International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:7, No:7, 2013

An Investigation into the Application of Artificial Neural Networks to the Prediction of Injuries in Sport

J. McCullagh and T. Whitfort

***Abstract*—**Artificial Neural Networks (ANNs) have been used successfully in many scientific, industrial and business domains as a method for extracting knowledge from vast amounts of data. However the use of ANN techniques in the sporting domain has been limited. In professional sport, data is stored on many aspects of teams, games, training and players. Sporting organizations have begun to realize that there is a wealth of untapped knowledge contained in the data and there is great interest in techniques to utilize this data. This study will use player data from the elite Australian Football League (AFL) competition to train and test ANNs with the aim to predict the onset of injuries. The results demonstrate that an accuracy of 82.9% was achieved by the ANNs' predictions across all examples with 94.5% of all injuries correctly predicted. These initial findings suggest that ANNs may have the potential to assist sporting clubs in the prediction of injuries.

***Keywords*—**Artificial Neural Networks (ANNs), data, injuries, sport.

I. INTRODUCTION

IT is the coaches and sports scientists' job to train athletes at the optimal level to produce peak performance while ensuring that they do not over train and risk the onset of injury (Fig. 1). Numerous studies [1]-[5] have involved investigating various types of training, assessing training loads, monitoring physiological and psychometric data and using these factors to find a relationship with the injury rates. Many of these studies reported an increase in the frequency of injuries with an increased player workload. Gabbett and Jenkins [3] demonstrated that training load was significantly related to overall injury rates in professional rugby league players and that high strength and power training loads may contribute indirectly to field injuries. Dennis et al. [5] monitored workloads and occurrence of injuries of forty four Australian junior cricket fast bowlers and found that injured bowlers were bowling significantly more frequently than uninjured bowlers. Faigenbaum [4] reported a number of risk factors related to the onset of injury for youths undertaking resistance training.

These included excessive load and volume, poor exercise technique, muscle imbalances and inadequate recovery. The most frequent injury in adolescent athletes who participated in a resistance training program was low back pain.

Despite the strong relationship between excessive training loads and the occurrence of injury, there is evidence that under training can increase the risk of injury. Shaw et al. [6] demonstrated that the likelihood of sustaining an injury is least when training for a total of 8 to 10 hours per week for non-elite tri-athletes, and that athletes training at lower levels and higher levels than this were more likely to sustain an injury. A study on sub elite rugby players [7] also found an increased injury risk occurred with reduced aerobic fitness. The study highlighted the importance of an appropriate pre-season preparation to provide athletes with adequate physiological capacities to endure the demands of competition.

Fig. 1 Hamstring injury in AFL

ANNs (Artificial Neural Networks) are well suited to problems with large amounts of data and complex relationships between inputs and outputs. Such problems are often difficult to solve using traditional techniques. Numerous studies have demonstrated the effectiveness of using ANN techniques to estimate various parameters in a variety of domains [8]-[10]. While there have been many successful applications of ANNs, their use in the sporting domain has been limited. ANNs are well suited to the prediction of injuries in sport due to the large amount of data available and the complex mapping between these factors and the occurrence of injury. One of the main reasons sporting organizations have been slow to use advanced methods such as ANNs to improve predictions is that such methods can be difficult to master and require considerable expert knowledge to successfully apply. The following studies describe some of the ANN applications from the sporting domain. The applications take a variety of forms including prediction of outcomes, talent identification, evaluation of game strategies, and prediction of training loads, injuries, team and individual performance.

Dr J. McCullagh is Senior Lecturer in Physical Education at La Trobe University, PO Box 199 Bendigo, Australia 3550 (phone:+61354447800 fax:+61354447800, e-mail: j.mccullagh@latrobe.edu.au).

T. Whitfort is a Lecturer in Information Technology in the Science and Technology Faculty at La Trobe University PO Box 199 Bendigo, Australia 3550 (e-mail: t.whitfort@latrobe.edu.au).

Kahn [11] examined the predictive power of ANNs to forecast future US National Football League team performance based on past performance. Factors such as time in possession, total yardage, rushing yardage and turnover differential were used to train the network. The ANN produced an output for each team's performance and achieved an accuracy of 75%, outperforming the experts. ANNs have also been used successfully in college football as an alternative to the traditionally used techniques to rank football teams [12]. The ANN model offered enough flexibility in its parameter settings to take into account many of the different factors involved in the decision making process.

