



# Prediction of athletes performance using neural networks: An application in cricket team selection

Subramanian Rama Iyer \*, Ramesh Sharda

*Institute for Research in Information Systems, Department of Management Science and Information Systems, Spears School of Business, Oklahoma State University, Stillwater, OK 74078, United States*

## ARTICLE INFO

### Keywords:

Neural networks  
Sports  
Cricket  
Team-selection  
Decision support

## ABSTRACT

Team selection for international sports competitions requires predicting performance of individual athletes. We explore the use of neural networks to rate players and select specific players for a competition. We take cricket as an example. Cricket is a game with mass following in British Commonwealth Countries as well as some other countries. National teams visit other countries for bilateral matches as well as play in World Cup tournaments. We employ neural networks to predict each cricketer's performance in the future based upon their past performance. We classify cricketers into three categories – performer, moderate and failure. We collected data on cumulative player performance from 1985 onwards until the 2006–2007 season. The neural network models were progressively trained and tested using four sets of data. The trained neural network models were then applied to generate a forecast of the cricketer's near term performance. Based on the ratings generated and by applying heuristic rules we recommend cricketers to be included in the World Cup 2007. We evaluate the actual performance of the cricketers in the World Cup to validate the applicability of neural networks. The results show that the neural networks can indeed provide valuable decision support in a team selection process.

© 2008 Elsevier Ltd. All rights reserved.

## 1. Introduction

Predicting an individual athlete's performance based upon his/her past record can be critical in the selection of team members in international competitions. This process is highly subjective and usually requires much expertise and negotiative decision making. In this study, we explore the use of advanced non-linear modeling techniques such as neural networks to provide an analytical aid in such decision situations. We take the case of team selection in international contests in cricket.

In a team selection committee environment multiple members evaluate each player's performance and vote for inclusion/exclusion from the team. These selection committee members provide rankings for cricketers. Negotiations are then conducted to produce an agreement among the selection committee members as to which cricketer should be finally recommended to be selected to the World Cup team. We simulate this process by forecasting a player's performance using different models.

Cricket is a popular game played by a few countries (see Table 1 – Prominent Cricket Playing Nations). There are two versions of the game – Test Cricket which is played over five days and one-day cricket, which is obviously played over a day (Amy, 2007). One-

day cricket was introduced in the English domestic season of 1963 due to the growing demand for a shorter and more dramatic form of cricket to stem the decline in attendance. One-day, single-innings matches often took place before this, but the innovation was the limiting of each side's innings to an agreed number of overs (nowadays usually 50). The inaugural 1975 World Cup was a great success. The abbreviations ODI (One-day International) or sometimes LOI (Limited Overs International) are used for international matches of this type. Frequent nail-biting finishes and the impossibility of either side opting to play for a draw have seen ODI cricket gain many supporters (Beaudoin, 2003).

Often a Test series between nations has occupied the limelight with names of former Cricketers being attached to the matches. For example, The Ashes is a bi-annual contest, and Border – Gavaskar Trophy – is awarded to the winner of the official Test matches played between India and Australia. Test cricket is played over five days, with three two-hour sessions per day. Sessions are usually interspersed with a 40-min break for lunch and 20-min break for afternoon tea. Each team has eleven players.

The popularity of cricket has increased tremendously over the last 10 years due to the influx of television and cable networks broadcast of cricket matches all over the world. The cricket match broadcasts have been watched by millions. Cricket team's performance has become a symbol of national pride. Thus team selection assumes utmost importance.

\* Corresponding author. Tel.: +1 405 334 2351; fax: +1 405 744 5180.

E-mail addresses: [subbusrama@yahoo.com](mailto:subbusrama@yahoo.com), [subramanian.iyer@okstate.edu](mailto:subramanian.iyer@okstate.edu) (S.R. Iyer).

**Table 1**  
Prominent cricket playing countries

<i>Full members</i>
Australia
Bangladesh
England
India
New Zealand
Pakistan
Srilanka
South Africa
West Indies
Zimbabwe
<i>Associate members</i>
Argentina
Belgium
Bermuda
Botswana
Canada
Cayman Islands
Denmark
Fiji
France
Germany
Gibraltar
Hong Kong
Ireland
Israel
Italy
Japan
Kenya
Kuwait
Malaysia
Namibia
Nepal
Nigeria
Papua New Guinea
Scotland
Singapore
Tanzania
Thailand
The Netherlands
Uganda
United Arab Emirates
United States of America
Zambia

As is true with most team sports, it takes individual brilliance as well as team cohesion to make a winning team. To create this perfect fit, cricketing nations have different programs to identify and nurture talented cricketers in their early years. In spite of this effort only a few individuals achieve greatness or are able to play continuously for a long period. Thus the team selection committee members face a difficult task of first predicting each team candidate's future performance, and then applying a team/portfolio concept to select the final team. The International Cricket Council (ICC) has started ranking cricketers as well as teams numerically by assigning points to them based on various factors. However, these rankings have not been used extensively to select cricketers. In this study, we use neural networks to address the first issue of predicting individual cricketer's performance in the future.

A neural network is a biologically inspired information processing system which is capable of modeling extremely complex functions. A neural network is composed of a large number of highly interconnected elements (neurons) working together to solve specific problems. Neural networks are capable of learning from existing data and apply the learned knowledge to new observations. Neural networks are applicable where a relationship may exist between independent variables and dependent variables. Some of the fields where neural networks have been applied successfully are medicine (Anonymous, 2006; Papik et al., 1998), stock markets

(Altay & Satman, 2005; Jasic & Wood, 2004), genetics (Ritichie, 2005) and credit rating (Kumar & Bhattacharya, 2006). However, none of these include applications in rating cricketers and selecting a cricket team from an eligible pool of cricketers. This paper is one of the first attempts in cricketer performance rating and selection. This present study originated when Mr. Mahendra Mapagunaratne, a cricket enthusiast and free-lance journalist, contacted the second author after media coverage of the movie forecasting research (Sharda & Delen, 2006).

Our test environment is the recently concluded ICC World Cup played in West Indies. From the ten full members of the ICC we chose eight countries namely Australia, England, India, New Zealand, Pakistan, South Africa, Sri Lanka and West Indies. The choice of the above mentioned eight countries was based on the opinions of cricket experts. The game has undergone rapid technological and governing changes after 1985 so we chose to include only those players who were active after the 1985 season. In other words the data contains players who may have retired after the 1985 season along with players who made their debut after the 1985 season. The data on the cumulative performance of thus chosen cricketers was collected until the 2006–2007 season. The data organization is explained in detail in Sections 3.4.1 and 3.4.2. However, in the 2006–2007 season data we chose to exclude the World Cup performance data and use the World Cup as an avenue to measure the effectiveness of neural networks in predicting the cricketers' performance. We were also able to obtain the World Cup probable list of cricketers published by different countries. The neural networks were used to generate performance predictions for the World Cup probable cricketers. Some heuristics were then applied to recommend cricketers for each country's World Cup team. Some of our recommendations were made before the World Cup tournament and deposited with an independent party. After the World Cup, the performance of the each player was collected and prediction accuracy of neural networks was investigated. We present the method of this process as well as the performance of the model. The paper is organized as follows. Section 2 provides a brief literature review. Section 3 describes our methodology. Section 4 presents our results. Section 5 discusses the limitations and concludes the paper with comments on further applications.

## 2. Literature review

Very little research has been done on sports performance prediction using neural networks. Following is a brief literature review of the neural network applications in sports performance predictions.

Wilson (1995) used neural networks to model performance and rank NCAA college football teams. NCAA college football team winner is determined by voting. The voting system is subject to politicking and lobbying. The neural networks generated objective rankings which could be used to declare the college football winner for each season.

Bhandari et al. (1997) used data mining techniques to identify interesting patterns in National Basketball Association games in USA. The coaches can use these patterns to identify mistakes and make improvements to their game.

