Journal of Quantitative Analysis in Sports

Volume 5, *Issue* 1 | 2009 | *Article* 7

Predicting NBA Games Using Neural Networks

Bernard Loeffelholz, *Air Force Institute of Technology*
Earl Bednar, *Air Force Institute of Technology*
Kenneth W. Bauer, *Air Force Institute of Technology*

Recommended Citation:
Loeffelholz, Bernard; Bednar, Earl; and Bauer, Kenneth W. (2009) "Predicting NBA Games Using Neural Networks," *Journal of Quantitative Analysis in Sports*: Vol. 5: Iss. 1, Article 7.

DOI: 10.2202/1559-0410.1156

©2009 American Statistical Association. All rights reserved.

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

Predicting NBA Games Using Neural Networks

Bernard Loeffelholz, Earl Bednar, and Kenneth W. Bauer

Abstract

In this paper we examine the use of neural networks as a tool for predicting the success of basketball teams in the National Basketball Association (NBA). Statistics for 620 NBA games were collected and used to train a variety of neural networks such as feed-forward, radial basis, probabilistic and generalized regression neural networks. Fusion of the neural networks is also examined using Bayes belief networks and probabilistic neural network fusion. Further, we investigate which subset of features input to the neural nets are the most salient features for prediction. We explored subsets obtained from signal-to-noise ratios and expert opinions to identify a subset of features input to the neural nets. Results obtained from these networks were compared to predictions made by numerous experts in the field of basketball. The best networks were able to correctly predict the winning team 74.33 percent of the time (on average) as compared to the experts who were correct 68.67 percent of the time.

KEYWORDS: feed-forward neural networks, radial basis functions, probabilistic neural network, generalized regression neural networks, Bayesian belief networks, fusion, signal-to-noise ratio, basketball

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

1. Introduction

Since the winter of 1891, following the day Dr. James Naismith nailed a peach basket into a gym wall, basketball has evolved into a true American game (NMHOF, 2008). Nearly 270,000 people, each game day, attend basketball arenas around the country to watch the best of the best sweat, hustle, and entertain (Ibisworld, 2008). Along with watching the games, millions of fans are involved with the ever growing arena of fantasy basketball leagues and other gambling alternatives. Involvement in these leagues and gambling precipitates the desire to know which team will win before they even participate in a game.

In this paper we examine the use of neural networks as tool for predicting the success of basketball teams in the National Basketball Association (NBA). Further, we investigate which subset of features input to the neural nets are the most salient features for prediction.

2. Previous Research

The research performed in this paper is similar in nature to that found in the works of Purucker (1996), Kahn (2003), and Calster et al. (2008). Purucker (1996) applied back-propagation, self-organizing maps (SOMs) and other neural structures to predict the winner of a football game in the National football League (NFL). Purucker investigated different training methods and determined back-propagation would be best to develop a model with greater predictive accuracy than various experts in the field of football predictions. Purucker achieved 61% accuracy as compared to the 72% accuracy of the experts.

Kahn (2003) extended the work of Purucker (1996) to strengthen the model developed to predict NFL football games. Back-propagation was also employed by Kahn, but a different learning mechanism as well as network structure was applied. Kahn was able to attain 75% accuracy, performing far better than Purucker and slightly better than the experts in the field. Both authors used differential statistics from the box scores rather than raw statistics.

Calster et al. (2008) applied Bayesian Belief Networks (BBNs) in the area of professional soccer. These authors investigated causes of unwanted draws that occur in soccer. These authors confirmed that Bayesian networks are beneficial if prior information or probabilities are known.

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

3. Materials and Methods

3.1 Data

The database used in this research consists of box scores from NBA games played in the 2007-2008 season. 620 games of the season are used for the training and test samples and 30 games are used as the validation set to represent "un-played" games. With 30 teams being present in the league, this amounts to each team playing about 20 games or a quarter of a season. This amount of games per team should be representative of how each team is expected to perform on a future game. The first 650 games of the season were chosen since the least amount of transactions and injuries occur during this time. The trade deadline is not until around mid-season, therefore the average statistics should reflect each team accurately. Also, staying within the season was important to capture each team's averages and avoiding any off-season trades or improvements that may drastically change a particular team's performance. The box scores were downloaded from www.espn.com using Microsoft Excel to organize and separate the data. Figure 1 displays the typical box score downloaded.

