variance and thus allow us to prevent overfitting. By penalizing correlated regression coefficients, the ridge regression method optimizes the bias-variance trade-off, avoids overfitting, and increases the predictive ability of our model.

The dataset included imbalanced classes with 70% professional and 30% elite data points. To handle this issue, we introduced class membership weights for each data point that was used during cross-validation and the fitting of the model. The weights were calculated as follows: the ratio of number of members for each class and the total number of members (observations) subtracted from the total number of observations in the dataset. Thus, professionals got a weight of 0.3, and elites got a weight of 0.7, effectively downweighing the influence of the larger group in the data analysis.

For the analysis, we used R version 3.6.3 (2020-02-29) and glmnet (Hastie & Qian, 2016; https://cran.rproject.org/web/packages/glmnet/index.html; https://d.docs.live.net/e19a8dc52966804f/Documents/BrainsFirst/) and caret packages (Kuhn, 2008; https://CRAN.R-project.org/package=caret). We randomized row order in the dataset before splitting it into a training (80%) and testing (20%) set. To optimize model performance, we checked whether any of the variables had zero variance, before including them as predictors in the model. By default, the glmnet package standardizes the predictor variables by centering them around zero. This minimizes penalizing of the intercept coefficient. After that, we applied 10-fold cross-validation to the training dataset for each ridge regression regularization parameter to choose the best one which would minimize the cost function of the model. We applied this parameter and the corresponding coefficients, estimated from the training dataset, to a separate test data to assess model's ability to successfully differentiate elite and professional players based on measured EFs. We repeated this procedure 1000 times. Therefore, data points were randomly assigned to train and test data for each iteration. We saved the 1000 fitted models and their performance metrics.

Predictive performance

To assess the model's performance we calculated sensitivity, specificity, accuracy and F1 score (Altman & Bland, 1994a, 1994b). In the model, elite was treated as a positive class and professional – as a negative class. F1 score is a harmonic mean of precision and recall: F1-score= 2 * Precision * Recall/(Precision + Recall). Precision shows the proportion of true positives (elite players) from all the predicted positives by the model, and recall, also called the sensitivity, gives the proportion of true positives out of the actual positives in the dataset. Because our dataset had imbalanced classes and we only had two classes (binary classification), we used Matthews Correlation Coefficient (MCC) as a primary indicator of model's performance (Chicco & Jurman, 2020; Powers, 2011). MCC measures the correlation between predicted and actual classes and is interpreted as the Pearson product moment correlation coefficient with the range of -1 and 1 (Field, Miles, & Field, 2012). The value of 1 is

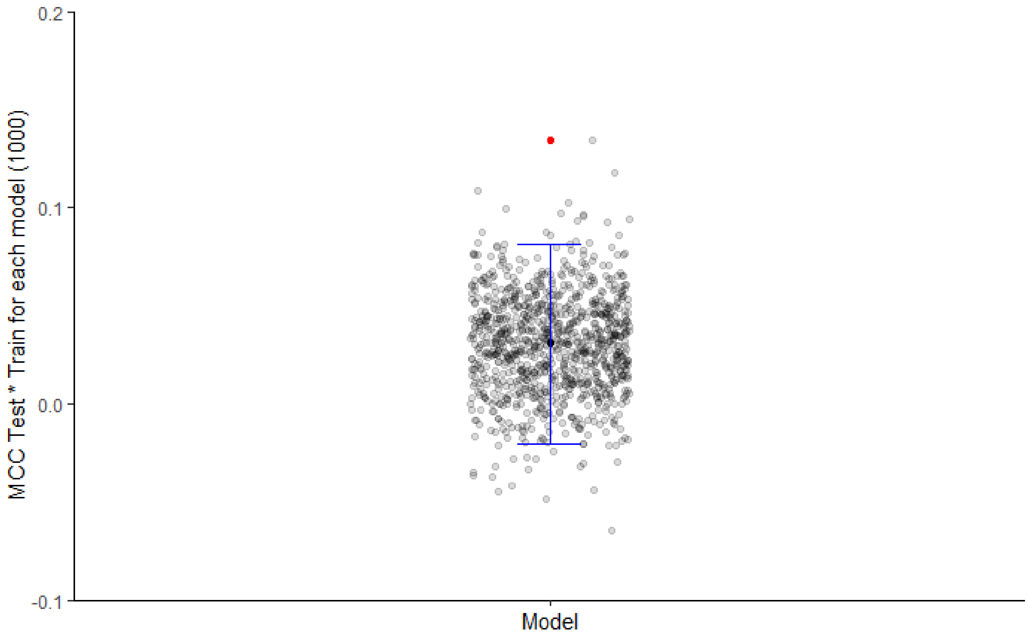


Figure 3. MCC test * MCC train for each model depicted with standard deviation (error bars) and best model (in red)

indicative of perfect classification and the value of -1 indicates the absolute misclassification, while MCC = 0 indicates model's performance at chance level (Chicco & Jurman, 2020; Laerd Statistics, 2020). Unlike model accuracy, MCC is insensitive to class imbalance and is not affected by the order of positive and negative classes in the model. MCC, calculated for train and test data sets, was used to select the best model from 1000 generated models (Figure 3). We selected a model with the maximum value of the multiplication product of train and test MCC values (max (trainMCC * testMCC)). By using the maximum value of the product, we select the best possible model with the balance between performance on a train dataset and generalizability on the test data.

Statistical analysis

The fitted model was applied to the whole dataset to calculate class probabilities as model scores. Low scores (less than 0.5) indicated elite players and higher scores indicated professional players. Pearson's r was calculated to test the correlation between the model scores and market values of the players. Finally, we selected data points with model scores above and below of two standard deviations and took the subjects with these scores as the prototypical professional and elite players and plotted their EF components against each other for comparison.

Results

For the final analysis we selected 76 elite (160 observations) and 153 professional players (373 observations). Each test session was treated in the analysis as a separate data point that resulted in 533 observations in total. Mean current age of included players was 22.34 (SD 3.22), mean age during the test was 19.25 (SD 3.84). As for the predictor variables, after checking for zero variance, three variables were removed: working memory composure, filtering and, "control_accuracy_repeat_4_trials" which includes raw scores measuring the accuracy in no switch condition of Connect game, and is the subcomponent of the automatic control variable. 66 predictor variables were included in the analysis.