Condon, Golden, and Wasil (1999) examined Olympic success of countries using 17 independent variables. They built 27 different models – three models using ordinary least squares (OLS) linear regression models and 24 neural network models and found that neural network models outperformed regression models. The Olympic success was predicted using socio-economic factors. The neural networks were found to do better than OLS methods. These networks could be used by nations, in the

long term, to predict and improve their own success at Olympics. There is little a country could do, in the short term, to improve the success at Olympics. These networks could be used to compare different countries and closely examine the difference in performance between countries. In the long run countries with less resources could change their training methods, by analyzing the successful countries, and turn in a good performance at Olympics.

Maier, Wank, and Bartonietz (2000) examined the possibility of predicting javelin flights using multi-layer perceptron neural networks. Maps of distances reached by javelins were generated for different combinations of release parameters. They found that the most important parameter was the release velocity and that a moderate side angle of attack should be used to attain the longest throws to compensate for rotation of the javelin on release.

Kahn (2003) used a back propagation multi-layer perceptron neural network to predict the winner of National Football League (NFL) winners. The network had 3% more accuracy than the predictions by ESPN sportscasters.

Jürgen and Arnold (2003) used a Kohonen Feature Map (KFM) – a type of neural network to recognize patterns in Table Tennis and Rowing. Some of the variables used in this table tennis study are grip (shakehand/penhold), left/right hander, type of player (offensive/defensive) and so on. The model comprised very detailed information on various aspects of table tennis. Match videos were evaluated afterwards by two experienced table tennis players. The high-dimensional patterns with a lot of recorded information were reduced to manageable patterns using neural networks. The new models did reduce the amount of information used to recognize striking patterns and features. Qualitative visual analysis of the striking features of any table tennis player can be analyzed to identify types of players and tactical structures.

Flitman (2006) created a neural network model to predict winners in Australian Football League and compared the results with human tippers and other predictive models. The results were encouraging. The neural network and the linear program model performed well when compared with other human predictors as well as with other models.

One of the neural network applications in cricket has been to model bowler action (Bartlett, 2006). However, none of the studies appears to have focused in providing a decision aid to the team selection committee by providing a forecast of each individual cricketer's performance. This paper is one of the first to apply this concept in cricket.

### 3. Methodology

We obtained data from [www.Cricketarchive.com](http://www.Cricketarchive.com) (2007) as well as from <http://www.Cricinfo.com> (2007) for our purposes. The data was segregated into bowling and batting i.e., a cricketer is assumed to represent just one category of cricket – bowling or batting and the cricketers were rated subjectively. Although wicket-keeper is a vital part of the team there is not enough data to evaluate wicket keepers, therefore we did not develop a model to predict performance of wicket keepers. Batsmen and bowlers were assigned subjective ratings based on heuristic rules, which are explained in Sections 3.1.2 and 3.1.3. Let us call these assigned subjective ratings as “primary ratings”. Two experiments were performed using different neural networks. Neural networks in each experiment were trained and tested using primary ratings. After training and testing, each neural network generated its own ratings for all players.

Based on heuristics and using the neural network generated ratings we recommended players to respective World Cup teams. Finally, we compare and analyze the prediction of neural networks

(in both experiments) with the actual performance of cricketers in the recently concluded World Cup. Chart 1 describes the whole methodology.

#### 3.1. Primary rating generation process

##### 3.1.1. Dependent variable

The dependent variable is our characterization of a cricketer's performance or in other words ratings. The ICC provides rankings of cricketers; however, there is a need for selection committees to predict each cricketer's performance. After discussion with cricket connoisseurs who we had been in contact with, we decided that it would be sufficient to predict a cricketer's performance i.e., one of the three categories – Performer, Moderate and Failure. As the words imply, each player is classified into one of these classes. Since such classification is not formally available, we assigned these classes to each player based upon their past performance. We employed the following rules in categorizing cricketers into – “Performer”, “Moderate”, and “Failure”. These rules were created in consultation with cricket experts.

##### 3.1.2. Rules for rating batsmen

- A batsman is considered “Performer” if he has
  - a Scored more than or equal to 4000 runs across both versions of the game, and if he has played in 130 matches across both versions of the game and
  - b Played in at least 50 matches of each version.
- A batsman is considered “Moderate” if he has
  - a Scored more than or equal to 800 runs (but less than 4000 runs) across both versions of the game, and
  - b Played in at least 40 matches across both versions of the game
- A batsman is considered “Failure” if he has
  - a Scored less than 800 runs across both versions of the game, and
  - b Played in less than 40 matches across both versions of the game.

##### 3.1.3. Rules for rating bowlers

- A bowler is considered “Performer” if he has
  - a Bowled more than 7000 balls across both versions of the game, and
  - b Captured more than or equal to 150 wickets across both versions of the game, and
  - c Captured at least 30 wickets in each form of the game.
- A bowler is considered “Moderate” if he has
  - a Bowled more than 1000 balls (but less than 7000 balls) across both versions of the game, and
  - b Captured more than and equal to 40 wickets but less than 150 wickets across both versions of the game, and
  - c Captured at least 20 wickets in each version of the game.
- A bowler is considered “Failure” if he has
  - a Bowled less than 1000 balls across both versions of the game, and
  - b Captured less than 40 wickets across both versions of the game or
  - c Captured more than 40 wickets across both versions of the game but does not have at least 20 wickets in each version of the game.

We collected data on 234 batsmen and 256 bowlers from different countries. Primary ratings, assigned to batsmen and bowlers, were then used to train and test neural networks in both the experiments.