BOSTON CELTICS													
STARTERS	MIN	FGM-A	3PM-A	FTM-A	OREB	DREB	REB	AST	STL	BLK	TO	PF	PTS
Paul Pierce, SF	25	6-12	2-4	8-8	0	2	2	5	1	0	2	4	22
Kevin Garnett, PF	24	7-13	0-0	2-2	3	5	8	0	0	0	1	4	16
Kendrick Perkins, C	27	4-7	0-0	6-6	3	8	11	1	0	3	1	1	14
Ray Allen, SG	26	2-7	1-3	6-7	2	4	6	3	3	1	2	4	11
Rajon Rondo, PG	32	5-9	0-0	6-8	0	3	3	4	1	0	2	3	16
BENCH	MIN	FGM-A	3PM-A	FTM-A	OREB	DREB	REB	AST	STL	BLK	TO	PF	PTS
James Posey, SF	23	2-7	1-6	4-4	0	6	6	2	2	0	1	2	9
Glen Davis, C	11	2-2	0-0	0-2	0	0	0	1	0	1	1	2	4
Eddie House, PG	22	3-8	1-3	1-1	1	4	5	5	0	1	3	1	8
Leon Powe, PF	15	3-6	0-0	4-4	0	3	3	0	0	2	0	0	10
Tony Allen, SG	21	0-4	0-1	0-0	1	3	4	2	1	0	4	2	0
P.J. Brown, FC	6	1-2	0-0	1-1	2	4	6	0	1	0	2	2	3
Brian Scalabrine, PF	8	1-2	1-1	0-0	0	1	1	0	0	1	1	1	3
TOTALS		FGM-A	3PM-A	FTM-A	OREB	DREB	REB	AST	STL	BLK	TO	PF	PTS
		36-79	6-18	38-43	12	43	55	23	9	9	20	26	116
		45.6%	33.3%	88.4%									Team TO (pts off): 21 (20)

Figure 1. Layout of typical box score

The information extracted from the box score came from the bottom three lines. Table 1 provides clarification for each statistic. Information on the team totals in the game, as well as their home/away situation, were the only features used to conduct the neural network analysis. Table 2 displays an example set of 6

DOI: 10.2202/1559-0410.1156

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

games where each exemplar consists of the away team statistics, the home team statistics, and the winner of the game (1 for away and 2 for home).

Table 1. Statistics and their Abbreviations

Abbreviation	**Statistic**	**Abbreviation**	**Statistic**
FG	Field Goal %	Stl	Steals
3P	Three Point %	Blk	Blocks
FT	Free Throw %	TO	Turnovers
Oreb	Offensive Rebounds	PF	Personal Fouls
Dreb	Defensive Rebounds	PTS	Points
Ast	Assists		

The statistics shown in Table 1, and different subsets of these statistics, were used since they are common to basketball and the typical fan understands what each statistic represents. Other statistics could be used, such as pace, efficiency, or the per 48 minute (length of a typical basketball game) statistic, but these are more complicated to understand. Page et al. (2007) recommend examining the per 48 minute statistic in determining which players contribute the most to a team's win, but this paper examines whether or not the team wins and which statistics proved most insightful.

Table 2. Sample Set of 6 Games

Percentage Input

Away												Home												
FG	3P	FT	Oreb	Dreb	Ast	Stl	Blk	TO	PF	PTS	Away	FG	3P	FT	Oreb	Dreb	Ast	Stl	Blk	TO	PF	PTS	Home	Winner
50.0	46.2	76.5	8	32	15	1	4	17	19	97	0	47.1	25.0	69.2	12	28	21	8	4	9	19	106	1	2
52.2	45.5	83.3	16	40	24	9	7	19	26	117	0	41.6	26.1	68.4	7	30	19	8	9	20	28	96	1	1
45.9	27.3	67.7	12	37	23	10	5	19	30	95	0	42.1	25.0	60.0	11	26	18	16	3	12	22	93	1	1
36.4	30.0	71.1	23	33	15	6	6	18	30	110	0	41.3	44.8	83.3	14	42	20	12	6	17	33	119	1	2
37.9	23.5	59.1	18	30	16	4	4	15	28	83	0	43.8	54.5	74.3	9	33	15	2	9	10	22	102	1	2
45.8	41.2	60.9	16	31	22	3	6	18	17	97	0	49.4	56.3	88.2	8	27	23	8	6	11	20	106	1	2

No set of statistics provides the exact knowledge on which team will be victorious. Fantasy basketball league participants are constantly in search of statistics that provide a success almost every time. One of the most common statistics used is a formula that calculates a player's or team's efficiency rating (Whitling, 2007). This formula encompasses 9 of the 11 box score statistics collected in this research. In this research, different subsets are created based only on shooting, defense, and efficiency.

To produce features for the prediction of "un-played" games, current season averages were used. The same features that are collected for the training

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

set are used in the prediction or validation set of the neural network. A specific team's average statistics typically provide insight into performance. This also allows for the model to be easily updated as the season is played. In this case, the future games that have been played are entered into the training set and the new current season averages are used for the games not yet played.