Silva et al [13] used ANN technology to model swimming performance. A sample of 138 young swimmers of national level was submitted to a test battery including strength and flexibility, swimming functional and technique evaluation and kin-anthropometric evaluation. A mapping between these factors and performance in the 200 medley and 400 crawl were developed. The testing results indicated that an error lower than 0.8% was achieved between the true and estimated performances demonstrating that the neural network approach can offer a resolution to complex problems such as performance modeling and talent identification in swimming.

McCullagh and Whitfort [14] used ANNs to investigate the relationships between anthropometric and fitness tests from the Australian Football League (AFL) Draft camp and the career progression of these players in AFL Football. The results demonstrated that ANNs have the potential to provide information to enhance the selection process. A neural network program called Advanced Scout was used to seek out interesting patterns in basketball game data. The information derived proved successful in assisting coaches to assess game strategy as well as formulate game plans for future games [15].

Passos et al. [16] used ANNs to reconstruct the 3D performance space in a typical rugby pattern of play. The results suggest that ANNs may be instrumental in identifying pattern formations is team sport and that they might be a suitable method for improving the analysis of decision making on sports tasks.

This study will involve assessing the injury risk of players competing in Australian Rules Football. Australian Rules Football consists of two teams of 22 players (18 playing and 4 interchange). It is a contact sport in which players can tackle opponents using their arms or whole body. Frequent physical contests and fast movement of both players and the ball ensure that the physical demands on the players are high. Players are required to have great speed, endurance and strength to enable them to compete at the elite level in the AFL national competition. The aim of this study is to investigate the applicability of ANNs to assist in the prevention of injury in Australian Rules Football.

II. Artificial Neural Networks

ANNs are loosely modelled on the human brain. One type of ANN used to model many real life situations is the back propagation neural network. A back propagation neural network is a multi-layered non-linear feed-forward network trained by the back propagation learning algorithm. It contains an input layer, one or more hidden layers and an output layer. Each layer consists of one or more nodes (artificial neurons). All nodes (other than the input nodes) generate an internal activation which is calculated by summing all of the input weight products as outlined in (1).

$$Activation = \sum_{i=1}^{n} w_i x_i + w_B \quad (1)$$

where $x_1 ... x_n$: are the inputs, $w_1 ... w_n$: are the associated connection weights, w_B: is the bias weight.

A bias acts exactly as a weight on a connection from a node whose activation is always one. Increasing the bias increases the net input to the unit [17]. The structure of the network including the number of nodes used, their organization into layers and the connections between them is referred to as its architecture. The most common architecture is the fully interconnected multi-layered network (see Fig. 2). The back propagation neural network is trained by being presented with numerous examples where each example consists of a set of inputs and their corresponding output(s). The "learning" that takes place in back propagation neural networks occurs by the adjustment of the connection weights which are represented by lines in between the nodes in Fig. 2 [18].The weights are adjusted to minimize the error between the neural network outputs and the correct output values.

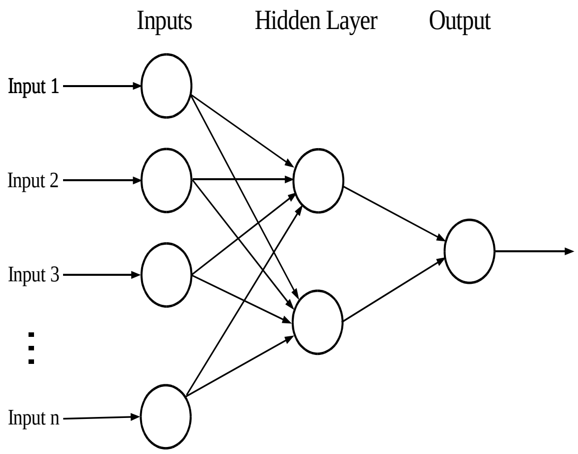

Fig. 2 ANN showing inputs, a hidden layer and output

Many factors influence the success that can be achieved by a back propagation neural network on a particular problem, this includes the architecture and the parameters. The architecture chosen for a network is influenced by the complexity of the mapping between inputs and outputs. A complex mapping will often require a greater number of hidden nodes in the architecture. A number of parameters must be specified before the back propagation neural network can be used, including the momentum, learning rate, initial weight size, epoch size and the number of passes. The weights in a back propagation neural network indicate the strength of the connections between neurons and represent the learning that has been achieved [19]. The number of passes indicates how many times the training data is presented to the network. The network may be set to run the full number of passes or an error criterion may be used to stop the network. The ANNs used in this study will be back propagation neural networks.

An example presented to the ANN represents a particular instance of interest such as a player's attributes on a particular week. Each example used by an ANN consist of inputs and one or more outputs. The inputs are attributes that may be relevant to predicting the output(s). For example a players workload and squeeze test score for a week (inputs) may be relevant to predicting whether the player is likely to be injured or not (the output).