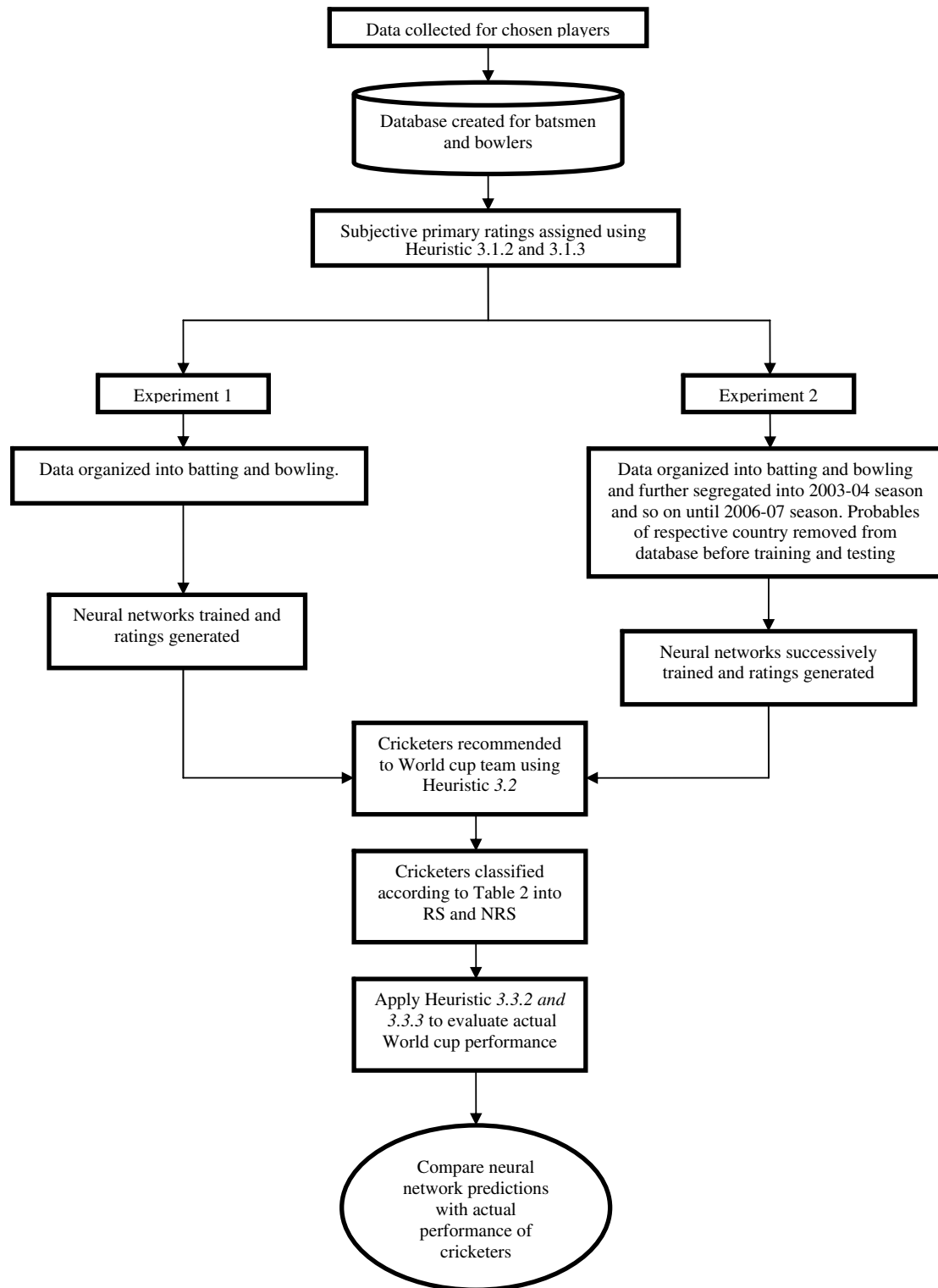


Chart 1. Methodology flow chart.

### 3.2. Heuristic rules for recommending cricketers to the world cup team

Two experiments were performed after the generation of primary ratings. Experiments are explained in Section 3.4. Heuristics mentioned below were used in both the experiments to recommend cricketers to the World Cup team.

- (i) Batsmen – Rule No. 1: If a batsman has received at least 2 “Performer” rating and at least 1 “Moderate” rating, and if the batsman did not receive any “Failure” rating he was recommended to be included in the team. Rule No. 2: If a batsman has received all “Moderate” ratings, and if he did not receive any “Failure” rating he was included in the team.

**Table 2**  
Class matrix of cricketers

Class 1 – RS Cricketers recommended and selected to the world cup team	Class 3 – NRNS Cricketers not recommended and not selected to the world cup team
Class 2 – NRS Cricketers not recommended but selected to the world cup team	Class 4 – RNS Cricketers recommended but not selected to the world cup team

- (ii) Bowler – Rule No. 1: If a bowler has received at least 2 “Performer”, and at least 1 “Moderate” rating he was recommended to be included in the team. Rule No. 2: If a bowler has at least 1 “Performer” rating, and 4 “Moderate” ratings for the rest, he was recommended to be included in the team. Rule No. 3: If a bowler received all “Moderate” ratings, and no “Failure” rating he was recommended to be included in the team.

### 3.2.1. Classification matrix

Using the performance ratings predicted by the neural networks and the heuristics mentioned in the Section 3.2, we recommended cricketers. The selection committees could select any cricketer from within their respective country. Some cricketers could be in our pool, others could be completely out of our pool. Thus, the teams consisted of players selected and also recommended by our models as well as players not recommended by the model but selected by the selection committees. On the other hand, there are players recommended by neural networks but not selected by the selection committee. Finally, there are players ignored by both, the neural networks and the selection committees. Hence four different classes of cricketers are in the test.

Since we cannot observe the performance of those cricketers who were recommended but not selected, the focus shifts to only two classes, which are classes RS and NRS. The probabilities were classified as per Table 2 and the classification was provided to Mr. Mahendra Mapagunaratne before the World Cup began, as an independent safe-keeper. As mentioned in Section 1, Mr. Mahendra Mapagunaratne is free-lance cricket journalist. Cricketers are classified according to Table 2 for both experiments. The World Cup becomes a true test of the prediction capability of neural network models. Analysis of the results is presented in Section 4.

### 3.3. Heuristic rules for analyzing world cup performance

#### 3.3.1. The total pool of probable cricketers

A total of 235 cricketers were considered for the World Cup team by all the selection committees of respective countries (Anonymous, 2007; Cricinfo, January 9, 2007; Cricinfo, January 10, 2007; Cricinfo, January 12, 2007; Cricinfo, January 13, 2007; Cricinfo, January 14, 2007; Seepersad, 2007; Vasu, 2007). The selection committees chose cricketers from this pool. This total pool was segregated into batsmen and bowlers; some cricketers are termed as an All-rounder since they are considered as a batsman as well as a bowler. There were 115 cricketers who were classified as batsmen and 107 cricketers were classified as bowlers. However, there were a few cricketers for whom data was not available from any source; these cricketers were excluded from the total pool of players we considered. Players recommended using neural networks and players selected by selection committees were compared and placed in the RS and NRS class according to Table 2.

#### 3.3.2. Heuristic rules for batting

According to the heuristic rules explained in Section 3.1.2, a batsman is rated a “Performer” if he has scored more than or equal

to 4000 runs across both versions of the game; if he has played in 130 matches across both versions of the game; and played in at least 50 matches in each version of the game. Then the approximate average score (Ave) by a batsman turns out to be 30. For a batsman, higher the average the better. To analyze the performance of batsmen in the World Cup, we use a benchmark average score of 30. If a batsman has an average higher than 30 we propose that the batsman has performed well. A batsman can play in just one match and score 30 runs. However, this performance alone does not lead to conclude that the batsman has performed well. Hence we also place another filter that the batsman should have scored at least 150 runs during the World Cup. In summary, the heuristic rule used to evaluate batsman performance are

- A batsman is said to have performed well if he has an average of at least 30, and if he has scored more than or equal to 150 runs.

#### 3.3.3. Heuristic rules for bowling

According to the heuristic rules explained in Section 3.1.3, a bowler is rated a “Performer” if he has bowled more than 7000 balls across both versions of the game; captured more than or equal to 150 wickets across both versions of the game, and captured at least 30 wickets in each version of the game. The approximate strike rate (S rate) by a bowler then turns out to be 47. For a bowler, lower the strike rate the better. To analyze the performance of a bowler in the World Cup, we use a benchmark strike rate of 47. If a bowler has a strike rate of lower than 47 we propose that the bowler has performed well. A bowler could capture just one wicket in 30 balls, which would give him a strike rate of 30. However, this performance alone does not lead to conclude that the bowler has performed well. Hence we also place another filter that the bowler should have captured at least 5 wickets. In summary, the heuristic rules used to evaluate bowler performance are

- A bowler is said to have performed well if he has captured at least 5 wickets with a strike rate less than 47.

### 3.4. Experiments

We used two different experiments to generate ratings and recommend players to the world cup team of respective countries from the pool of probables. Following is a brief description of the experiments.

#### 3.4.1. Experiment 1

Data was collected until the end of December 2006. Ratings were assigned to these cricketers using heuristics explained in Sections 3.1.2 and 3.1.3. While rating, careful consideration was given to cricketers who were active during 2006–2007 season. For example, cricketers may have made their debut before 2006–2007 season and could have been active during the 2006–2007 season. Cricketers who fall into this classification may not have achieved enough to be categorized as a “Performer”. However, these cricketers were rated as “Moderate” because we believe that the cricketers still have a fair chance to prove their potential and cannot be written off as “Failure”. If these “Moderate” cricketers faded away in the following season then the heuristics explained in Sections 3.1.2 and 3.1.3 can be used to assign a rating. Neural networks were employed to generate ratings for the World Cup probables. Some of the World Cup probable cricketers were also present in the original data set used for training and testing the neural networks. After generating the ratings, heuristics in Section 3.2 were used to recommend cricketers to their respective world cup



team. A summary and prediction accuracy of the neural networks is provided in Section 3.5.1.

### 3.4.2. Experiment 2

As opposed to the methodology followed in Experiment 1 we chose to segregate the data into four sets. The first set contains cumulative performance of players until 2003–2004 season; the second set contains cumulative performance until 2004–2005 season (which also includes the 2003–2004 season) and so on until 2006–2007 season. Ratings were then assigned to these cricketers using heuristics explained in Sections 3.1.2 and 3.1.3. While rating, careful consideration was given to cricketers who were active during each season. For example, cricketers may have made their debut just before 2004–2005 season and could have been active during the 2004–2005 season. Cricketers who fall into this classification may not have achieved enough to be categorized as a “Performer”. However, these cricketers were rated as “Moderate” because we believe that the cricketers still have a fair chance to prove their potential and cannot be written off as “Failure”. However, if these “Moderate” cricketers faded away in the 2005–2006 season then the heuristics explained in Sections 3.1.2 and 3.1.3 were used to assign a rating for 2005–2006 data set.