Two other schemes, with respect to predicting the "un-played" games, were also applied to the model but were found to be no more beneficial than the overall current season averages. The first technique employed the current season average of each team with respect to being either at home or away. This information provided insight into a team's performance in the home stand as well as in an opponent's arena. The second technique used only the average of the previous five games played by each team. This data set allowed considering those teams that might be on hot or cold streaks. It also attempts to capture any devastating injuries or sicknesses.

The final data collected were the experts' opinions on who would win. Numerous experts exist in the field and any could be chosen. If one was interested in betting on the game, the Las Vegas Line might be considered the expert to beat. Rather than betting on the line, this research used the experts' opinions as to who was favored to win. The experts' opinions were derived from USA Today (2008) by examining the favorite to win each game. Figure 2 displays a typical layout of the sports information odds page provided by USA Today (2008). The matchup is printed with the home team always on the bottom. As discussed previously, other alternatives in measuring the success of an expert's opinion exist, but using the favorite and underdog proved to be the most common method. Any one of the five betting sources could be used, but in this research, the team favored by a majority of the experts was considered the favorite. The item that is of interest is the highlighted number. Only the sign of a highlighted number was considered. A negative sign indicates the home team is the favorite to win while a positive number indicates the away team is the favorite. For instance, Figure 2 reflects that Detroit and San Antonio are the experts' opinions on who should win.

DOI: 10.2202/1559-0410.1156

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

5/26 8:30 PM ET	Opening		betED		Sportsbook		5Dimes		SPORTSBETTING	
Boston	174.5o	-110	176o	-110	176o	-110	176o	-107	176o	-110
Detroit	-5.5	-110	-6.5	-110	-6	-110	-6.5	100	-6	-110

5/27 9:00 PM ET	Opening		betED		Sportsbook		5Dimes		SPORTSBETTING	
L.A. Lakers	192o	-110	192o	-110	192o	-110	192o	-110	192o	-110
San Antonio	-4.5	-110	-4.5	-110	-4	-110	-4	-110	-4	-110

Figure 2. Typical Odds Page Layout (USA Today, 2008)

3.2. Dimensionality Reduction: Preprocessing

When examining a high dimensional data set it is desired to reduce the dimensionality of the data set without sacrificing useful information. Such reductions can reduce the noise level in data and eliminate redundant features. When dealing with a large data set, several pre-processing and post-processing techniques are available for use in reducing the dimensionality. One of the most popular pre-processing methods is Principal Component Analysis (PCA). This technique reduces the dimension of the data set without losing any of the intrinsic information provided by the data set (Zhang, 2000). Other pre-processing dimensionality reduction techniques such as Factor Analysis exist. However, our datasets were not of high dimension so we endeavored to employ feature selection on the backend of the analysis to gain insight into the important discriminators. The next section will present a post-processing technique used in this research.

3.3. Neural Networks

Neural networks are a powerful tool in pattern recognition (Verikas et al., 2002). Computers have advanced tremendously in recent years allowing for quicker results to be obtained from neural networks using massive data sets. The elements upon which a neural network is constructed is three-fold: The structure of the network, the training method, and the activation function (Purucker, 1996).

Four different neural networks are examined in this research. The first neural network examined was the feed-forward network. We applied a log-sigmoid transfer function, given as Equation 1 below, as the activation function (Matlab©).

$$\log sig(n) = \frac{1}{(1+e^{-n})} \tag{1}$$

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

To construct the network, one hidden layer was created. Accounting for the number of hidden neurons within this layer is an "art" but some research has been accomplished in this area; see for instance (Steppe, et al., 1996). Most researchers appear to resort to simple trial and error, as did we, given the relatively small dimensionality of our problem. A large disadvantage of the feed-forward network is its relatively slow training. Figure 3 provides a graphical representation of a typical neural network. A noise node is added to the input layer, which is primarily used in the SNR feature selection method explained below.