Elite football brain model

The main question was whether we could create a model which would allow for accurate prediction of the current football level - adult elite or adult professional, based only on the historical cognitive data. To create the elite football brain model, we used logistic ridge regression algorithm with the football players' current performance level – adult elite or adult professional, as an outcome variable and EFs previously measured on NeurOlympics games, as predictors.

The first model was created from the entire data set, including players who were either adults or youth players when their cognitive data was collected. We used 10-fold cross-validation for choosing the best regularization parameter before fitting the model, and this procedure was run 1000 times to allow random grouping of data points into train and test datasets. The best model was chosen based on maximum MCC for both train and test sets (Figure 4). To be able to compare to the other performance metrics of the model (i.e., accuracy), MCC was normalized (normMCC = (MCC + 1) / 2). Normalized MCC has a range from 0 to 1 instead of -1 to 1, where value of 0.5 is indicative of a random classification and the closer the value gets to 1 the better the performance of a model (Chicco & Jurman, 2020). We were primarily interested to see how the model generalizes to the new (test) data rather than the performance of the model on the training dataset. Therefore, we decided to present the performance metrics like MCC in the results only for the test data as it is informative of the model's predictive ability. The correlation as calculated with Matthews Correlation Coefficient, between actual and predicted classes on test dataset for this model was significant (*MCC* = .38, *p* <.001).

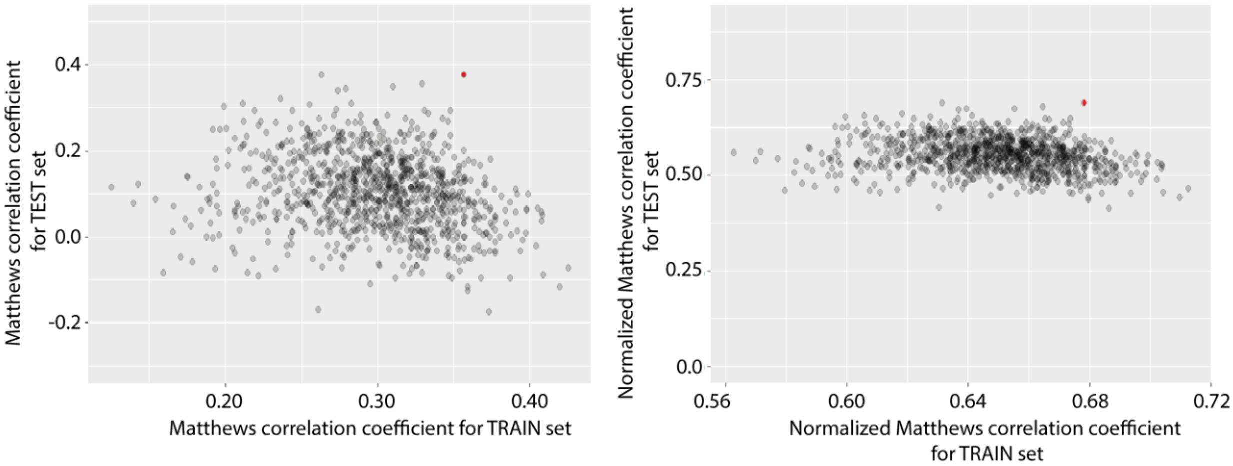


Figure 4. Elite football brain model: MCC and normMCC for test and train datasets depicted for all 1000 models. Best model, shown in red, has maximum value of test MCC multiplied with train MCC

Elite football brain model with session weights

After creating the first model, we performed several additional steps to test whether initial model could be improved. Except for the additional steps, the remaining procedure was the same (10-fold cross-validation, repeated 1000 times: Figure 5). We compared performance across models based on the values of MCC (normMCC), the higher the value the better the performance (Figure 6).

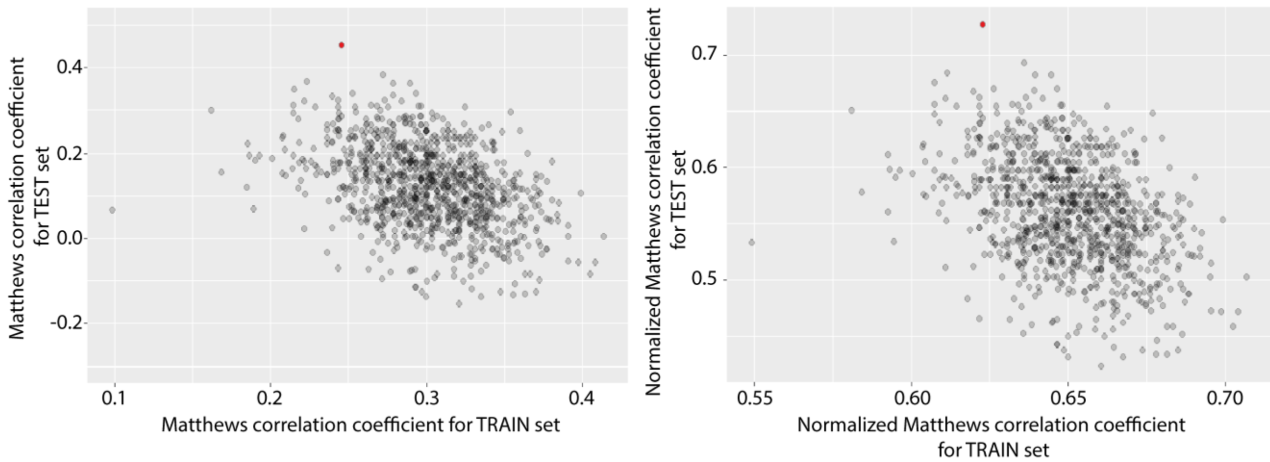


Figure 5. Elite football brain model with session weights: MCC and normMCC for test and train datasets depicted for all 1000 models. Best model, shown in red, has maximum value of test MCC multiplied with train MCC