The data used by an ANN is divided into training and testing examples, with each example randomly allocated to one of the two data sets. The ANN is trained on the training examples, where the correct result (output) is known and this is used to guide the learning of the ANN. The ANN aims to minimize the errors between the expected and predicted outputs. Once the ANN has been trained the testing set is used to evaluate the performance of the ANN on unseen data.

As this technique may be susceptible to the allocation of examples to a particular dataset (training or testing), cross validation, a form of resampling is used. In ten-fold cross validation the data is divided into ten approximately equal sized datasets. In the first run 9 datasets (datasets 1 to 9) are combined to form the training dataset and the remaining dataset (dataset 10) is used for testing. In the second run datasets (1 to 8 & 10) are used for training and the remaining dataset (dataset 9) is used for testing etc. In all, 10 runs are performed, with each dataset and example used 9 times for training and once for testing. The testing results are then averaged to give a measure of the performance on unseen data.

III. Experimental Design

A. Player Data

The data set contains player data from the pre-season and in-season elite AFL competition for the 2010 season. Thirty nine players were included in the data set resulting in a total of 1210 player examples (163 injured examples, 1047 not injured examples). The data recorded for each player include factors such as workloads, squeeze test data, soft tissue score, stress level, mood, sleep score, ankle flexibility, fatigue and player perceived performance. These measurements were taken weekly during the pre-season and in season. Other factors such as years played, player durability and age were also included. In total 30 player attributes were selected to input to the ANN.

The injury data recorded for each player includes the injury severity, body area injured, injury category and activity undertaken. A player who is currently injured will be excluded from the data set until they are assessed as able to return to playing games. The official AFL definition of an injury since 1997 has been an “injury or medical condition which causes a player to miss a match” [20]. This definition allows for a consistent classification of injuries without the need for interpretation, however it does not take into account the full spectrum of injuries which may lead to a reduced player performance in training and games. In this study an injury will be defined as a medical condition resulting in a player being placed on a modified training program, missing training sessions or games. In order to simplify the process for this initial study, players will be classified as either HIGH risk or LOW risk depending on ANNs injury prediction. The type or severity of the injury will not be used. The injury risk classifications used by the ANNs are outlined in Table I. It is anticipated that the types and severity of injury, as well as extending the one week prediction window would be included in future studies. An outline of player examples is shown in Table II.

TABLE I
INJURY RISK CLASSIFICATION

Injury Risk Classification	**Description**
HIGH	ANN predicts that the player will be injured in the following week.
LOW	ANN predicts that the player will remain uninjured in the following week.

TABLE II
PLAYER EXAMPLES

Player Number	**Inputs**				**Output**
	Week	*Workload*	*Squeeze Test*	…	*Injury Risk*
23	3	1247	240		LOW
23	4	890	180		HIGH
36	1	1782	290		LOW

B. ANN Design and Parameters

The ANN design is outlined in Fig. 3. It represents the ANN as a black box where a collection of inputs are presented to the neural network and one output is produced. Player attributes such as workloads, squeeze test results, sleep score, ankle flexibility and durability are supplied to the network and an output indicating the injury risk of the player is produced.

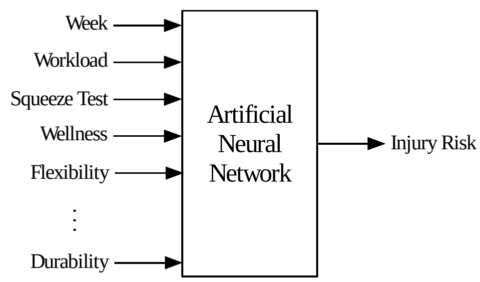

Fig. 3 ANN showing inputs and output for the AFL injury problem

The ANN was trained using the parameters specified in Table III. The parameters were derived by conducting a series of trials involving varying parameters and assessing the effect on the neural network's output. Ten-fold cross validation was used to train and assess the performance of the neural network. The performance was averaged across the ten testing sets.

TABLE III
ANN PARAMETERS

Parameter		Value
Architecture	Inputs	30
	Hidden Layer	15
	Outputs	1
Learning rate		0.3
Momentum		0.2
Epochs		500

C. Limitations to This Study

Due to the complex nature of the neural network development process, the researchers have decided to only use data from a single time instance in each player example. Future studies will use time series data which will allow trends across time to be included. For example the squeeze test data for each of the previous 4 weeks can be used as the trends may assist the neural networks prediction accuracy.