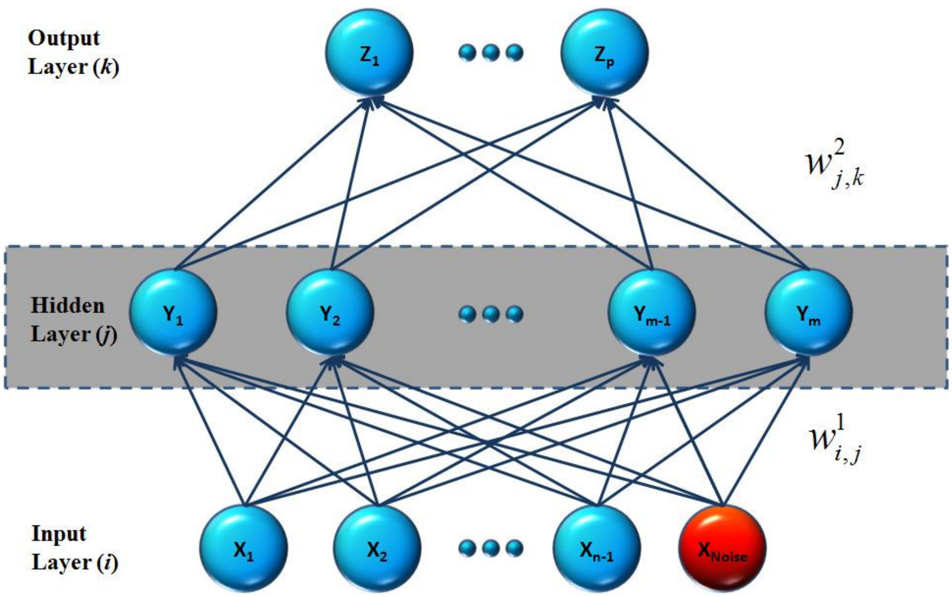

Figure 3. Example Neural Network with Noise Node

After feature data has been processed by a feed-forward neural network, post-processing via a feature selection technique is often applied. The feature reduction technique primarily applied in this research is known as the Signal-to-Noise Ratio (SNR) method. The SNR method examines the lower level weights of feed-forward neural network and computes a saliency measure for each feature. The higher the saliency measure, the more important role the particular feature plays. Bauer et al. (2000) proposed a ranking technique that augments a noise feature to the data set and determines which features play a more prominent role in the data set than random noise. This is effected through Equation 2 below:

DOI: 10.2202/1559-0410.1156

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

$$SNR_i = 10log_{10}\left(\frac{\sum_{j=1}^{J}(w_{i,j}^{1})^2}{\sum_{j=1}^{J}(w_{Noise,j}^{1})^2}\right) \tag{2}$$

where SNR_i is the value of the saliency measure for feature *i* and J represents the number of hidden nodes. All the weights are first layer weights and go from either node i to node j or from the noise node to node j. The use of $10\log_{10}$ of the summations transforms the saliency measure to a decibel scale. After training, the SNR is calculated for each feature. The feature with the smallest SNR is eliminated from the data set. This process is repeated until a smaller subset of the entire data set is generated without losing too much information.

Radial basis functions (RBFs) are a newer and more powerful network than the feed-forward approach; however, the two are similar (Looney, 1997: 96). Neurons in this network are used to create a smooth function to represent the data set. A network is produced with as many hidden neurons as there are feature vectors in the training set. These networks require some user provided information such as a parameter called the 'spread". Determining the spread, or, for that matter, the number of neurons in the network, is usually determined by trial and error or using heuristic procedures. Probabilistic neural networks (PNNs) and generalized neural networks (GRNNs) are types of radial basis networks and were also applied in this work.

3.4. Fusion

A single neural network can provide useful and powerful results in classifying new exemplars based on a training set. However, the use of fusion can help one neural network complement another neural network. The idea is to integrate the contributions of each neural network to obtain a new decision context that one would hope to be more accurate than the networks operating separately. Figure 4 provides a layout of a typical fusion technique.

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

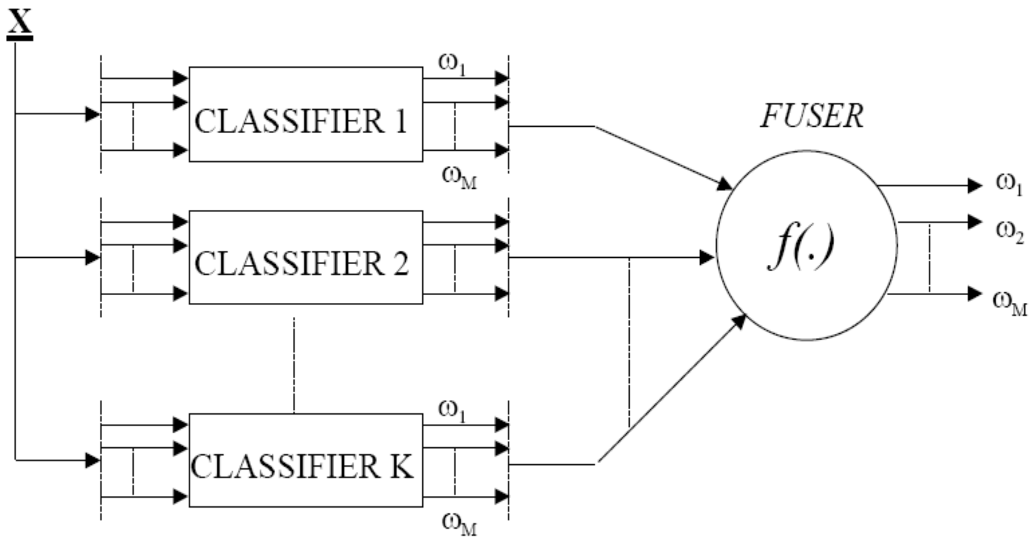

Figure 4. Integration of different classifiers (Rolli, 2002)