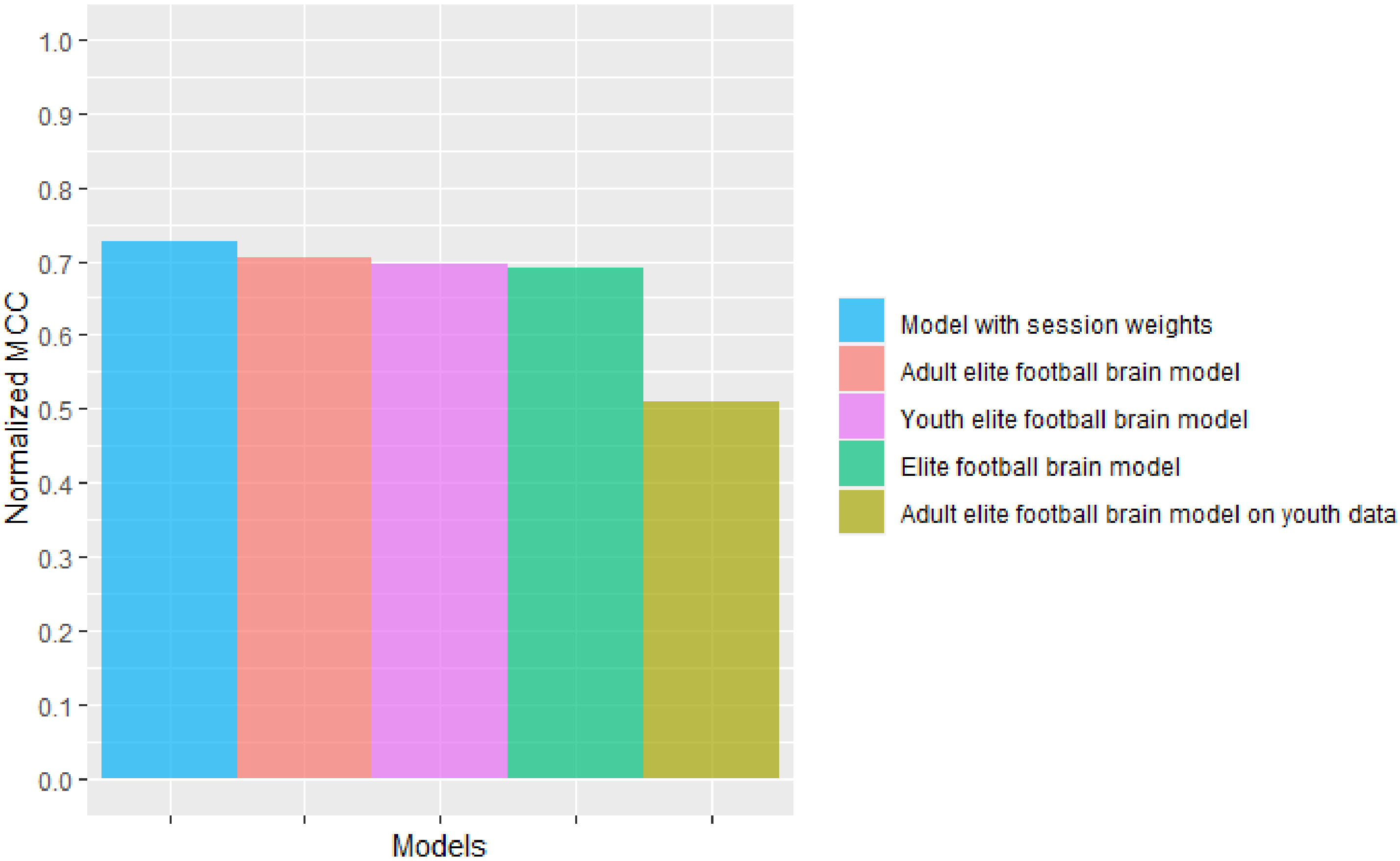


Figure 6. Comparison of performance of different models based on normalized MCC

The number of test sessions (data points) varied between players, ranging from 1 to 7. Therefore, to account for the variance between players we decided to add different weights to each data point based on the number of test sessions. The weights were calculated as a ratio of number of sessions and maximum number of sessions, for example players with maximum number of test sessions get a weight of 1 whereas players with only one session get a weight of 1/7 = 0.143. These weights were multiplied by class membership weights and were included during cross-validation and fitting of the model. The model with session weights performed better on new dataset than the first model with only class membership weights and showed higher correlation between actual and predicted classes (MCC = .45, p < .001), although the improvement was not statistically significant (p = .272).

Elite football brain scores and market value

Model scores calculated from the class probabilities showed a significant correlation with market value (r = -.29, p < .001). This indicates that elite football brain scores (lower scores) correlate with higher market values (Figure 7). For prototypical elite and pro EF components see, Figure 8.

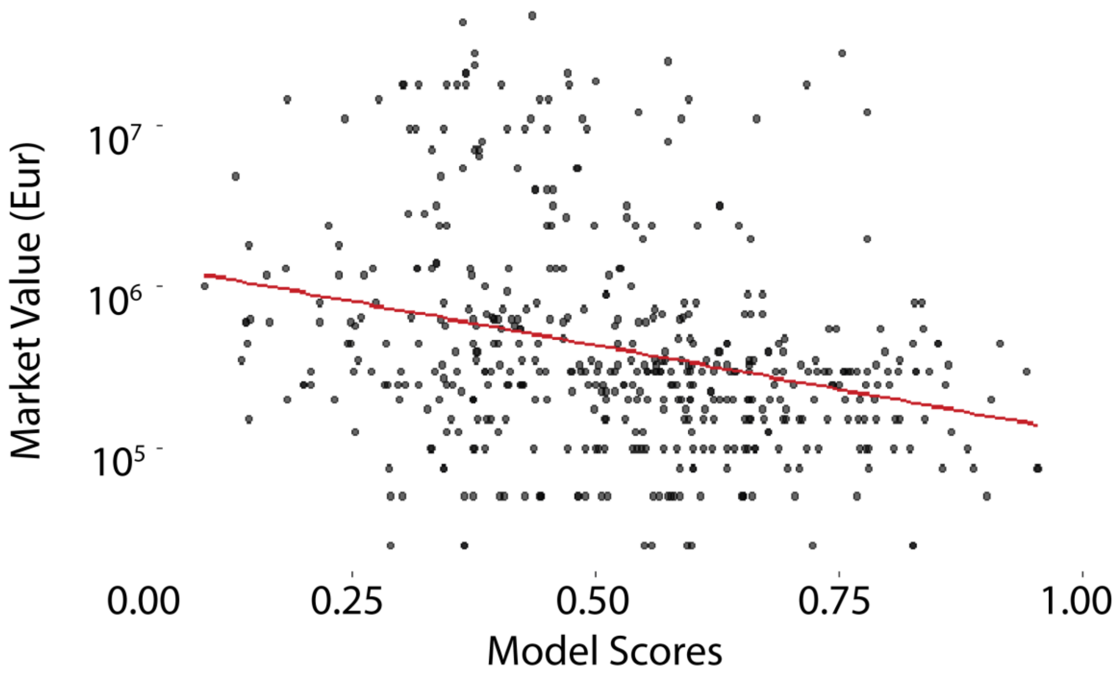


Figure 7. Correlation of elite football brain model scores and market value (log-transformed)