Further, we removed the probables for each country from the data set. For example, Australian probables were removed from data set and neural networks were created for batting and bowling. The rationale for performing experiment 2 is that we wanted to examine whether neural networks will perform even when multiple training is involved and further generate ratings on cricketers formerly alien to the networks.

These networks that were trained using four data sets were applied to generate ratings for Australian probables. This process ensured that the training data for neural networks did not include Australian probable cricketers. The same process, of removing probables from the data set and training neural networks was performed for each country. The 2006–2007 season data does not include the performance of the World Cup. Cricketers were then recommended to the World Cup team using heuristics explained in Section 3.2 and then classified into the four classes based on Table 2.

### 3.5. Data description

A cricketer develops through years of performance at the junior level and sub-national level. The performance of a player in the sub-national level catches the attention of team selectors who then vault him into the international level. A player participates in both ODI and Test versions in the sub-national as well as the international level. Due to the nature of available data we have named the sub-national level as youth level.

#### 3.5.1. Data for batsmen

The variables explained in this section appear for Youth One-Day, Youth-Test, One-Day (International Level) and Test (International Level) versions. Following is a short description of the independent variables:

**Matches** – This specifies the number of matches played by a batsman. A batsman gets better chances of scoring if he plays a higher number of matches. Conversely if a batsman plays more matches it could mean that he is a better batsman and scores higher.

**Innings** – Since there is a batting line-up for every team and each batsman specializes in playing at a certain position not all batsmen get an opportunity to bat in a match. Innings show the number of times a batsman had an opportunity to bat.

**Not out** – For a batsman being not out means that the opposition could not get him out which is an honor for the batsman.

**Runs** – A batsman's primary goal is to score runs for his team. A batsman is acknowledged by the number of runs he scores in a particular match as well as the overall runs he scores in his career.

**HS** – Stands for highest score. In the Test version of the game a batsman scoring 150 runs in an innings is not a rarity, however, in the One-Day version of the game surpassing 150 runs in an innings is considered to be good.

**Average** – Is the result of total runs per dismissal.

**Century (100)** – The number of times a batsman has scored 100 runs in an innings. A better batsman goes past the milestone on multiple occasions.

**Half-century (50)** – The number of times a batsman has scored 50 runs in an innings which is a very common milestone for a batsman.

**Year** – Shows the year in which a batsman started playing a particular version of the game. Longevity is another sign of a better batsman.

**Currently playing** – Suggests whether a batsman is currently a member of the national team. Somewhat correlates with the year the batsman made his first appearance. If the batsman has been around for many years, it could suggest that the batsman is a better batsman.

**Last played** – Many batsmen make their appearances and fade away with time and could not sustain their performance. For batsmen who are still part of the squad they will have “Active” entered in this variable.

**Dependent variable - Performance rating** – As described in Section 3.1.1 the performance rating is our dependent variable. Performance rating takes one of the three values: “Performer”, “Moderate”, and “Failure”.

There are three categorical variables among the above variables which are “Year”, “Currently Playing” and “Last Played”. If a batsman is currently playing; we use the year in which he started playing for the “Year” variable; for the “Currently Playing” variable we use “Yes”; and for the “Last Played” variable we use “Active”. If a batsman is currently not playing, we have used the year in which he started to play for the “Year” variable; for the “Currently Playing” variable we use “No” and for “Last Played” variable we use the actual year in which the batsman stopped playing.

Some batsmen may not have participated in a version of cricket or data is not available for their youth performance. In those cases we recorded a “0” or the null value for those data items.

Table 3 provides an idea regarding the data organization for batsmen. The same variables are used for a batsman's youth ODI, youth Test, national ODI, and national Test match performance. The variables are designated accordingly to identify each category.

#### 3.5.2. Data for bowlers

The variables explained in this section appear for Youth One-Day, Youth-Test, One-Day (International Level) and Test (International Level) versions. Following is a short description of the independent variables:

**Balls** – The number of times a bowler has bowled. The higher number of balls bowled by a bowler suggests that he could be a better bowler since he was chosen to bowl over others in the squad.

**Mdns** – An “over” is six attempts to bowl at a batsman by a bowler. The batsman may score in these six attempts. However, if a batsman is unsuccessful at scoring in any of these six “balls” (attempts) then the over is termed as a “Maiden” (Mdns).

**Runs** – The bowler aims at giving away the fewest number of runs during his ‘over’ and thus the lower number of runs given away suggest that the bowler is better.

**Table 3**  
Sample data for batsmen

Player name	Matches	Innings	Not out	Runs	HS	Average	Century	Half-century	Scoring rate	Catches	Year	Currently playing	Last played	Performance rating
BC Lara	277	268	30	9880	169	41.51	19	60	79.55	110	1990	Yes	Active	Performer
PA Wallace	33	33	0	701	103	21.24	1	2	60.9	11	1991	No	1999	Moderate
Jimmy Adams	127	105	28	2204	82	28.62	0	14	60.98	68	1992	No	2003	Performer
RIC holder	37	31	6	599	65	23.96	0	2	68.37	8	1993	No	1998	Moderate
SC Chanderpaul	201	189	24	5984	150	36.26	3	41	70.15	59	1994	Yes	Active	Performer

**Wickets** – The number of times a bowler is successful in ousting a batsman. The higher number of wickets suggests that a bowler is a better bowler.

**BB** – Best Bowling – A bowler may pick up wickets in an innings giving away different amount of runs. However, the best performance by a bowler will be the one where he picks up maximum wickets with the lowest runs given away.

**Ave** – Is the result of runs given away divided by the wickets captured.

**4 wI** – The number of occasions in which a bowler has captured four wickets in an innings. A bowler is considered to be better if he has a higher number to show against this parameter.

**5 wM** – The number of occasions in which a bowler has captured five wickets in a match. To capture five wickets in the ODI version of the game is a good performance as well as in Test matches, however, in a Test match the bowler has more overs to bowl at the opponent.

**10 wM** – The number of occasions in which a bowler has captured ten wickets in a match. This feat has not been achieved in an ODI match. However, it is not uncommon for bowlers to surpass this mark in a Test match.

**S rate** – is the total number of balls bowled to the total number of wickets captured by the bowler. If the bowler picks up wickets by bowling less number of balls, he is considered to be a better bowler.

**Econ rate** – is the total runs given away divided by total overs bowled by a bowler. A lower economy rate suggests that it is difficult to score runs of the bowler. For youth performance, data pertaining to this variable was not available hence this variable appears only for the ODI and Test match data.

**Year** – Shows the year in which a bowler started playing a particular version of the game. Longevity is another sign of a better bowler.

**Currently playing** – Suggests whether a bowler is currently member of the squad. Correlates with the year the bowler has made his first appearance. If the bowler has been around for many years it could suggest that the bowler is a better bowler.

**Last played** – A lot of bowlers made their appearances and faded with time and could not sustain their performance. For bowlers who are still part of the squad they will have the current year entered in this variable.

**Dependent variable - Performance rating** – As described in Section 3.1.1 the performance rating is our dependent variable. Performance rating takes one of the three values: “Performer”, “Moderate”, and “Failure”.

As we discussed regarding the batsmen, there are three categorical variables among the above variables which are “Year”, “Cur-

rently Playing” and “Last Played”. If a bowler is currently playing; we use the year in which he started playing for the “Year” variable; for the “Currently Playing” variable we use “Yes”; and for the “Last Played” variable we use “Active”. If a bowler is currently not playing then we use “No” for “Currently Playing” variable and for “Last Played” variable we use the actual year in which the batsman stopped playing. Missing data were handled the same way as with batsmen, by entering a “0”.