IV. RESULTS AND DISCUSSION

Table IV shows the percentage correct results for the LOW, HIGH and OVERALL classifications. A correct classification is achieved when the ANN predicts a HIGH classification and the player sustains an injury, or when the ANN predicts a LOW classification and the player remains uninjured in the following week. The OVERALL percent correct is the accuracy of the ANN to predict correctly from all 1210 player examples including injured and not injured examples. Table V shows the frequency for each of the ANN's predictions and their outcome. Actual refers to what did happen to the player the following week in terms of injury.

TABLE IV
ANN PREDICTIONS (PERCENTAGE CORRECT)

Injury Risk Classification	Percentage correct
HIGH	94.5%
LOW	81.1%
OVERALL	82.9%

TABLE V
ANN PREDICTIONS (FREQUENCY)

Actual	ANN Risk Prediction	Frequency
Injured	HIGH	154 (correct)
	LOW	9
Not Injured	LOW	849 (correct)
	HIGH	198

A percent correct result for the HIGH risk classification indicates that the ANN correctly predicted 94.5% (154 examples) of all of the injured examples occurring in 2010, while 5.5% (9 examples) of injuries occurring were not correctly predicted. It is anticipated that players who record a HIGH injury risk classification would be placed on a modified training/playing program in order to reduce the possibility of injury.

For the LOW risk classification, the ANN correctly predicted 81.1% of all of the not injured cases and made an incorrect prediction on 18.9% of cases. There were 1047 examples of player data where an injury did not occur in the following week. 849 of these were correctly predicted by the neural network whereas 198 were incorrectly predicted. The fact that 18.9% of cases were classified as HIGH risk and did not become injured in the following week is an area that requires improvement. Future studies where the severity of injury is included and the window for injury is increased from one week may assist to improve this result. When combining all of the injured and non-injured examples, an OVERALL injury risk classification of 82.9% was achieved.

Contact injuries have been shown to occur in physical collisions in the AFL and while they are thought to be largely unavoidable [21], the results from this study indicate that players classified as HIGH risk by the ANN are at an increased risk for both contact and non-contact injuries. Fig. 4 shows the percentage of each of the different injury mechanisms for the data used in this study. Contact injuries accounted for approximately 45% of all injuries while non-contact injuries accounted for approximately 55% of injuries.

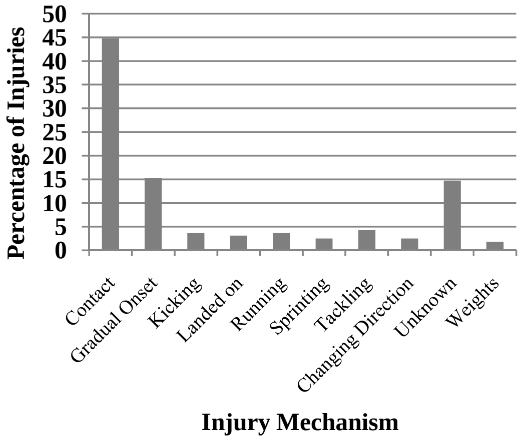

Fig. 4 Percentage of injuries by injury mechanism in AFL

The performance of the ANN on predicting injuries that resulted from contact and non-contact injuries is shown in Table VI. Where injuries resulted from contact, the ANN classified 97.3% of examples as having a HIGH risk of injury. For injuries that were from other causes (non-contact) the ANN classified 92.2% of examples as a HIGH risk of being injured.

TABLE VI
ANN INJURY MECHANISM PREDICTIONS (PERCENTAGE CORRECT)

Type of Injury	Percentage Correct
Contact	97.3%
Non-contact	92.2%

While the results indicate the potential for ANNs to predict both contact and non-contact injuries, further research is required to investigate ANNs which differentiate between different injury types.

V. CONCLUSION

The experiments conducted in this research are preliminary

ISSN: 2517-9969
Vol:7, No:7, 2013

in nature and it is difficult to derive any definite conclusions. However a percentage correct of 94.5% of injured examples correctly classified and overall percentage correct of 82.9% does indicate that ANNs are able to derive meaningful information from the vast amount of data available to assist in the injury prediction process. The percentage correct results of 97.3% and 92.2% for the contact and non-contact injuries respectively indicate the potential for predicting different types of injuries. The results demonstrate that ANNs may be useful as another decision making tool to assess the injury potential of a player along with medical staff, and other internal club techniques.