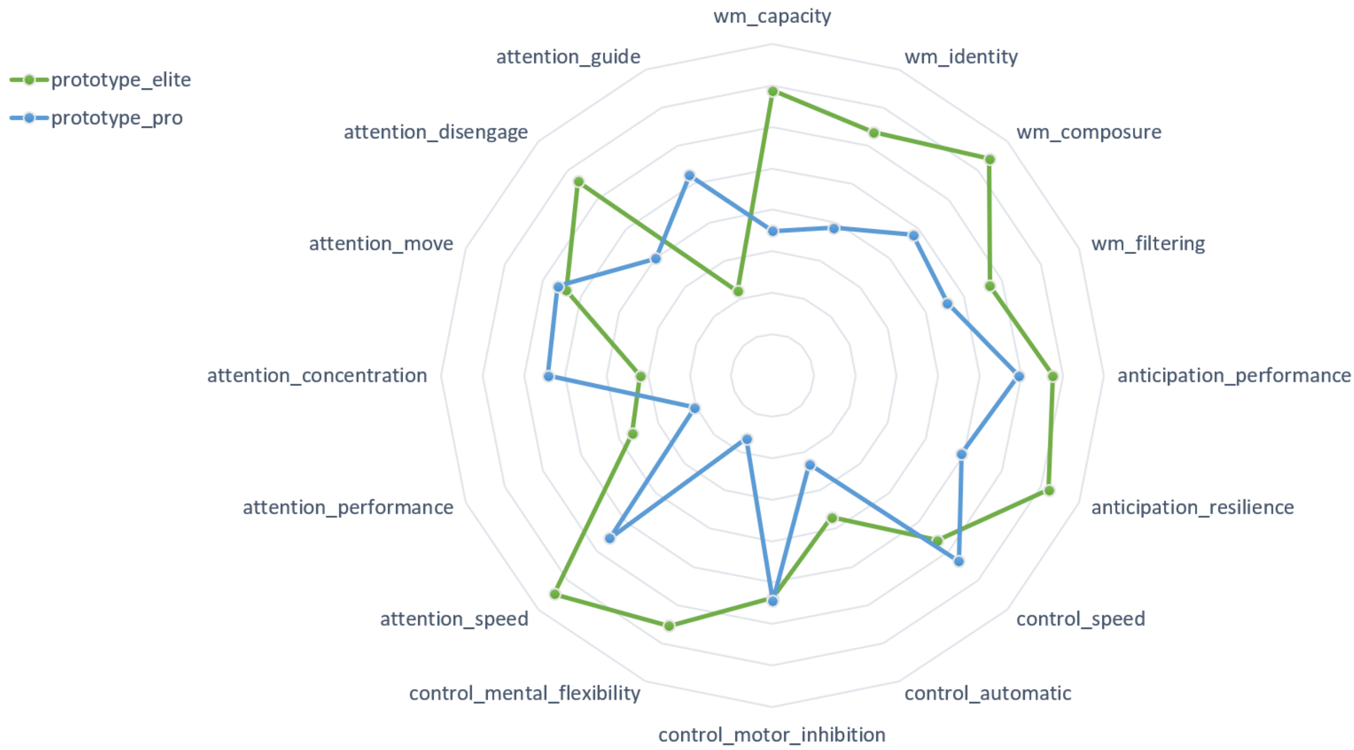


Figure 8. EF components, measured by NeurOlympics game, possessed by prototypical elite and pro football players. Prototypical elite was defined as average score of subjects scoring more than 2 SD below the mean: < .18; prototypical pro - above 2 SD: > .86 (13 data points in each group). We observed that elite players have superior scores on all working memory (wm) and anticipation components as well as on mental flexibility, attention speed, performance, concentration and

disengage. Professional players appear to outperform elites on the attention concentration and guide components

Exploratory analysis

Adult elite football brain model applied on youth data

Our dataset contained players whose cognitive data was collected during their youth, when the EFs are not yet completely matured. Therefore, it was logical to divide the data into players who were youth (272 data points) during the cognitive test and players who were already adults (261 data points) during the cognitive test. And to then train the model using only adult cognitive data to investigate how the model would generalize to youth cognitive data. In other words, can we successfully predict whether youth players would become adult elites or pros when the model is trained on adult cognitive data only. The cross-validation and fitting of the model was run only once since the train and test dataset were fixed in this case (adults and youth). The predictive performance of the model on youth data was almost at chance level with *MCC* = .02, *p* = .743. This implies that the cognitive profile distinguishing between elite vs. pro is not the same for youth cognitive data and adult cognitive data.

As a next step, we identified age as a confounding factor and tested whether controlling for it would improve the model. We removed the age effect by fitting a 4th order polynomial model to the whole data set and converting EF scores accordingly. This resulted in EF scores that are based on relative performance with respect to each age group. Even after controlling for the age effect the model did not show any improvement (*MCC* = .04, *p* = .256). This implies that the difference between adult cognitive data and youth cognitive data is driven by a non-linear difference between groups.