There are a variety of fusion rules available (Rolli, 2002). In this research, two different fusion rules were applied The first fusion method applied was the Bayesian belief network (BBN). Rodriguez et al. (2008) provide an explanation on how Bayesian model averaging (here the term refers to a BBN) merges several multi-class classifiers. The Bayes Net Toolbox (Murphy, 2007) for Matlab© was used to perform the computations. The user specifies the local conditional probability distribution (CPD) for a classification model, M_k, where k is one of K classifiers and M is the set of all classifiers. The CPD of each model M_k is $p(M_k|T)$, which represents the probability that a classification model will classify a target instance T. For example, given a target that is a home team win (or an away team win), $p(M_k|T = \textit{Home Team Win})$ represents the probability distribution over all of the possible classifications M_k could make, i.e., home team win or away team win. In our implementation, the confusion matrix, which represents the correct and incorrect classifications for a multi-class classifier, provides this information for each classifier.

The fusion process uses the classifications from the classification models (M), in conjunction with Bayes Rule, to compute the posterior probability for each target classification $T = c$ (c = home team win or away team win):

$$p(T = c \mid M) = \prod_{k=1}^{K} p(M_k \mid T = c)\, p(T = c)$$

The final classification is designated as the target classification $T = c$ with the highest probability. The prior probabilities $p(T)$ are calculated based on the number of home/away team wins used in the testing.

DOI: 10.2202/1559-0410.1156

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

Figure 5 illustrates an example Bayesian system used in this research. The four nodes at the top represent the classifiers and CPDs for each M_k. The Bayesian model node enables the merger of the results of the four models and makes the final classification.

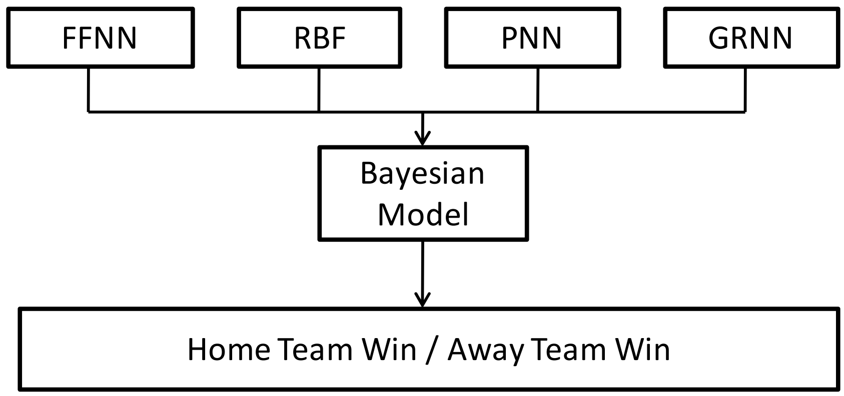

Figure 5. Bayesian Model Structure

The second fusion rule scheme in this research was probabilistic neural network (PNN) fusion (Leap et al., 2007). In this fusion technique, posterior probabilities are obtained from each neural network and then fused, as features, to a new PNN.

To conduct the analysis for this research, the neural network toolbox provided in Matlab© is used as a primary tool along with user developed code for fusion techniques. Matlab© provides further documentation in the understanding the workings of each neural network and instructions for implementation.

4. Results

The four neural networks discussed in the previous section were applied to the entire data to set a baseline for comparison. These networks were the FFNN, the RBF, the PNN and GRNN, respectively. 620 games were used to train and 30 "un-played" games were the validation set. This differs from subsequent fusion techniques in that the 620 games were split up into a training set and a test set. The test set was necessary to create posterior probabilities for the fusion rules. All four neural networks were used as classifiers for both fusion methods.

For purposes of the analysis, 10 different training and validation sets were created to obtain a fairly accurate estimate of each neural network's performance. The first training set and validation set were constructed deliberately to represent the first 620 games and then the next 30 games respectively. This allows for the idea that one would be able to predict the next week's game winners based on the

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

previous two months worth of games played. The next nine training and validation sets were constructed randomly using the same 650 games. However, no game was used more than once in the validation sets. This 10-fold cross validation should provide accurate estimates of the neural network performance. Results from the 10-fold average will be presented as well as results from the first validation set.