Future research will investigate increasing the complexity of the ANN to attempt to improve the prediction accuracy. This will involve a number of areas including windowing of the data to allow trends over time to be input to the network and increasing the number of risk classifications (eg: LOW, MEDIUM, HIGH). Specialist neural networks may also assist in improving prediction accuracy as it is likely that players who are different body shapes and play in different positions may exhibit characteristics that indicate a higher injury risk. This may take the form of creating a specialist network for key position players and a separate network for all other players. Further research is being conducted in this area.

References

[1] Kellmann, M., *Preventing overtraining in athletes in high-intensity sports and stress/recovery monitoring.* Scandinavian Journal of Medicine and Science, 2010. 20(s2): p. 95.

[2] Visnes, H. and R. Bahr, *Training volume as a risk factor for the development of jumper's knee? data from the patellar tendinopathy cohort 2006-2010.* British Journal of Sports Medicine, 2011. 45(4): p. 324.

[3] Gabbett, T. and D. Jenkins, *Relationship between training load and injury in professional rugby league players.* Journal of Science and Medicine in Sport, 2011. 14(3): p. 204.

[4] Faigenbaum, A.E.C., G.P.C. Myer, F.P.C. Naclerio, and A.M. Casas, *Injury Trends and Prevention in Youth Resistance Training.* Strength and Conditioning Journal, 2011. 33(3): p. 36.

[5] Dennis, R., C. Finch, and P. Farhart, *Is bowling workload a risk factor for injury to Australian junior cricket fast bowlers?* British Journal of Sports Medicine, 2005. 39(11): p. 843-846.

[6] Shaw, T., P. Howat, M. Trainor, and B. Maycock, *Training patterns and sports injuries in triathletes.* Journal of Science and Medicine in Sport, 2004. 7(4): p. 446-450.

[7] Gabbett, T.J. and N. Domrow, *Relationships between training load, injury, and fitness in sub-elite collision sport athletes.* Journal of sports sciences, 2007. 25(13): p. 1507-1519.

[8] Huang, Z., H. Chen, C.J. Hsu, W.H. Chen, and S. Wu, *Credit rating analysis with support vector machines and neural networks: a market comparative study.* Decision support systems, 2004. 37(4): p. 543-558.

[9] McCullagh, J. *A modular neural network architecture for rainfall estimation*. in *Artificial Intelligence and Applications.* 2005. ACTA Press.

[10] Smith, K. and J. Gupta, *Neural Networks in Business*, 2002, Hershey, PA: IDEA Group Publishing.

[11] Kahn, J., *Neural network prediction of NFL football games.* World Wide Web electronic publication, URL homepages. cae. wisc. edu/~ ece539/project/f03/kahn. pdf, 2003.

[12] Wilson, R.L., *Ranking college football teams: A neural network approach.* Interfaces, 1995. 25(4): p. 44-59.

[13] Silva, A.J., A.M. Costa, P.M. Oliveira, V.M. Reis, J. Saavedra, J. Perl, A. Rouboa, and D.A. Marinho, *The use of neural network technology to model swimming performance.* Journal of Sports Science and Medicine, 2007. 6(1): p. 117-125.

[14] McCullagh, J.A. and T.J. Whitfort, *An Investigation into the Application of Artificial Intelligence Techniques to the Player Selection Process at the AFL National Draft.* 2008.

[15] Bhandari, I., E. Colet, J. Parker, Z. Pines, R. Pratap, and K. Ramanujam, *Advanced scout: Data mining and knowledge discovery in NBA data.* Data Mining and Knowledge Discovery, 1997. 1(1): p. 121-125.

[16] Passos, P., D. Araújo, K. Davids, L. Gouveia, and S. Serpa, *Interpersonal dynamics in sport: The role of artificial neural networks and 3-D analysis.* Behavior research methods, 2006. 38(4): p. 683-691.

[17] Fausett, L.V., *Fundamentals of neural networks: architectures, algorithms, and applications.* 1994: Prentice-Hall Englewood Cliffs, NJ.

[18] Rumelhart, D.E. and J.L. McClelland, *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1. Foundations.* 1986.

[19] McClelland, J.L. and D.E. Rumelhart, *Explorations in Parallel Distributed Processing-Macintosh Version: A Handbook of Models, Programs, and Exercises.* 1989: MIT press.

[20] Orchard, J. and H. Seward, *Injury report 2010: Australian football league.* Sport Health, 2011. 29(2): p. 15.

[21] Gabbett, T., *The Development and Application of an Injury Prediction Model for Noncontact, Soft-tissue Injuries in Elite Collision Sport Athletes.* Journal of Strength and Conditioning Research, 2010. 24(10): p. 2593.