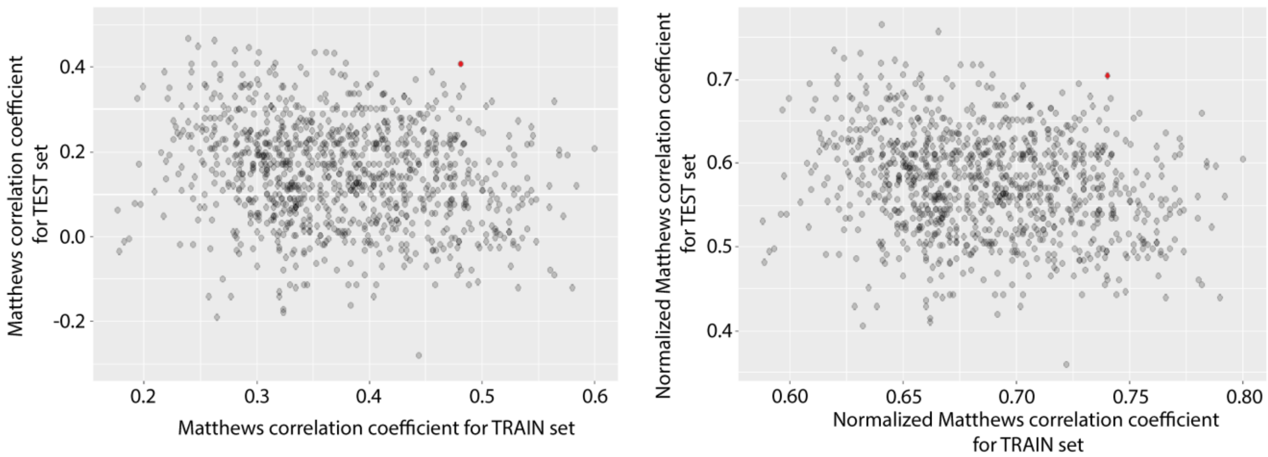


Figure 9. Elite football brain model with adult data: MCC and normMCC for test and train datasets depicted for all 1000 models. Best model, shown in red, has maximum value of test MCC multiplied with train MCC

Elite football brain model for adult and youth data separately

Finally, we tested whether the performance of the model on the whole dataset is driven only by the data of adults when being tested on cognition, or whether the model can successfully predict football performance level when the data contains only youth or adult cognitive variables as predictors. Again, we divided the data into youth and adults when the cognitive test was administered and applied the same logistic ridge regression for each group separately, with 1000 times iteration each (Figure 9 and

Figure 10. Elite football brain model with youth data**:** MCC and normMCC for test and train datasets depicted for all 1000 models. Best model, shown in red, has maximum value of test MCC multiplied with train MCC

10). The performance of these models on the new dataset was comparable to the model created on the whole data (for adults: *MCC* = .41, $p < .001$ and for youth: *MCC* = .39, $p < .001$). This indicates that the model can successfully classify elite and professional players on adult data as well as predict future football level in players whose data were collected while being youth. There was a significant correlation between market value and model scores calculated as class probabilities for adults ($r = -.22$, $p < .001$) and youth ($r = -.28$, $p < .001$) (Figure 11). Additionally, we plotted prototypical elite and pro EFs separately for youth and adults (Figure 12) and compared EFs of prototypical youth and adult elites (Figure 13). For additional performance metrics calculated across all models, see Table 1.

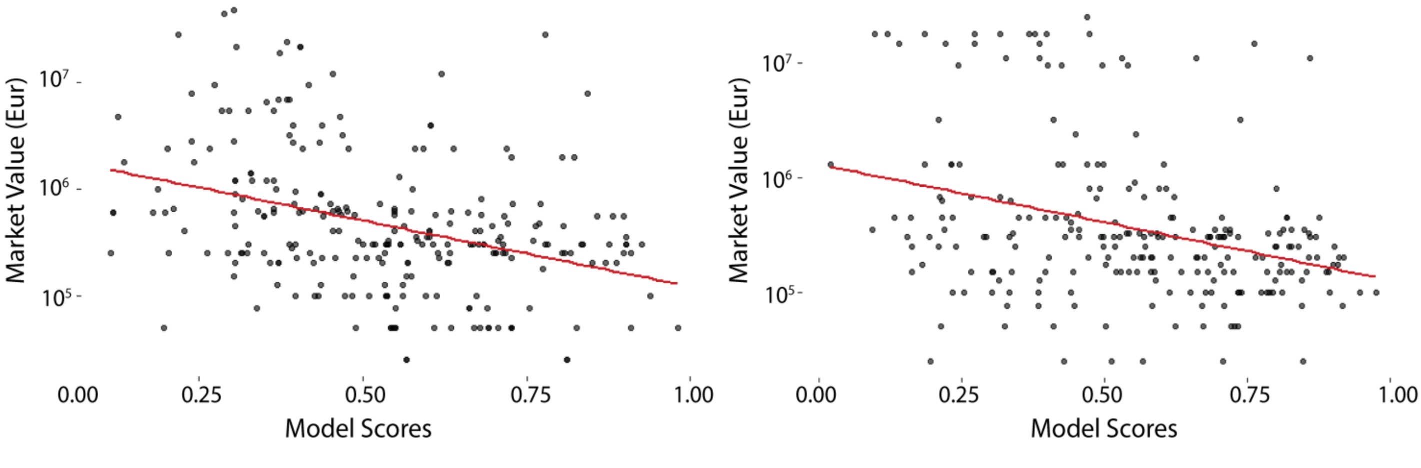
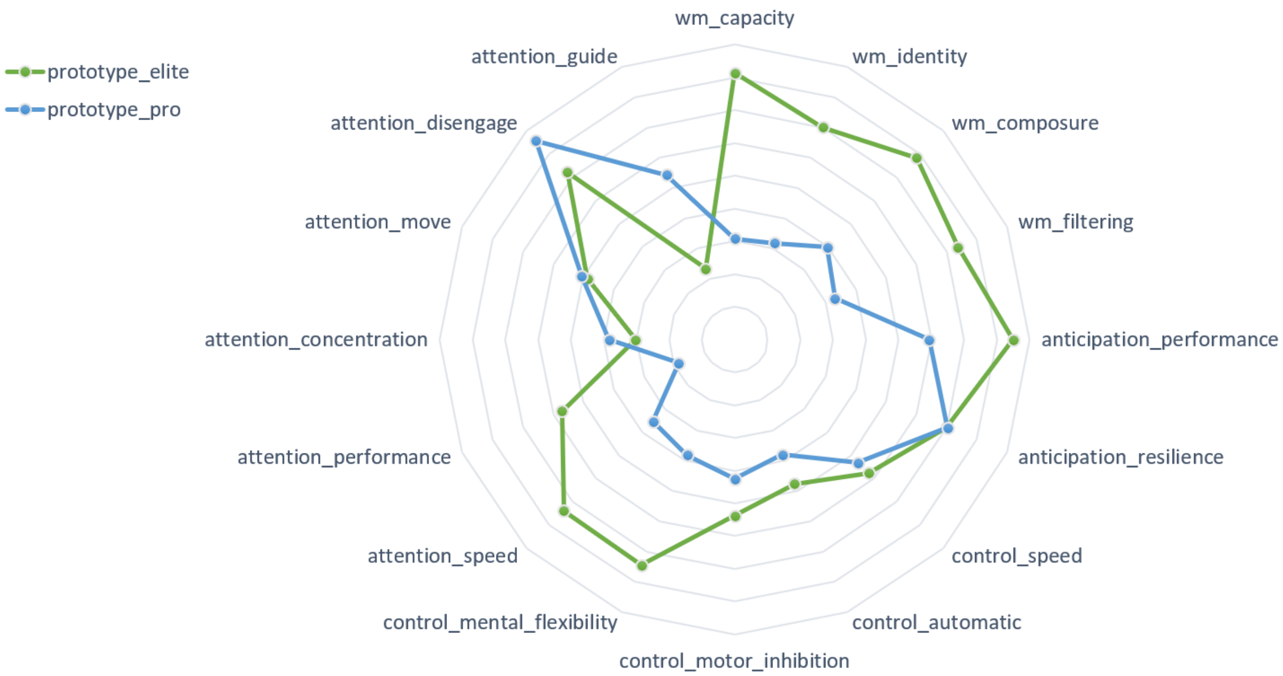


