

The application of k -Nearest Neighbour in the identification of high potential archers based on relative psychological coping skills variables

Zahari Taha¹, Rabi Muazu Musa^{1,2}, Anwar P P Abdul Majeed¹, Mohamad Razali Abdullah², Muhammad Muaz Alim¹ and Ahmad Fakhri Ab Nasir¹

¹Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

²Faculty of Applied Social Sciences, Universiti Sultan Zainal Abidin, 21300 Kuala Terengganu, Terengganu, Malaysia

E-mail: rabiumuazu@imamslab.com

Abstract. The present study aims at classifying and predicting high and low potential archers from a collection of psychological coping skills variables trained on different k -Nearest Neighbour (k -NN) kernels. 50 youth archers with the average age and standard deviation of (17.0 ± 0.56) gathered from various archery programmes completed a one end shooting score test. Psychological coping skills inventory which evaluates the archers level of related coping skills were filled out by the archers prior to their shooting tests. k -means cluster analysis was applied to cluster the archers based on their scores on variables assessed k -NN models, i.e. fine, medium, coarse, cosine, cubic and weighted kernel functions, were trained on the psychological variables. The k -means clustered the archers into high psychologically prepared archers (HPPA) and low psychologically prepared archers (LPPA), respectively. It was demonstrated that the cosine k -NN model exhibited good accuracy and precision throughout the exercise with an accuracy of 94% and considerably fewer error rate for the prediction of the HPPA and the LPPA as compared to the rest of the models. The findings of this investigation can be valuable to coaches and sports managers to recognise high potential athletes from the selected psychological coping skills variables examined which would consequently save time and energy during talent identification and development programme.

1. Introduction

Various researchers have endeavoured to provide insights on the impacts of psychological skills to the advancement of athlete's performance in different sports. Abdullah et al. studied the roles of psychological factors on the performance of elite soccer players and established that although, psychological factors alone could not forecast the performance of elite soccer players, but play an essential part in improving their performance [1]. Likewise, Lazarus [2] and Salim et al. [3] inferred that psychological elements such as stress and worry could be a significant constituent in hindering athletes performance. They further emphasised that athletes who are unable to cope with stress and adversity effectively might likely suffer from performance declining and overall psychological well-being issues. Furthermore, Pasaei et al. [4] stated that in the contemporary archery competitions there are presents of various noises, tensions, and anxiety, and as such selective attention through the



application of the psychological coping skills during the competition become paramount for better performance. This implies that the athletes with higher psychological coping skills are likely to outperform their opponents with lesser coping skills.

The employment of machine learning or artificial intelligence has gained popularity in prediction and classification in sports and exercise due to its superiority over conventional means. Bayesian network (BN) analysis was utilised to investigate a semi-professional team's performance and collective efficacy based on a number of psychological features [5]. In a different study, Fuster-Parra et al. [6] also employed BN in correlating negative psychological features with sportive behaviours of young team players in competitive sports.

It is evident from the literature that the employment of machine learning could be a useful tool for prediction as well as classification. The k-Nearest Neighbour (k-NN) is a non-parametric regression and classification method that was initially developed by Fix and Hodges [13]. It is regarded as one of the simplest forms of supervised machine learning algorithms that has been successfully utilised for a number of classification problems in different fields [15–19].

2. Material and methods

2.1. Participants

A complete number of 50 archers were selected to take part in in this study. The participants consisted of 37 male and 13 female youth archers between the age's range of 13-20 with a mean and standard deviation of (17.0 ± 0.56) drawn from different archery programmes in Malaysia. The archers were under a development program for preparing both at university and the state level and consequently targeted to be preferred to state and national archers respectively. Written consent was obtained, and all the archers signed consent forms. All the procedures, protocol, and equipment for this study were authorised by the Research Ethics Board of the Terengganu Sports Institute (ISNT) with a memo number 04-04/T-01/Jid 2.