Table 4 provides a sample of data bowlers in the one-day version of the game. The same variables are used for a bowler's youth ODI and national ODI performance. The variable 4 wI is replaced with 10 wM in the Test version of the game. These variables are used for a bowler's, youth Test, and national Test match performance. The variables are designated accordingly to identify each category.

### 3.6. Model development

Neural networks were created using a commercial software product Statistica 7.1 by Statsoft, Inc. These neural networks were generated by the Intelligent Miner mode in Statistica. Only the best networks were retained by the Intelligent Miner mode. Following is a short description of the different types of neural networks used to rate cricketers.

Multilayer perceptrons (Haykin, 1994; Turban, Aronson, Liang, & Sharda, 2007), commonly referred to as MLPs, are one of the most widely used neural networks. MLPs consists of multiple layers of computational units, usually interconnected in a feed-forward way. Neurons in one layer have directed connections to the neurons of the subsequent layer. Inputs are processed through successive layers of neurons. The input layer contains neurons which represent the predictor/independent variables in the data. The top layer is the output layer which includes the dependent variable, namely the player rating. The layers in between are referred to as ‘hidden’ layers. MLPs learn through a variety of techniques, most popular being *back propagation*. In back propagation the output variables are compared with the answers to compute the values of error functions. Using various methods these error functions are then fed back to the network. After performing this process over numerous learning cycles the network attains the stage where the error between predicted and actual classification is sufficiently small (Fig. 1).

Without a ‘hidden’ layer the perceptron can be called a linear network (Turban et al., 2007). Linear neural networks use a simple linear regression model for learning. They do not contain a hidden layer. Linear neural networks perform an important function of

**Table 4**  
Sample data for bowlers

Player name	Balls	Mdns	Runs	Wkts	BB	Ave	4 wI	5wM	S rate	Year	Currently playing	Last played	Performance rating
AC Cummins	6352	75	4508	168	05–16	26.83	6	3	37.8	1988	No	1996	Moderate
KCG Benjamin	1319	12	923	33	03–34	27.96	0	0	39.96	1992	No	1997	Moderate
CE Cuffy	2153	41	1436	41	04–24	35.02	1	0	52.51	1994	No	2003	Moderate
VC Drakes	1640	22	1293	51	05–33	25.35	3	2	32.15	1994	No	2003	Moderate

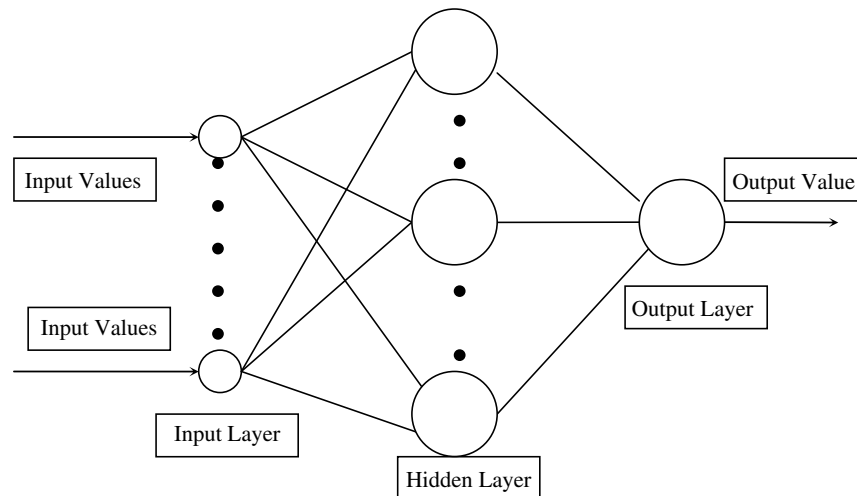


Fig. 1. Graphical representation of MLP neural network architecture.

benchmarking other neural networks with a linear model. If the problem is linear, then neural network based non-linear solutions cannot be expected to produce a superior result.

Radial basis function (RBF) neural networks (Bose & Liang, 1996; Haykin, 1994) employ a hidden layer of radial units and an output layer of linear units. RBF networks are fast to learn and reasonably compact. RBF networks have three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer. The neurons in the hidden layer contain functions whose outputs are inversely related to the distance from the center of the neuron. The output layer consists of the dependent variable which in this case is the cricketer rating. RBF networks are advantageous over MLP networks and linear networks since RBFs can model any non-linear function using a single hidden layer (Fig. 2).

### 3.6.1. Experiment 1 – summary of networks

**3.6.1.1. Neural networks – batting.** The five neural networks chosen by Statistica Intelligent Miner mode for rating are – two MLPs, one linear neural network, and two RBF networks. There were 50 input neurons. A summary of the input layers and hidden layers is provided in Table 5.

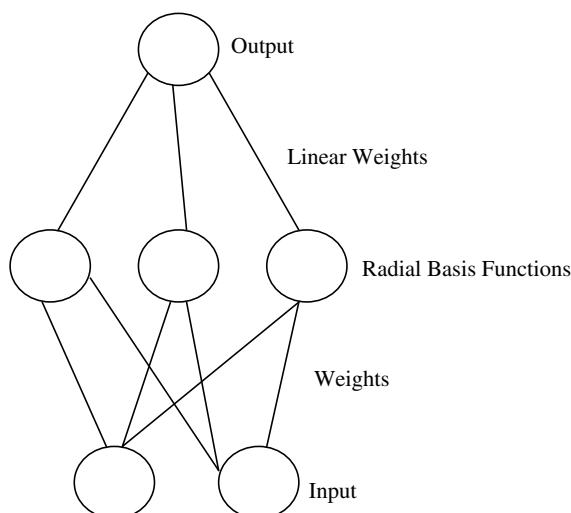


Fig. 2. Graphical representation of RBF neural network architecture.

- **MLP** – The first MLP neural network used 41 input neurons and 17 neurons in the hidden layer whereas the second MLP used 41 input neurons and 19 neurons in the hidden layer. The MLP 1 chose to leave out Matches YO, Runs YO, HS YO, 100 YO, Currently Playing YO, Not Out YT, Average YT, Currently Playing YT, and Catches O variables. MLP 2 chose to leave out Matches YO, HS YO, Currently Playing YO, Last Played YO, Runs YT, HS YT, Catches YT, Currently Playing YT, and Last Played YT variables. Majority of the input variables which were not used by neural networks are of the Youth performance category for which data is not available for many batsmen.
- **Linear** – The linear neural network used 49 input neurons, left out Last Played YT variable and employed no hidden layers.
- **RBF** – The first RBF neural network used 33 input neurons and 3 neurons in the hidden layer, whereas the second RBF neural network used 33 input neurons and 7 neurons in the hidden layer. RBF 1 and RBF 2 chose to leave out Matches YO, Innings YO, Not Out YO, Runs YO, HS YO, 100 YO, 50 YO, Catches YO, Currently Playing YO, Matches YT, Innings YT, Not Out YT, Runs YT, Average YT, 100 YT, 50 YT and Currently Playing YT variables. Similar to the MLP network RBF chose to leave out most of the Youth performance category for which data is not available for many batsmen.

The neural networks and the number of input neurons in each neural network is based upon the best network in each category as determined through the training and validation process by Statistica in the Intelligent Miner mode.

**3.6.1.2. Neural networks – bowling.** We have trained five neural networks for classification – two MLPs, one linear neural network, and two RBF networks. There were 50 input neurons. A summary of the input layers and hidden layers is provided in Table 6.