Figure 11. Correlation of elite football brain model scores and market value (log-transformed) in adults **(left)** and youth **(right)** at the test administration

A



B

Figure 12. EF components measured by NeurOlympics game, possessed by prototypical elite and pro adults **(A)** and youth **(B)**. Prototypical adult elite was defined as average score of subjects scoring more than 2 SD below the mean: < .18; prototypical adult pro - above 2 SD: > .90 (elite n = 6, pro n = 7); prototypical youth elite model score < .16 and prototypical youth pro model score: > .91 (elite n = 7, pro n = 7). While adult elite players are observed to score higher than adult pros across most of the tested EFs, with some exceptions of anticipation resilience and attention subcomponents, youth elites and pros do not show the similar degree of contrast. Youth elites only differ from youth pros on working memory identity, attention move, disengage, and guide

Figure 13. EF components measured by NeurOlympics game, possessed by prototypical adult and youth elites. Youth elites show better scores than adult elites only in few EF subcomponents: control speed, attention speed and concentration. Other EF components are either comparable between adult and youth elites, or adults show the superiority

Table 1. Performance metrics of models on test set: accuracy, sensitivity, specificity, F1-score and MCC

Model	*accuracy*	*sensitivity*	*specificity*	*F1-score*	*MCC (normMCC)*
Elite football brain model	.70	.72	.69	.59	.38 (.69)
Elite football brain model with session weights	.77	.59	.85	.61	.45 (.73)
Adult elite football brain model on youth	.57	.40	.62	.31	.02 (.51)
Adult elite football brain model on youth age corrected	.51	.54	.51	.35	.04 (.52)
Elite football brain model with adult data	.73	.58	.82	.61	.41 (.70)
Elite football brain model with youth data	.70	.77	.68	.56	.39 (.70)

Discussion

The purpose of this study was to investigate whether cognitive functions can predict whether a football player would reach elite-level or professional-level football performance 0 to 6 years after cognitive test administration. Moreover, to assess whether players that were classified as elite-level football players had a more successful football career, their elite football brain scores were correlated to their current market value. We hypothesised that it would be possible based on historical EF data to classify the player as a future elite or professional, and that the generated models would show a modest correlation with their present market value. Furthermore, we explored the generalizability of the model to youth data when being trained only using the adult EFs as predictors. Finally, we tested the predictive capacity of the elite football brain model trained on the youth and adult data separately and compared the EF profiles for players classified as prototypical elites and pros.

In line with our hypothesis, the elite football brain model proved successful in differentiating between football levels, based solely on the players' cognitive performance. Therefore, this signals that executive functions could serve as additional criteria for the future talent identification. Even though statistically significant, the strength of the correlation between actual and predicted classes was moderate. It was expected that the models' performance would not be higher than that, because it was built with only cognitive variables as predictors of football performance. However, cognitive function, albeit important, is only a part of many other factors that, taken together, determine successful football career. For instance, an athlete's psychological and physiological profiles as well as tactical and technical skills and motivation are shown to be important aspects defining current and future performance levels in sport (Abade et al., 2014; Murr et al., 2018; Sarmento et al., 2018; Williams & Reilly, 2000). We argue that the predictive ability of the model could be improved by adding these factors along with EFs and future studies should pursue this possibility. It is important to note that because we used several data points for each player, the performance of the model could be inflated, and these results should be interpreted with this limitation in mind. The alternative approach would have been to average cognitive scores of each player, resulting in unique observations per player. However, this solution is suitable only with large datasets which was not the case for us. For future studies with limited data size, a possible solution would be to ensure that all the data points from the same player are grouped together in either train, validation, or test data sets, rather than assigning data points to these datasets at random. This would provide a clearer view on the generalizability of the model.

As predicted, we observed a modest correlation between elite-level model scores and the market value of football players. We deemed market value to be an objective and personal parameter of the football success level. In other words, taking a football club rank, or league level as a marker of successful career, would group players with different performance level together, while the market value gives us the opportunity to avoid this variability between players and to assess the individual performance level. Thus, the positive relationship between a higher market value and elite-level model scores indicates the usefulness of the model to identify and assign elite class to the individual players who objectively have top-level career. Moreover, these findings add to the growing body of research suggesting that superior executive functions are important contributors in elite-level sport performance (Scharfen & Memmert, 2019a; Voss et al., 2010). Both Vestberg et al. (2017) and Huijgen et al. (2015) showed that youth elite players possessed higher cognitive capabilities and outperformed lower-level football players (Huijgen et al., 2015; Vestberg et al., 2017). Our findings suggest that the difference between elite-level and professional-level EFs can be used as a classifier to differentiate adult elite from professional players. More importantly, we were able to show that based on this difference, not only current but future football level can be predicted, years after the cognitive data was collected.