2.2. Psychological coping skills inventory assessment

The Athletic Coping Skills Inventory (ACSI), an instrument for assessing athlete's psychological skills, produced by Smith et al. [7] was selected. This instrument is considered apt for evaluating the psychological skills of the archers in this study because of its connection to the nature of archery game as an individual sport rather than a team sport. The instrument was distributed to the archers prior to the shooting test, and their responses were compiled and analysed. A simulated shooting competition area was set up, and all the archers' shoot six arrows (one end) over a distance of 50 meters. All the archers were given trials of four arrows shot before recording the final six arrows scores.

2.3. Data analysis

2.3.1. Clustering: k-means cluster analysis. In the current study, the k-means clustering algorithm, which is essentially a type of unsupervised learning, is used to separate the classes of the related performance variables assessed via Orange 2.7. It has been reported in the literature that the k-means clustering method is more reliable than the hierarchical agglomerative clustering in catering large datasets [8]. This is primarily due to the fact that it operates on actual observations rather than the dissimilarity measures employed in hierarchical clustering. The number of clusters, k is selected to be two as the data is consists of HPPA and LPPA.

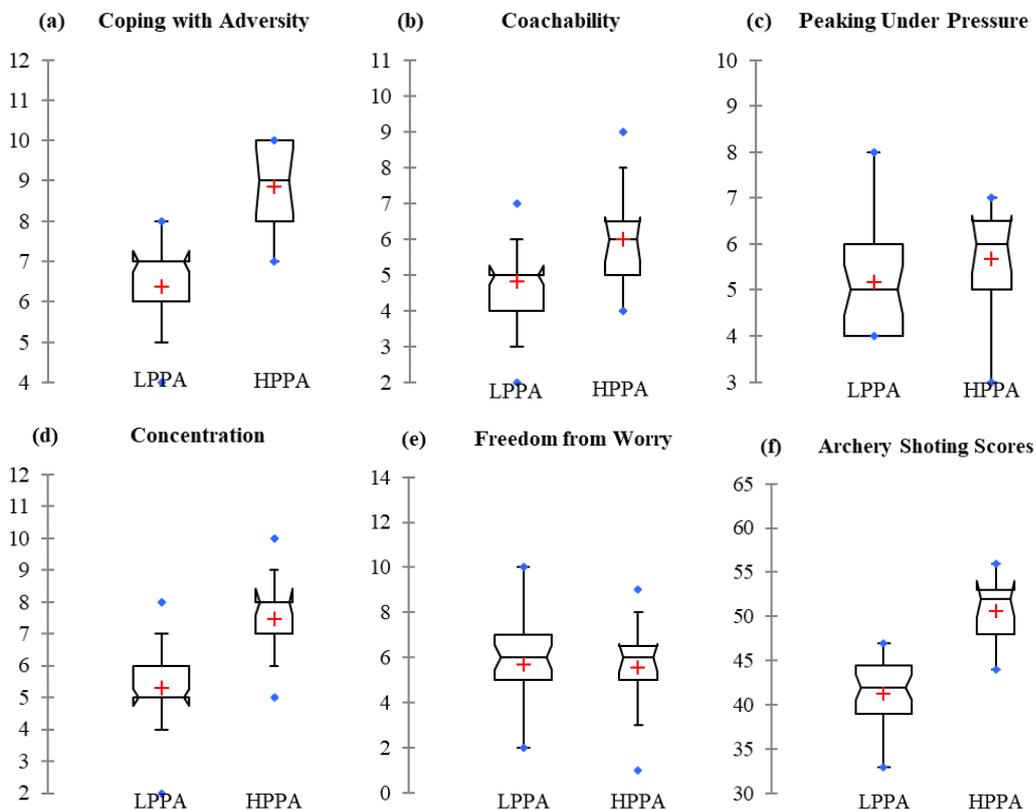
2.3.2. Classification: k-Nearest Neighbour (k-NN). In this study, six variations of k-NN are examined i.e. fine, medium, coarse, cosine, cubic and weighted. The number of neighbours, k for fine and coarse are 1, and 100, respectively whilst for the rest of the variations is selected to be 10. The Euclidean distance metrics is employed in the fine, medium, coarse and weighted k-NN variations, whilst the cubic and cosine utilises a special case of the Minkowski distance and the cosine distance, respectively. The no weights are attributed to the distance for all variations of k-NN with an exception for the weighted k-NN, in which, the weight is the squared inverse of the distance.