For each validation set, an average is gained from expert opinion to determine how well the neural networks compare against expert opinion. For the first validation set, the experts correctly picked 70% of the ultimate winners. This means that the experts incorrectly "guessed" 9 games out of the possible 30. In terms of all 10 experiments run, the experts were correct 68.67% of the time.

The following sections report the results for three different sets of analyses. The first set details the neural network analysis, including fusion, using all the statistics collected from the box scores as the feature set. The second analysis considers the feature set as suggested by the SNR method. The final analysis considers two different subsets of the feature set, each consisting only of shooting statistics, as suggested by experts. One subset will use all six shooting statistics while the second subset only uses four of the six statistics.

Using the entire feature set (all 22 game statistics as features), a maximum (average) accuracy of 71.67% was achieved for the baseline testing while the first validation set achieved the same results as the experts. Table 2 summarizes each of the individual neural network results as well as the fusion results. Once again, all 620 games were used to train each network and the 30 "un-played" games were used as a validation set to determine the accuracy. For the fusion rules, 400 games were used for the training set and the remaining 120 games were used in the test set to determine posterior probabilities. The same 30 "un-played" games were used for the validation set. The results are given in Table 3 and are presented in percentage values.

Table 3. Baseline Results (all 22 features used)

	Validation Sets										
	1	2	3	4	5	6	7	8	9	10	**Average**
FFNN	70	76.67	70	73.33	63.33	70	76.67	73.33	80	63.33	71.67
RBF	66.67	73.33	70	73.33	63.33	63.33	70	73.33	73.33	60	68.67
PNN	70	76.67	70	73.33	63.33	70	76.67	70	80	63.33	71.33
GRNN	70	76.67	70	73.33	63.33	70	76.67	70	80	63.33	71.33
PNN Fusion	70	76.67	70	73.33	63.33	70	76.67	73.33	80	63.33	71.67
Bayes Fusion	70	76.67	70	73.33	63.33	70	76.67	73.33	80	63.33	71.67
Experts	70	73.33	70	73.33	63.33	63.33	70	70	73.33	60	68.67

Following this initial assessment, steps were taken to reduce the dimensionality of the data set in hopes to either reduce the data set size while

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

maintaining similar accuracy, or to reduce the data set size and increase the overall accuracy of the prediction model. The first technique applied was the use of SNR. After implementing SNR on the data, a new subset consisting of TO and PTS for both the away team and the home team was suggested: This allowed us to reduce the number of features to only four variables, showing an emphasis on turnovers and points per game. Overall, the performance for the average accuracy experienced a slight decline from the previous feature set, but each neural network and fusion method tested proved to be better than the expert opinion. Also, in terms of validation set 1, every neural network now achieved the same results as expert opinion. Results for the SNR tests are shown below in Table 4.

Table 4. SNR Results (Features are TO and PTS)

	Validation Sets										
	1	2	3	4	5	6	7	8	9	10	Average
FFNN	70	73.33	76.67	73.33	63.33	70	76.67	63.33	76.67	63.33	70.67
RBF	70	66.67	76.67	63.33	63.33	70	76.67	63.33	76.67	63.33	69.00
PNN	70	73.33	76.67	73.33	63.33	66.67	70	60	76.67	60	69.00
GRNN	70	73.33	76.67	73.33	63.33	66.67	70	60	76.67	60	69.00
PNN Fusion	70	73.33	76.67	73.33	63.33	70	76.67	63.33	76.67	63.33	70.67
Bayes Fusion	70	73.33	76.67	73.33	63.33	70	76.67	63.33	76.67	63.33	70.67
Experts	70	73.33	70	73.33	63.33	63.33	70	70	73.33	60	68.67

Another data reduction based simply on using shooting statistics, as suggested by different experts (Whitling, 2007; Page et al., 2007), which infers a good offense wins basketball games, was made and tested. Here, only FG, 3P, and FT were used for each team, reducing the entire data set to just 6 variables. The results obtained from this subset will be used in comparison to the experts' opinions on the future games as well as the baseline and SNR results. Table 5 describes the optimal accuracy obtained from each network and fusor using only the 6 shooting statistics.