The elite football brain model was able to accurately classify 72% of elite players in the new (test) data set. After controlling for the variance created by the different number of test sessions between players, we managed to improve the correlation strength between the actual and predicted classes (correlation coefficient from .38 to .45). However, this attempt decreased model's sensitivity from .72 to .59. It is known that only a small fraction (less than 5 percent) of young players end up performing at the elite level, therefore it is crucial to develop reliable criteria to identify "hidden" talents from an early age (Hol & Wolfert, 2011). Using EFs for this purpose provides an additional opportunity to accurately select talented youth players when other criteria fail to be informative. Therefore, sensitivity is an important metric for our model because we are primarily interested in correctly identifying as many elite-level players as possible based on their EFs. Although the performance metric such as MCC indicates better performance of the model with session weights, we argue that the first model with higher sensitivity proves to be more useful when the goal is to identify elite-level players. On the other hand, the elite football brain model with session weights could correctly retrieve from test data 85% of players who were not elites (pros). The high specificity of this model could serve as a confirmatory tool in screening out youth players who do not show elite-level qualities.

The elite football brain model built with adult EFs, could not distinguish between future elites and pros when tested on youth data, with the performance at chance level. This may be indicative that the mature, adult EFs do not generalize well to youth EF data. The earlier research has shown that different EFs reach their peak of development at different phase of adolescence (Anderson et al., 2001; Luciana et al., 2005; Verburgh, Königs, Scherder, & Oosterlaan, 2014). Thus, there can be an inherent difference between the EFs important for adult elite-level players and the EFs necessary for young players who are still at the stage of development. Another explanation of this finding could be the nonlinear relationship between adult and youth EFs necessary for top-level football which could not be captured by the linear machine learning algorithm applied here. More complex models should be employed by future studies to explore whether the translation between adult and youth EFs is possible.

Interestingly, the elite football brain model created from only the adult data could successfully classify adult elite and professional players. Similarly, a model built with the youth data could predict current football level of adult players based on their youth cognitive variables, recorded several years ago. However, when prototypical adult and youth elite and pro EFs were inspected visually, the cognitive profiles showed different patterns. We observed that adult elites show superior EFs across almost all EF components, while the youth elite and pro EF profiles are not that drastically different from each other (Figure 12 and 13). Thus, the fully developed adult elite and pro-level EFs are clearly demarcated from each other, while still maturing youth EFs make the distinction between elite-level from pro-level EFs more subtle. This effect can be explained by the fact that fully developed adult elite EFs are showing their full potential, while youth EFs have different maturation rates (Beavan, Spielmann, Mayer, et al., 2019) and the effect could be diminished because different age groups are included in youth data. Therefore, predictive models built with youth cognitive variables can be more informative if different age groups are considered separately.

Furthermore, we can argue that EFs responsible for the high-level performance during the adulthood differs from the EFs that define the development of youth into high -level adult athletes. It has been proposed that the development of EFs are associated with the amount and intensity of practice in sport even throughout the adult life (Ericsson, Krampe, & Tesch-Romer, 1993; Karbach & Unger, 2014). The players who were adult elites or professionals while taking the cognitive tests, could have had very distinct exercise regimes and intensity according to their performance level and thus showed very different EF profiles. In contrast, youth players taking the tests were not yet developed into elite or professional football players and in youth academies their training quality was mostly similar. Therefore, their EFs showed to be less contrasting than those of the adults. On the other hand, there are studies suggesting the genetic predisposition for the superior EFs (Friedman et al., 2008). Thus, it is possible that despite the homogeneity of the training regime among young players, there were some genetically determined individual differences in cognitive performance that could be picked up by the classifier. Future studies with bigger sample size in each age group could explore subtle differences between still developing EFs. Furthermore, the overall developmental trajectory of each EF component could be tracked over the years from childhood to adulthood. As a next step, it would be important to assess the predictive value of each EF component for each age group to see which cognitive variables are prominent in the development of elite-level players.

In conclusion, we showed that future football performance level can be predicted based on cognitive functions. While the predictive value is similar for youth cognitive data and adult cognitive data, different cognitive functions seem to be important in predicting future performance when youth cognitive data compared to adult cognitive data are used. Overall, these findings corroborate the previous work that elite and professional-level EFs are indeed different from each other and as an additional step, we demonstrate that players can be successfully classified based on these differences. Therefore, EFs can be the additional valid decisive factors for the selection and development of elite-level talents.

Acknowledgments

I want to thank my supervisor dr. Ilja Sligte for giving me the opportunity to do my masters' research project at BrainsFirst. I am grateful for his guidance throughout the project and for the much-needed support during the lockdown period. I also want to express my gratitude towards the team of BrainsFirst for their warm welcome. Lastly, special thanks to Rianne Lam for her assistance and insights during the data analysis and for the helpful comments on the earlier version of this report.

References

Abade, E. A., Gonçalves, B. V., Silva, A. M., Leite, N. M., Castagna, C., & Sampaio, J. E. (2014). Classifying young soccer players by training performances. *Perceptual and Motor Skills*. https://doi.org/10.2466/10.25.PMS.119c31z8

Altman, D. G., & Bland, J. M. (1994a). Statistics Notes: Diagnostic tests 1: Sensitivity and specificity. *BMJ*, *308*(6943), 1552. https://doi.org/10.1136/bmj.308.6943.1552

Altman, D. G., & Bland, J. M. (1994b). Statistics Notes: Diagnostic tests 2: Predictive values. *BMJ*, *309*(6947), 102. https://doi.org/10.1136/bmj.309.6947.102

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental Neuropsychology*. https://doi.org/10.1207/S15326942DN2001_5

Beavan, A. F., Spielmann, J., & Mayer, J. (2019). Taking the First Steps Toward Integrating Testing and Training Cognitive Abilities Within High-Performance Athletes; Insights From a Professional German Football Club. *Frontiers in Psychology*, *10*(December), 1–5. https://doi.org/10.3389/fpsyg.2019.02773

Beavan, A. F., Spielmann, J., Mayer, J., Skorski, S., Meyer, T., & Fransen, J. (2019). Age-Related Differences in Executive Functions Within High-Level Youth Soccer Players. *Brazilian Journal of Motor Behavior*, *13*(2), 64–75. https://doi.org/10.20338/bjmb.v13i2.131

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 6. https://doi.org/10.1186/s12864-019-6413-7

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Ericsson, K. A., Krampe, T. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100, 363–406.*

Fan, J., Gu, X., Guise, K. G., Liu, X., Fossella, J., Wang, H., & Posner, M. I. (2009). Testing the behavioral interaction and integration of attentional networks. *Brain and Cognition*, *70*(2), 209–220. https://doi.org/10.1016/j.bandc.2009.02.002

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *NeuroImage*, *26*(2), 471–479. https://doi.org/10.1016/j.neuroimage.2005.02.004

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. *SAGE Publications Ltd* (Vol. 50). https://doi.org/10.5860/choice.50-2114

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual Differences in Executive Functions Are Almost Entirely Genetic in Origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Gold, J. M., Fuller, R. L., Robinson, B. M., Mcmahon, R. P., Braun, E. L., & Luck, S. J. (2006). Intact Attentional Control of Working Memory Encoding in Schizophrenia. *Journal of Abnormal Psychology*, *115*(4), 658. https://doi.org/10.1037/0021-843X.115.4.658

Hastie, T., & Qian, J. (2016). Glmnet Vignette. *Retrieved from Https://Web.Stanford.Edu/~hastie/Glmnet/Glmnet_alpha.Html*.