2.3.3. Model training and testing. A fivefold cross-validation method was utilised in this study. This form of validation technique is desirable as it mitigates the notion of overfitting through partitioning the dataset into a number of folds and estimating the accuracy of each fold. The data (50 observations) is randomly split into five subsets, and for each iteration, one of the five subsets are used as the testing data, whilst the remaining four will be used as the training data [9]. Then, the average performance over all the folds is then computed. The SVM analysis and assessment were conducted by means of MATLAB 2016a (Mathworks Inc., Natick, USA).

2.3.4. Model evaluation. The variations of the SVM employed in this study are evaluated by means of classification accuracy (ACC), sensitivity (SENS), specificity (SPEC), precision (PREC), error rate (ERR) as well as Matthew's correlation coefficient (MCC). The confusion matrix allows the observation of correctly classified and misclassified observations that transpires between the defined classes.

3. Results

3.1. Clustering



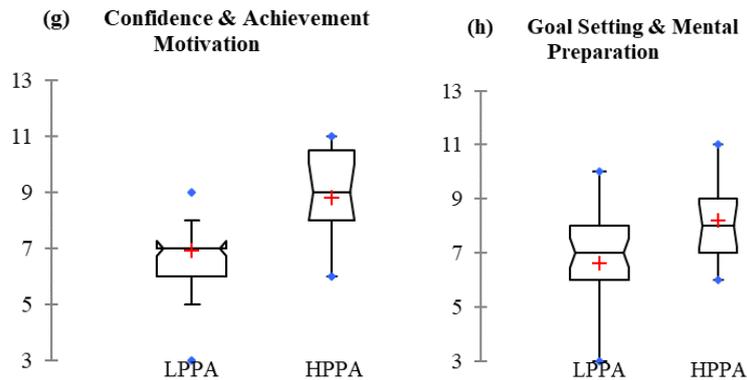


Figure 1. Comparisons of performance differences of the archers based on ACSI (a) Coping with Adversity; (b) Coachability; (c) Peaking Under Pressure; (d) Concentration; (e) Freedom from Worry; (f) Archery shooting score; (g) Confidence and Achievement Motivation; (h) Goal Setting and Mental Preparation.

The performance differences of the archers based on the eight performance variables assessed that were clustered via *k*-means are shown in figure 1. It could be seen from the box plots that the mean performances of HPPA are greater than LPPA across all the seven psychological coping skills measured in the study except for ‘freedom from worry’ (figure 1(e)). This observation is non-trivial as worry is often observed to be inversely proportional towards performance. Therefore, the aforementioned variables are essential attributes that allow for an accurate discrimination between the HPPA and LPPA.

3.2. Classification

Table 1. Model evaluation.

<i>k</i> -NN variation	ACC (%)	SENS (%)	SPEC (%)	PREC (%)	ERR (%)	MCC
Fine	92	97.37	75	92.5	8	0.7729
Medium	86	100	41.67	84.44	14	0.5932
Coarse	76	100	0	76	24	-
Cosine	94	97.87	83.33	94.87	6	0.8320
Cubic	84	100	33.33	82.61	16	0.5248
Weighted	90	100	58.33	88.37	10	0.7180

It could be observed from table 1 that the cosine *k*-NN variation is able to produce exceptional classification through the evaluation of all assessment parameters as compared to the rest of the *k*-NN variations. It is also apparent that the fine-based model could offer a reasonably accurate classification with an error rate of 8 %. Conversely, the coarse *k*-NN model is rather unsuitable for predicting the correct classification of HPPA and LPPA, as it produces the highest error rate at 24% on top of a poor negative class accuracy shown via the SPEC evaluation that in turn produces an uncorrelated model as demonstrated through the MCC. It is evident through the present study that the cosine-based model may be used for the purpose of talent identification for archers based on the predefined psychological coping skills. The confusion matrix of the evaluated *k*-NN models are illustrated in figure 2.