Table 5  
Neural network profile – batting

Neural network profile	Number of neurons in input layer	Number of neurons in hidden layer	Number of neurons in output layer
MLP 1	41	17	1
MLP 2	41	19	1
Linear	49	0	1
RBF 1	33	3	1
RBF 2	33	7	1



**Table 6**

Neural network profile – bowling

Neural network profile	No of neurons in input layer	No of neurons in hidden layer	Output layer
MLP 1	40	21	1
MLP 2	41	20	1
Linear	42	0	1
RBF 1	15	4	1
RBF 2	15	8	1

- **MLP** – The first MLP neural network used 40 input neurons and 21 neurons in the hidden layer whereas the second MLP used 41 input neurons and 20 neurons in the hidden layer. MLP 1 chose to leave out BB YO, Ave YO, Mdns YT, BB YT, Mdns O, BB O, Econ Rate O, Currently playing O, BB T and Econ Rate T variables. MLP 2 chose to leave out BB YO, Currently playing YO, BB YT, Ave YT, BB O, Econ Rate O, Currently playing O, BB T and Econ Rate T variables.
- **Linear** – The Linear neural network used 42 input neurons and no hidden layers. The linear network chose to leave out BB YO, BB YT, 5wI YT, 10wM, YT, BB O, Econ Rate O, BB T and Econ Rate T variables.
- **RBF** – The first RBF neural network used 15 input neurons and 14 neurons in the hidden layer, whereas the second RBF neural network used 15 input neurons and 8 neurons in the hidden layer. The RBF networks chose to leave out Balls YO, Mdns YO, Runs YO, Wkts YO, BB YO, Ave YO, 4wI YO, 5wM YO, s rate YO, Year YO, Last played YO, Balls YT, Mdns YT, Runs YT, Wkts YT, BB YT, Ave YT, 5wI YT, 10wM YT, Year YT, Last played YT, BB O, Ave O, 5wM O, s rate O, Econ Rate O, Year O, Currently playing O, Last played O, BB T, Ave T, 5wI T, 10wM T, S rate T, and Econ Rate T variables.

The neural networks and the number of input neurons in each neural network is based upon the best network in each category as determined through the training and validation process by Statistica in the Intelligent Miner mode. Majority of the input variables which were not used by the neural networks are of the Youth performance category for which data is not available for many bowlers.

**3.6.1.3. Overall prediction comparison of neural networks with assigned ratings.** The objective of this test was to check whether the neural networks were capable of learning and forecasting using the learned model.

- Batting accuracy** – The ratings generated by the neural networks were compared with the ratings assigned using the heuristic rules explained in Section 3.1.2. For example, a batsman may subjectively be rated as “Performer” and if the neural network rates him as a “Performer” we call that an accurate rating, but neural network could rate him as “Moderate”, or even as a “Failure” which we call an inaccurate rating. We ran the training and testing process through 10-fold cross validation with 80% in training and 20% in testing. Table 7 reports the testing accuracy based on the average of these 10 runs. Among the 234 batsmen, MLP1 neural network agreed in 196 ratings when compared with the subjective ratings, which translates to 84% accuracy. Similarly MLP2 had an accuracy of 86% and so on. As we can see, the networks performed reasonably well with good accuracy, which confirms our belief, that neural network based approach can be used in selecting cricketers for national teams.

**Table 7**

Summary of accuracy – neural networks – batting

	MLP1	MLP2	Linear	RBF1	RBF2
No of cases in agreement	196	202	185	170	184
Total players	234	234	234	234	234
Testing accuracy (%)	84	86	79	73	79

**Table 8**

Summary of accuracy – neural networks – bowling

	MLP1	MLP2	Linear	RBF1	RBF2
Cases of correctly predicted	206	203	178	166	180
Total cases	256	256	256	256	256
Testing accuracy (%)	80	79	70	65	70

- Bowling accuracy** – The ratings generated by the neural networks were compared against the ratings assigned using the heuristic rules explained in Section 3.1.3. For example a batsman may subjectively be rated as “Performer” and if the neural network rates him as a “Performer” we call that an accurate rating, but neural network could rate him as “Moderate”, or even as a “Failure” which we call an inaccurate rating. We ran the training and testing process through 10-fold cross validation with 80% in training and 20% in testing. Table 8 reports the testing accuracy based on the average of these 10 runs. Among the 256 bowlers, MLP1 neural network agreed in 206 ratings when compared with the subjective ratings, which translates to 80% accuracy. Similarly MLP2 had an accuracy of 79% and so on. We can see from the reasonably accurate ratings by the neural networks that neural networks were able to learn and train well. These neural networks were used to generate ratings for World Cup probabilities – batsmen and bowlers. Heuristics were subsequently used to recommend cricketers to the World Cup team and what follows is an explanation of those heuristics.

### 3.6.2. Experiment 2 – summary of neural networks

In experiment 2 the probabilities for each country were removed from the four data sets and neural networks were created for batting and bowling for each country. For example, five neural networks each for batting and bowling were created for Australia. These neural networks were trained using four data sets of cumulative performance starting from 2003–2004 season to the 2006–2007 season. These neural networks were then used to generate ratings for Australian probabilities. We ended up creating 80 neural networks for eight countries. Owing to considerations of brevity we are not providing the summary of neural networks and overall prediction accuracy that were constructed for each country. These neural networks were used to generate ratings for the respective probabilities for each country. Cricketers were then recommended to the World Cup team using heuristics explained in Section 3.2 and then classified into the four classes based on Table 2. The analysis of the performance of cricketers in the World Cup is provided in Section 4.

## 4. Results

### 4.1. Analysis and comparison of results

The selection committees chose cricketers to be in the probable list from which they chose the final team. The probable cricketers consisted of batsmen, bowlers, and all-rounders, who can bat and bowl. We discussed with cricketing experts and segregated the

cricketers into batsmen and bowlers. We classified 115 cricketers as batsmen and out of which 70 batsmen were selected to the World Cup team. We classified 107 cricketers as bowlers and out of which 65 bowlers were selected to the World Cup team. As exposted in Section 3.2.1.1 our focus is on Class RS (cricketers who were recommended by neural networks and selected) and Class NRS cricketers (cricketers who were selected but not recommended by neural networks). The performance data of the players during the world cup was collected and classified as “Performers” according to heuristics explained in Sections 3.3.2 and 3.3.3.

#### 4.1.1. Pre world cup recommendation accuracy – batting

We begin the analysis by looking at the recommendation accuracy of the neural networks vis-à-vis the selection committee choices. Table 9 provides an account of the batsmen recommended by the neural networks using the experiments compared to the total number of batsmen selected to the World Cup team.

In both experiments we can see that the neural networks performed a reasonably accurate task of identifying the batsmen to be recommended to the respective World Cup team. In other words neural networks were able to identify and recommend the likely performers prior to the world cup. Neural networks and the selection committee members were in agreement in more than 70% (see column 3) cases for both experiments.

#### 4.1.2. Pre world cup recommendation accuracy – bowling

Table 10 provides an account of the bowlers recommended by the neural networks using the experiments vis-à-vis the total number of bowlers selected to the World Cup team.

In both experiments we can see that the neural networks performed a reasonably accurate task of identifying the bowlers to be recommended to the respective World Cup team. Neural networks and the selection committee were in agreement in more than 60% (see column 3) bowlers for experiment 1. Neural net-

works and selection committee were in agreement in more than 45% (see column 3) bowlers for Experiment 2.

#### 4.1.3. Post world cup analysis – batting

A list of batsmen who qualified as performers according to the heuristics explained in Section 3.3.2 was created for both experiments. Similar lists were generated in the bowling category for both experiments. We are interested in the performance of batsmen in Class RS and Class NRS post World Cup. Table 11 provides a summary of the performers in each class.

In experiment 1 Class RS contained 54 batsmen, these were batsmen recommended by neural networks as well as selected by the selection committee. However, selection committees selected 16 others players who would not have been recommended by neural networks. Whereas 50% of the batsmen recommended by neural networks performed well, only 25% of the batsmen selected by the selection committees, but not recommended by neural networks, performed well. Similarly in experiment 2, 46% of the batsmen selected by neural networks performed well whereas 40% of batsmen selected by selection committees, but not recommended by neural networks, performed well. Neural networks clearly enjoy superior performance over the selection committees. Chart 2 provides the percentage comparison of neural networks with selection committees for both the experiments.