Hol, F., & Wolfert, R. (2011). Dromen in Duigen. *Elsevier. Retrieved from Https://Ziladoc.Com/Download/Dromen-in-Duigen_pdf*.

Huijgen, B. C. H., Leemhuis, S., Kok, N. M., Verburgh, L., Oosterlaan, J., Elferink-Gemser, M. T., & Visscher, C. (2015). Cognitive functions in elite and sub-elite youth soccer players aged 13 to 17 years. *PLoS ONE*, *10*(12), 1–13. https://doi.org/10.1371/journal.pone.0144580

Isoda, M., & Hikosaka, O. (2007). Switching from automatic to controlled action by monkey medial frontal cortex. *Nature Neuroscience*, *10*(2), 240–248. https://doi.org/10.1038/nn1830

Johnson, M. K., McMahon, R. P., Robinson, B. M., Harvey, A. N., Hahn, B., Leonard, C. J., … Gold, J. M. (2013). The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. *Neuropsychology*, *27*(2), 220–229. https://doi.org/10.1037/a0032060

Karbach, J., & Unger, K. (2014). Executive control training from middle childhood to adolescence. *Frontiers in Psychology*. Frontiers Research Foundation. https://doi.org/10.3389/fpsyg.2014.00390

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5). Retrieved from http://www.jstatsoft.org/

KyllingsbæK, S., & Bundesen, C. (2009). Changing change detection: Improving the reliability of measures of visual short-term memory capacity. *Psychonomic Bulletin and Review*, *16*(6), 1000–1010. https://doi.org/10.3758/PBR.16.6.1000

Laerd Statistics. (2020). Pearson's product moment correlation. Statistical tutorials and software guides. Retrieved July, 23, 2020, from https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php.

Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The Development of nonverbal working memory and executive control processes in adolescents. *Child Development*, *76*(3), 697–712. https://doi.org/10.1111/j.1467-8624.2005.00872.x

McNab, F., & Dolan, R. J. (2014). Dissociating distractor-filtering at encoding and during maintenance. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 960–967. https://doi.org/10.1037/a0036013

McNab, F., Zeidman, P., Rutledge, R. B., Smittenaar, P., Brown, H. R., Adams, R. A., & Dolan, R. J. (2015). Age-related changes in working memory and the ability to ignore distraction. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(20), 6515–6518. https://doi.org/10.1073/pnas.1504162112

Murr, D., Raabe, J., & Höner, O. (2018). The prognostic value of physiological and physical characteristics in youth soccer: A systematic review. *European Journal of Sport Science*. https://doi.org/10.1080/17461391.2017.1386719

Neubert, F. X., Mars, R. B., Buch, E. R., Olivier, E., & Rushworth, M. F. S. (2010). Cortical and subcortical interactions during action reprogramming and their related white matter pathways. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(30), 13240–13245. https://doi.org/10.1073/pnas.1000674107

Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation. *Journal of Machine Learning Technology*, *2*(1), 37–63.

Sakamoto, S., Takeuchi, H., Ihara, N., Ligao, B., & Suzukawa, K. (2018). Possible requirement of executive functions for high performance in soccer. *PLoS ONE*, *13*(8), 1–11. https://doi.org/10.1371/journal.pone.0201871

Sarmento, H., Anguera, M. T., Pereira, A., & Araújo, D. (2018). Talent Identification and Development in Male Football: A Systematic Review. *Sports Medicine*, *48*(4), 907–931. https://doi.org/10.1007/s40279-017-0851-7

Scharfen, H. E., & Memmert, D. (2019a). Measurement of cognitive functions in experts and elite athletes: A meta-analytic review. *Applied Cognitive Psychology*, *33*(5), 843–860. https://doi.org/10.1002/acp.3526

Scharfen, H. E., & Memmert, D. (2019b). The relationship between cognitive functions and sport-specific motor skills in elite youth soccer players. *Frontiers in Psychology*, *10*(APR), 1–10.

https://doi.org/10.3389/fpsyg.2019.00817

Verburgh, L., Königs, M., Scherder, E. J. A., & Oosterlaan, J. (2014). Physical exercise and executive functions in preadolescent children, adolescents and young adults: A meta-analysis. *British Journal of Sports Medicine*, *48*(12), 973–979. https://doi.org/10.1136/bjsports-2012-091441

Verburgh, L., Scherder, E. J. A., Van Lange, P. A. M., & Oosterlaan, J. (2014). Executive functioning in highly talented soccer players. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0091254

Vestberg, T., Gustafson, R., Maurex, L., Ingvar, M., & Petrovic, P. (2012). Executive functions predict the success of top-soccer players. *PLoS ONE*, *7*(4), 1–5. https://doi.org/10.1371/journal.pone.0034731

Vestberg, T., Reinebo, G., Maurex, L., Ingvar, M., & Petrovic, P. (2017). Core executive functions are associated with success in young elite soccer players. *PLoS ONE*, *12*(2), 1–13. https://doi.org/10.1371/journal.pone.0170845

Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S., & Roberts, B. (2010). Are expert athletes “expert” in the cognitive laboratory? A meta-analytic review of cognition and sport expertise. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.1588

Williams, A. M., & Reilly, T. (2000). Talent identification and development in soccer. *Journal of Sports Sciences*, *18*(9), 657–667. https://doi.org/10.1080/02640410050120041

Xie, W., Li, H., Ying, X., Zhu, S., Fu, R., Zou, Y., & Cui, Y. (2017). Affective bias in visual working memory is associated with capacity. *Cognition and Emotion*, *31*(7), 1345–1360. https://doi.org/10.1080/02699931.2016.1223020