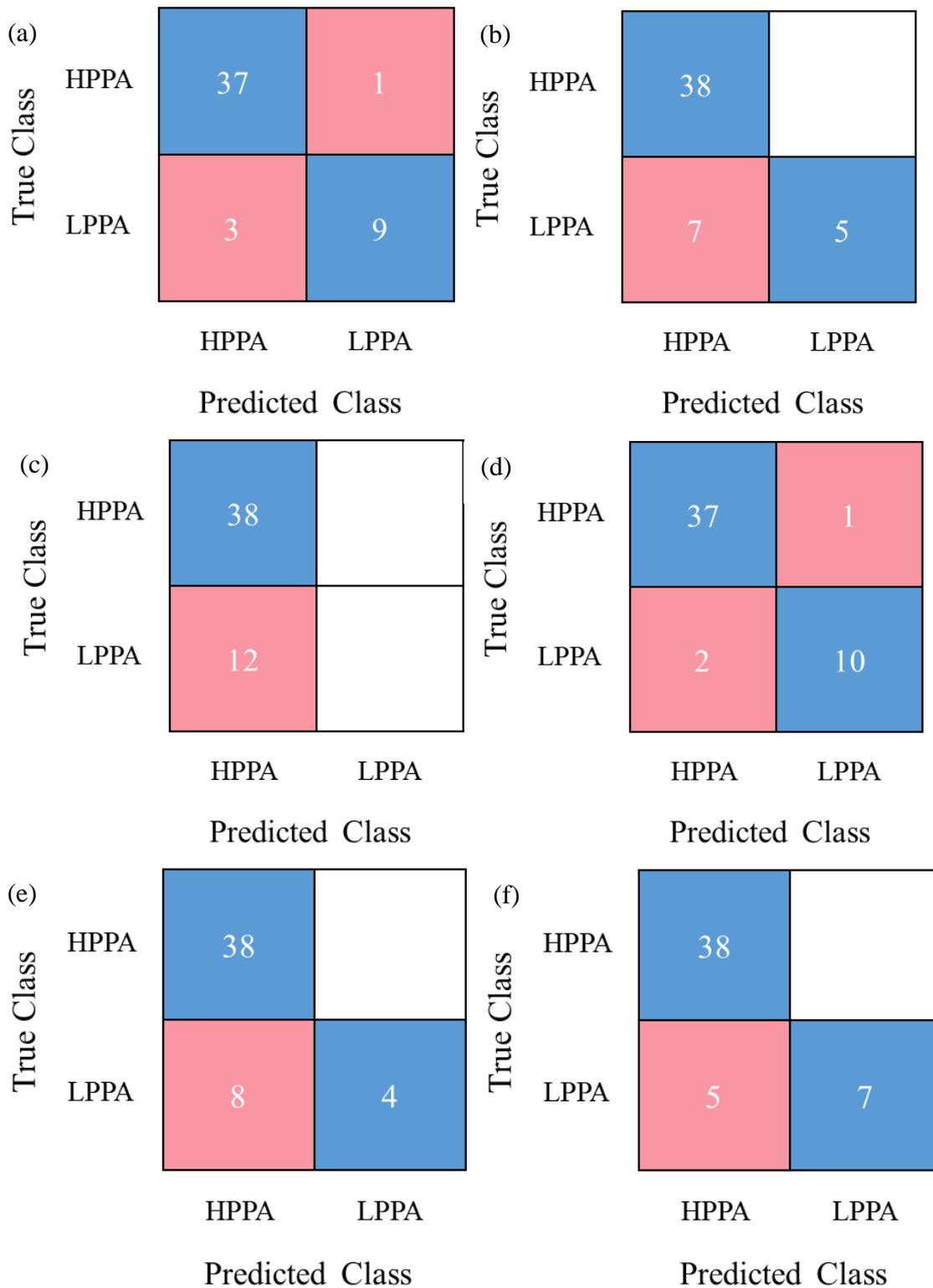


Figure 2. Confusion matrix. (a) Fine; (b) Medium; (c) Coarse; (d) Cosine; (e) Cubic; (f) Weighted

4. Discussion

The present study has demonstrated that the selected psychological coping skills variables established can predict or classify well the performance of the archers either as LPPA or HPPA. The shooting scores allow for the clustering of the archers with respect to measured psychological coping skills. This was carried out by means of *k*-means as illustrated in the box plot (Figure 1). It was shown that the cosine-based model is able to classify the LPPA and HPPA well.