According to the heuristic explained in Section 3.3.2 there were 31 performers in total. Let us consider experiment 1; there were 27 performers in Class RS and only 4 in Class NRS. Neural networks were able to identify 87% of actual performers before the World Cup. In experiment 2 there were 23 performers identified by neural networks prior to the World Cup whereas 8 batsmen performed well from Class NRS. Neural networks were able to identify 74% of the performers prior to the World Cup. Charts 2 and 3 present a consolidated picture of the percentage of batsmen in Class RS and Class NRS for both the experiments.

**Table 9**

Recommendation accuracy of neural networks – batting

(1)	Class RS – recommended and selected		Class NRS – not recommended yet selected		Total selected (6) = (2) + (4)
	(2) No. of players	(3) Percentage = (2)/(6)	(4) No. of Players	(5) Percentage = (4)/(6)	
Experiment 1	54	77%	16	23%	70
Experiment 2	50	71%	20	29%	70

**Table 10**

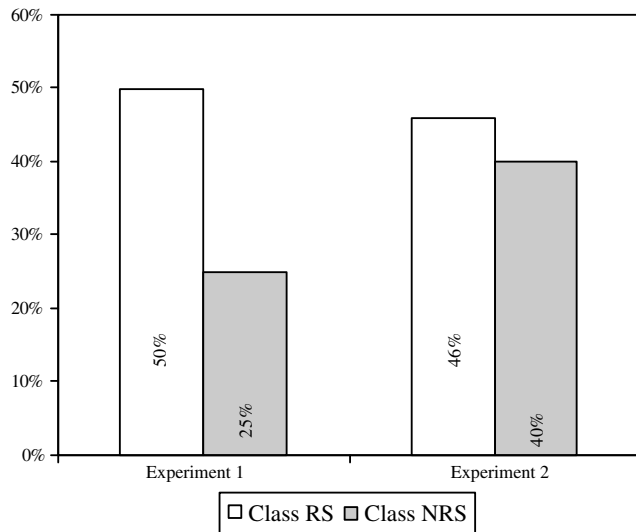
Recommendation accuracy of neural networks – bowling

(1)	Class RS – recommended and selected		Class NRS – not recommended yet selected		Total selected (6) = (2) + (4)
	(2) No. of players	(3) Percentage = (2)/(6)	(4) No. of players	(5) Percentage = (5)/(6)	
Experiment 1	41	63%	24	37%	65
Experiment 2	32	49%	33	51%	65

**Table 11**

Post World Cup Performance – Batting

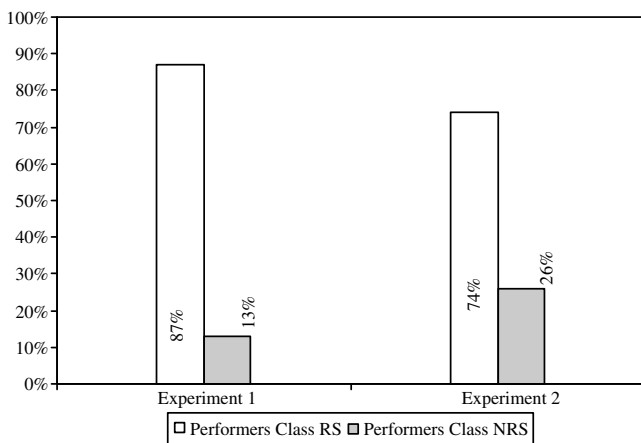
Post world cup performance – batting					Total performers	Total selected
Experiments	Class RS		Class NRS			
	No. of performers	Total players	No. of performers	Total players		
Experiment 1	27	54	4	16	31	70
Experiment 2	23	50	8	20	31	70



**Chart 2.** Success rate of neural network and selection committees in identifying performers.

There were a total of 70 batsmen selected by selection committees from different countries. Out of these, 31 batsmen performed well which means that selection committees were able to identify 44% of the all performers. By contrast, neural networks in both the experiments could identify more than 73% of the performers. In other words neural networks had a value addition of at least 30%.

A significant result in the above comparison is that a high percentage of batsmen in Class RS performed well as compared to a lower percentage of performing batsmen in Class NRS. In other words neural networks successfully identified most of the likely performers before the World Cup and the post World Cup results clearly show that the neural network based approach can yield markedly positive results in terms of predicting performers.

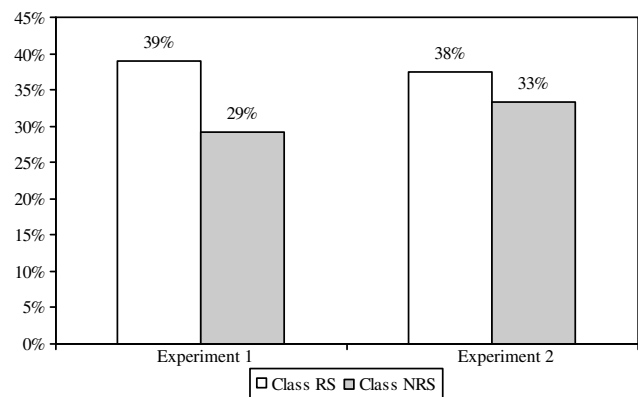


**Chart 3.** Percentage of performers in each class – batting.

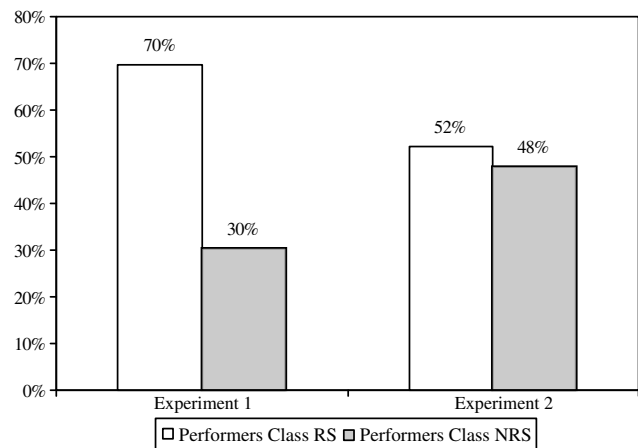
#### 4.1.4. Post world cup performance – bowling

We are interested in the performance of bowlers in Class RS and Class NRS post World Cup. Table 12 provides a summary of the performers in each class.

In experiment 1, Class RS contained 41 bowlers, these were bowlers recommended by neural networks as well as selected by the selection committee. However, selection committees selected 24 players who would not have been recommended by neural networks. When 39% of the bowlers recommended by neural networks performed well, only 29% of the bowlers selected by the selection committees, but not recommended by neural networks, performed well. Similarly in experiment 2, 38% of the bowlers recommended by neural networks performed well, whereas 33% of bowlers selected by selection committees, but not recommended by neural networks, performed well. Neural networks were able to identify a higher percentage of the performing bowlers than



**Chart 4.** Success rate of neural network and selection committees in identifying performers.



**Chart 5.** Percentage of performers in each class – bowling.

**Table 12**  
Post world cup performance – bowling

Experiments	Class RS		Class NRS		Total performers	Total selected
	No. of performers	Total players	No. of performers	Total players		
Experiment 1	16	41	7	24	23	65
Experiment 2	12	32	11	33	23	65

the selection committees in both experiments. Chart 4 provides the percentage wise comparison of neural network comparison with selection committees for both the experiments.

According to the heuristic explained in Section 3.3.2 there were 23 performers in total. Let us consider experiment 1; there were 16 performers in Class RS and only 7 in Class NRS. Neural networks were able to identify 70% of actual performers before the World Cup. In experiment 2 there were 12 performers identified by neural networks prior to the World Cup, whereas 11 bowlers turned out to be performers from Class NRS. Neural networks were able to identify 52% of the actual performers before the World Cup. Charts 4 and 5 present a consolidated picture of the percentage of batsmen in Class RS and Class NRS for both the experiments. Chart 5 provides a consolidated picture of the comparison of performers in both classes for both the experiments.

There were a total of 65 bowlers selected by selection committees from different countries and out of these 23 bowlers performed well which means that selection committees were able to identify 35% of the all performers. By contrast, neural networks in both the experiments could identify 70% of the performers in experiment 1 and 52% of the performers in experiment 2. In other words neural networks had a value addition of 35% in experiment 1 and 17% in experiment 2.