Table 5. Shooting Results (using 6 variables)

	Validation Sets										
	1	2	3	4	5	6	7	8	9	10	Average
FFNN	80	73.33	76.67	66.67	66.67	76.67	70	66.67	83.33	66.67	72.67
RBF	73.33	73.33	73.33	66.67	63.33	70	63.33	60	76.67	60	68.00
PNN	80	73.33	76.67	63.33	66.67	76.67	70	66.67	83.33	66.67	72.33
GRNN	80	73.33	76.67	63.33	66.67	76.67	70	66.67	83.33	66.67	72.33
PNN Fusion	80	73.33	76.67	66.67	66.67	76.67	70	66.67	83.33	66.67	72.67
Bayes Fusion	80	73.33	76.67	66.67	66.67	76.67	70	66.67	83.33	66.67	72.67
Experts	70	73.33	70	73.33	63.33	63.33	70	70	73.33	60	68.67

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

As seen by the results in Table 5, every neural network and fusion method, with the exception of RBF, saw around a four percent increase in accuracy over the 10 validation sets compared to the experts. Also, each of these networks and fusion methods saw an increase compared to using all the features or the SNR feature set. It should also be noted that in validation set 1, all the neural networks and fusion methods performed better than the experts (most by 10%).

Further reduction was performed on the six-variable data set. Factor Analysis was conducted and it appeared that the true dimensionality of this subset was four. FG and 3P for both the home and way team are thought of as one variable, with an emphasis placed on FG due to a heavier loading. This is expected since the field goal percentage is affected by the three point percentage. Using this information, the data set was further reduced to only include FG and FT for each team. In this case, all the neural networks and fusion methods saw an increase in performance, most notably the RBF. Table 6 presents the results obtained using only the four shooting variables in the feature set.

Table 6. Shooting Results (using 4 variables)

	Validation Sets										
	1	2	3	4	5	6	7	8	9	10	Average
FFNN	83.33	76.67	73.33	66.67	70	80	76.67	76.67	76.67	63.33	74.33
RBF	70	70	73.33	66.67	70	80	73.33	76.67	76.67	63.33	72.00
PNN	83.33	76.67	66.67	66.67	66.67	80	76.67	76.67	76.67	63.33	73.34
GRNN	83.33	76.67	66.67	66.67	66.67	80	76.67	76.67	76.67	63.33	73.34
PNN Fusion	76.67	76.67	70	66.67	70	80	73.33	76.67	73.33	63.33	72.67
Bayes Fusion	80	76.67	73.33	66.67	70	80	76.67	76.67	76.67	63.33	74.00
Experts	70	73.33	70	73.33	63.33	63.33	70	70	73.33	60	68.67

It is also important to note that in validation set 1, an increase of 13.33% was seen achieving the highest accuracy of all feature sets tested. This accuracy reflects that only five games were misclassified out of all 30 “un-played” games. In fact, all five of these games are victories by the away team and four of them were considered *upsets* (in terms of records) by the experts.

To help better understand the results, Table 7 examines the results of the PNN and GRNN in terms of validation set 1 and using 6 shooting variables as the feature set. As seen in the table, three of the games were upsets, one game contained two even teams (by expert opinions and records), and the other two games were simply misclassified (most likely since the away team won).

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

Table 7. Missed Games

Game	Winner	Reason	Game	Winner	Reason	Game	Winner	Reason
SA v MI	1		SAC v SEA	1		NJN v DEN	2	
PX v CLE	1		IN v MIL	2		SAC v UT	2	
CTT v CHI	1	Even	NJ v GS	2		IND v MIA	2	
HOU v POR	1	Upset	ORL v DET	2		LAC v MEM	2	
ATL v SEA	1	Missed	MEM v WAS	2		BOS v ORL	2	
PHI v CTT	1	Upset	MIL v TOR	2		WAS v MIL	2	
NO v SA	1	Missed	MIN v BOS	2		ATL v POR	2	
PHX v CHI	1		PHI v NYK	2		DEN v DAL	2	
CLE v LAL	1	Upset	LAC v NO	2		NJN v MIN	2	
UT v HOU	1		LAL v DAL	2		NYK v GS	2	

5. Conclusions

As seen in the results sections, FFNNs, PNNs and GRNNs appear to be good neural network models to use in predicting the outcome of future NBA games. Overall, all four networks were able to match or perform better than the experts' opinions/models.

Table 8 summarizes the use of neural network models coupled with fusion, with respect to all tested data sets in terms of the average of all 10 validation sets and Table 9 summarizes the results using validation set 1. Even though fusion did not yield a better accuracy than the networks alone, its use could still be applied in future research. The fusion model may not have provided better results due to the small size of the validation set, as well as less data to train. Very few games were misclassified using the optimal data set and fusion was unable to correctly classify these games due to small numbers. It is also important to note that the optimal accuracy obtained in this research was using only four statistics to obtain an average accuracy of 74.33% (using FFNN) and in terms of validation set 1, an optimal of 83.33% was achieved.