The result accentuated the need for mental aptitudes, for example, dependence and self-motivational strategies in the sport. Archery is closed skill-based sport, in which the archer is required to compete against another archer. For an archer to get a higher score, he or she should have the ability to maintain confidence as well as self-motivation. Evidence in sports and exercise science demonstrates that one's attention to capability or self-assurance is the significant factor prompting the achievement of any type of sport [11].

Additionally, it has been depicted that confidence and self-motivation are the amongst the qualities of a winner. They are the secret elements that all superior athletes seem to achieve, regardless of what level they contest since positive mental attitude keeps an athlete working hard irrespective of how often he/she may fail or how many hindrances might come in his/her direction [12]. Confidence and self-motivation could give a team or an ordinary athlete the determination and focus on defeating a stronger opponent. These mental abilities could lead athletes to the greater level of performance in their domain. Likewise, inadequate or lack of confidence and self-motivation could render athletes perform below their natural potential. Poor confidence and self-motivation could diminish an athlete's indulgence of the sport and turn him/her into a dropout and a frequent loser [1].

5. Conclusion

The present study has successfully shown that the selected psychological coping skills variables, i.e. coping with adversity, coachability, peaking under pressure, concentration, freedom from worry, confidence and achievement motivation, goal setting and mental preparation does to a certain degree influence the performance of the archers. The study also has demonstrated that the employment of such machine learning methods allow coaches to appropriately identify high potential athletes in the sport of archery by taking into consideration the aforementioned elements. The current findings are non-trivial to coaches and sports managers in identifying talents with significantly little effort and cost.

Acknowledgement

This work is funded by the National Sports Institute of Malaysia (ISNRG: 8/2014-12/2014).

References

- [1] Abdullah M R, Musa R M, Maliki A B H M B, Kosni N A and Suppiah P K 2016 Role of psychological factors on the performance of elite soccer players *J. Phys. Educ. Sport* **16** 170
- [2] Lazarus R S 2000 How emotions influence performance in competitive sports *Sport Psychol.* **14** 229–52
- [3] Smith R E and Christensen D S 1995 Psychological skills as predictors of performance and survival in professional baseball *J. Sport Exerc. Psychol.* **17** 399–415
- [4] Pasaei N, Abdoli B, Kazem S, Mousavi V and Aslankhani M 2013 Does Predict Personality Introversion-Extraversion Dimension the Archer's Performance? *Middle-East J. Sci. Res.* **13** 760–3
- [5] Fuster-Parra P, García-Mas A, Ponseti F J and Leo F M 2015 Team performance and collective efficacy in the dynamic psychology of competitive team: A Bayesian network analysis *Hum. Mov. Sci.* **40** 98–118
- [6] Fuster-Parra P, García-Mas A, Ponseti F J, Palou P and Cruz J 2014 A Bayesian network to discover relationships between negative features in sport: a case study of teen players *Qual. Quant.* **48** 1473–91

- [7] Smith R E, Schutz R W, Smoll F L and Ptacek J T 1995 Development and validation of a multidimensional measure of sport-specific psychological skills: The Athletic Coping Skills Inventory-28 *J. Sport Exerc. Psychol.* **17** 379–98
- [8] Collins B M, Lydersen J M, Fry D L, Wilkin K, Moody T and Stephens S L 2016 Variability in vegetation and surface fuels across mixed-conifer-dominated landscapes with over 40 years of natural fire *For. Ecol. Manage.* **381** 74–83
- [9] Taha Z, Musa R M, P.P. Abdul Majeed A, Alim M M and Abdullah M R 2018 The identification of high potential archers based on fitness and motor ability variables: A Support Vector Machine approach *Hum. Mov. Sci.* **57** 184–93