A high percentage of bowlers in Class RS were successful when compared to a low percentage of successful bowlers in Class NRS. In other words neural networks successfully identified most of the likely performers much before the world cup. The post World Cup results lead us to conclude that the neural network based approach can indeed be applied in team selection situations.

## 5. Discussion and future directions

Our study, perhaps, is the first attempt in cricket performance prediction. This humble beginning shows us that neural networks can indeed be applied to performance prediction generation in Cricket and perhaps other sports as well.

This study being an exploratory project has some shortcomings. First, we recognize that fielding is an important aspect of cricket and we have not included fielding performance of any player into our study. Batsmen and bowlers can be valuable fielders and the selection of a player is often influenced by his fielding abilities. However, we cannot quantify a cricketer's fielding abilities other than by the number of catches, which we believe is not a good measurement of fielding ability. Secondly, wicket-keeping is another important aspect of the game and we have not included wicket-keeping analysis in our study. However, we are fully aware that a wicket-keeper is imperative in a team primarily for his wicket-keeping abilities as well as for his batting abilities. The lack of a considerable sample hinders the generation of ratings for wicket-keepers.

Our study shows that neural networks can indeed learn to classify a bowler or a batsman based on his past performance into pre-defined categories. The results are encouraging and we hope to continue our effort in this direction. Selection of players for team sports is extremely important since each player's collective effort makes a winning team. Neural networks rely only on data and thus are purely objective in rating. These can be used to rank player performance in any team sports. Many factors influence the selection of cricketers into national teams, but neural networks can be used as an objective way to aid the selection process.

Neural networks could be used to identify potential "performers" early in their careers by analyzing the data from their sub-national and provincial performance. Selectors and scouts can use neural networks to identify cricketers before they make a national appearance.

Marketers spend a lot of money on endorsements and sponsorships. For marketers the biggest challenge is to identify a "Performer". If marketers can identify a player sufficiently early in his career, the player can be contracted at a lower cost than spending a higher amount later. Neural networks can be used to identify and nurture talented sportspersons early in their career. This has two implications, firstly for marketers it will be a boon if they could contract a person early in his career, and secondly scouts and selectors can nurture a talented sportsperson. We hope to channel our efforts in the direction of identifying "Performers" early on in their careers.

## Acknowledgements

We would like to express our gratitude to Mr. Mahendra Mapagunaratne who initiated this study. Mr. Mapagunaratne contacted the second author after the media coverage surrounding the movie forecasting project. In experiment 1 we predicted individual players' performance rating (i.e., "Performer", "Moderate", "Failure") and submitted the names of players predicted as performers to Mr. Mapagunaratne.

## References

- Altay, E., & Satman, M. H. (2005). Stock market forecasting: Artificial neural network and linear regression comparison in an emerging market. *Journal of Financial Management and Analysis*, 18(2), 18–33.
- Amy, M. (2007). The history of cricket, eSSORTMENT. [http://www.essortment.com/hobbies/historycricket\\_sngj.htm](http://www.essortment.com/hobbies/historycricket_sngj.htm).
- Anonymous (2006). Lund University: Acute coronary syndrome patient triage is aided with artificial neural network, *Medicine and Law Weekly*, June 23 (pp. 543–543).
- Anonymous (2007). Squads for Commonwealth Bank Series and 2007 ICC World Cup, (January 5, 2007). <http://cricket.com.au/default.aspx?s=newsdisplay&id=38511>.
- Bartlett, R. (2006). Artificial intelligence in sports biomechanics: New dawn or false hope? *Journal of Sports Science and Medicine*, 5(4), 474–479.
- Beaudoin, D. (2003). The best batsmen and bowlers in one-day cricket. Thesis. Canada: Laval University.
- Bhandari, I., Colet, E., Parker, J., Pines, Z., Pratap, R., & Ramanujam, K. (1997). Advanced scout: Data mining and knowledge discovery in NBA data. *Data Mining and Knowledge Discovery*, 1, 121–125.
- Bose, N. K., & Liang, P. (1996). *Neural network fundamentals with graphs, algorithms and applications*. New York: McGraw-Hill Inc.
- Condon, E. M., Golden, B. L., & Wasil, E. A. (1999). Predicting the success of nations at the summer Olympics using neural networks. *Computers and Operations Research*, 26(13), 1243–1265.
- Cricinfo.com (2007) is part of the Wisden Group. World Wide Web Address is <http://www.cricinfo.com/>.
- Cricinfo (2007). Mills and Oram in World Cup preliminary squad, (January 14, 2007). <http://content-usa.cricinfo.com/wc2007/content/story/276390.html>.
- Cricinfo (2007). Arnold and Chandana handed World Cup lifeline, (January 12, 2007). <http://content-usa.cricinfo.com/wc2007/content/story/276263.html>.
- Cricinfo (2007). Giles included in England World Cup probables, (January 13, 2007). <http://content-usa.cricinfo.com/wc2007/content/story/276318.html>.
- Cricinfo (2007). Pakistan include Shoaib and Asif, (January 10, 2007). <http://content-usa.cricinfo.com/pakistan/content/story/276038.html>.
- Cricinfo (2007). South Africa hope to meet racial quotas, (January 9, 2007). <http://content-usa.cricinfo.com/southafrica/content/story/275937.html>.
- Cricket Archive (2003–2007). World Wide Web Address is <http://cricketarchive.co.uk/>.
- Flitman, A. M. (2006). Towards probabilistic footy tipping: A hybrid approach utilizing genetically defined neural networks and linear programming. *Computers and Operations Research*, 33(7), 2003–2022.
- Haykin, S. (1994). *Neural networks – a comprehensive foundation*. New York: MacMillan College Publishing Company.
- Jasic, T., & Wood, D. (2004). The profitability of daily stock market indices trades based on neural network predictions: Case study for the S&P 500, the DAX, the TOPIX and the FTSE in the period 1965–1999. *Applied Financial Economics*, 14(4), 285–297.
- Jürgen, P., & Arnold, B. (2003). Application of neural networks to analyze performance in sports. *Proceedings of the 8th annual congress of the European college of sport science* (pp. 342). Salzburg: ECSS.
- Kahn, J. (2003). *Neural network prediction of NFL football games*. University of Wisconsin – Electrical and Computer Engineering Department.
- Kumar, K., & Bhattacharya, S. (2006). Artificial neural network vs. linear discriminant analysis in credit ratings forecast; a comparative study of prediction performances. *Review of Accounting and Finance*, 5(3), 216–227.

- Maier, K. D., Wank, V., & Bartonietz, K. (2000). Neural network based models of javelin flight: Prediction of flight distances and optimal release parameters. *Sports Engineering*, 3(1), 57–63.
- Papik, K., Molnar, B., Schaefer, R., Dombovari, Z., Tulassay, Z., & Feher, J. (1998). Application of neural networks in medicine – a review. *Medical Science Monitor*, 4(3), 538–546.
- Ritichie, M. D. (2005). Genetic programming optimized neural networks for identifying gene–gene interactions. Dissertation. Nashville, Tennessee: Vanderbilt University.
- Seepersad, R. (2007). Pollard included in World Cup probables, (January 13, 2007). <http://content-usa.cricinfo.com/wc2007/content/story/276307.html>.
- Sharda, R., & Delen, D. (2006). Predicting box office success of motion pictures with neural networks. *Expert Systems with Applications*, 30(2), 243–254.
- Statsoft Inc. Tulsa, OK, USA. (2007). The World Wide Web Address is [www.statsoft.com](http://www.statsoft.com).
- Turban, E., Aronson, J. E., Liang, T. P., & Sharda, R. (2007). *Decision support and business intelligence systems*. New Jersey: Pearson Prentice Hall.
- Vasu, A. (2007). Sehwaq out, Ganguly picked for ODIs, (January 12, 2007). <http://content-usa.cricinfo.com/wc2007/content/story/276198.html>.
- Wilson, R. (1995). Ranking college football teams: A neural network approach. *Interfaces*, 25(4), 44–59.