Table 8. Summary (Average Accuracy)

	Feature Set			
	Baseline (all)	**SNR (4 dim)**	**Shooting (6)**	**Shooting (4)**
FFNN	71.67	70.67	72.67	74.33
RBF	68.67	69.00	68.00	72.00
PNN	71.33	69.00	72.33	73.34
GRNN	71.33	69.00	72.33	73.34
PNN Fusion	71.67	70.67	72.67	72.24
Bayes Fusion	71.67	70.67	72.67	71.57
Experts	68.67	68.67	68.67	68.67

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

Table 9. Summary (Validation Set 1)

	Feature Set			
	Baseline (all)	SNR (4 dim)	Shooting (6)	Shooting (4)
FFNN	70	70	80	83.33
RBF	66.67	70	73.33	70
PNN	70	70	80	83.33
GRNN	70	70	80	83.33
PNN Fusion	70	70	80	76.67
Bayes Fusion	70	70	80	80
Experts	70	70	70	70

The neural networks in this research were capable of using common box score statistics to accurately classify the outcome of an "un-played" game. Extensions to this research can be made off the baseline model to add or delete more features. The models created could also be adjusted to determine if the classification can beat the Las Vegas Line rather than classify which team will win. Overall, these models have shown that they can achieve up to a 5.66 percent better (average) accuracy using FFNN of only four variables or a 13.33 percent better accuracy (in validation set 1) than the experts in the sport of basketball.

6. References

Bauer, K.W., Alsing, S.G., and K.A. Greene, Feature Screening Using Signal-to-Noise Ratios, *Neurocomputing*, 2000, 31, 29-44.

Calster, B.V., Smits, T., and S.V. Huffel, The Curse of Scoreless Draws in Soccer: The Relationship with a Team's Offensive, Defensive, and Ooverall Performance, *Journal of Quantitative Analysis in Sports*, 2008, 4, 1, article 4.

Ibisworld Inc., United States Basketball League, Incorporated- Company Report US, 2008, available on http://www1.ibisworld.com/enterprise/retail.aspx?entid=51530&chid=1&rcid=1, (accessed on May 25, 2008).

Kahn, J., Neural Network Prediction of NFL Football Games, 2003, available on http://homepages.cae.wisc.edu/~ece539/project/f03/kahn.pdf, (accessed on April 1, 2008).

Leap, N.J., Clemans, P.P., Bauer, K.W., and Oxley, M.E., An Investigation of the Effects of Correlation and Autocorrelation on Classifier Fusion and Optimal Classifier Ensembles, *International Journal of General Systems-Intelligent Systems Design*, 2007, forthcoming.

Looney, C.G., *Pattern Recognition Using Neural Networks*, 1997, Oxford University Press, New York.

DOI: 10.2202/1559-0410.1156

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM

Murphy, K., Bayes Net Toolbox for Matlab, 2007, available on www.cs.ubc.ca/_murphyk/Software/BNT/bnt.html.

Naismith Museum and Hall of Fame, n.d, available on http://www.naismithmuseum.com/naismith_drjamesnaismith/main_drjamesnaismith.htm, (accessed May 25, 2008).

NBA Scoreboard, 2008, available on http://scores.espn.go.com/nba/scoreboard.

Page, G.L., Fellingham, G.W., and Reese, C.S., Using Box-Scores to Determine a Position's Contribution to Winning Basketball Games, *Journal of Quantitative Analysis in Sports*, 2007, 3, 4, article 1.

Purucker, M.C, Neural Network Quarterbacking, 1996, *IEEE Potentials*.

Rodriguez, B.M., Peterson, G.L., and Bauer, K.W., Fusion of Steganalysis Systems Using Bayesian Model Averaging, *Advances in Digital Forensics IV*, eds. Shenoi, S., and Ray, I., 2008, forthcoming.

Rolli, F., Fusion of Multiple Classifiers, *5th International Conference on Information Fusion*, July 2002, Annapolis MD.

Steppe, J. M., Bauer, K. W., and Rogers, S. K., Integrated Feature and Architecture Selection, *IEEE Transactions on Neural Networks*, 1996, 7, 4, 1007-1014.

Game Guide- Live Odds, 2008, available on http://covers.usatoday.com/data/odds.aspx#nba.

Verikas, A. and Bacuskiene, M., Feature Selection with Neural Networks, *Pattern Recognition Letters*, 2002, 23.

Whitling, J., Inside Obscure Stats, Oct. 18, 2007, available on http://www.nba.com/fantasy/features/obscure_051018.html?rss=true, (accessed on May 25, 2008).

Zhang, G.P., Neural Networks for Classification: A Survey, *IEEE Transactions on Systems, Man, and Cybernetics- Part C: Applications and Reviews*, 2000, 30, 4.

Brought to you by | Cambridge University Library
Authenticated
Download Date | 11/9/15 1:22 PM