Human Movement Science 57 (2018) 184–193

Contents lists available at ScienceDirect

Human Movement Science

journal homepage: www.elsevier.com/locate/humov

Full Length Article

The identification of high potential archers based on fitness and motor ability variables: A Support Vector Machine approach

Zahari Taha[a], Rabiu Muazu Musa[a,b,*], Anwar P.P. Abdul Majeed[a], Muhammad Muaz Alim[a], Mohamad Razali Abdullah[b]

[a] *Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia*

[b] *Faculty of Applied Social Sciences, Universiti Sultan Zainal Abidin, 21300 Kuala Terengganu, Terengganu, Malaysia*

ARTICLE INFO

Keywords:
Archery
Support Vector Machine
Fitness variables
Motor ability

ABSTRACT

Support Vector Machine (SVM) has been shown to be an effective learning algorithm for classification and prediction. However, the application of SVM for prediction and classification in specific sport has rarely been used to quantify/discriminate low and high-performance athletes. The present study classified and predicted high and low-potential archers from a set of fitness and motor ability variables trained on different SVMs kernel algorithms.

50 youth archers with the mean age and standard deviation of 17.0 ± 0.6 years drawn from various archery programmes completed a six arrows shooting score test. Standard fitness and ability measurements namely hand grip, vertical jump, standing broad jump, static balance, upper muscle strength and the core muscle strength were also recorded. Hierarchical agglomerative cluster analysis (HACA) was used to cluster the archers based on the performance variables tested. SVM models with linear, quadratic, cubic, fine RBF, medium RBF, as well as the coarse RBF kernel functions, were trained based on the measured performance variables.

The HACA clustered the archers into high-potential archers (HPA) and low-potential archers (LPA), respectively. The linear, quadratic, cubic, as well as the medium RBF kernel functions models, demonstrated reasonably excellent classification accuracy of 97.5% and 2.5% error rate for the prediction of the HPA and the LPA.

The findings of this investigation can be valuable to coaches and sports managers to recognise high potential athletes from a combination of the selected few measured fitness and motor ability performance variables examined which would consequently save cost, time and effort during talent identification programme.

1. Introduction

Archery is a fine and gross motor skill sport, wherein a success is defined by the capacity to shoot a target repeatedly with tremendous precision and accuracy. It has been stated that to achieve victory in the sport, an optimum level of physical fitness and motor ability are required (Musa, Abdullah, Maliki, Kosni, & Haque, 2016). Previous findings documented that the mastery of specific physical fitness and motor skill variables such as core body strength, upper body strength, hand grip, leg power and static balance could have a desirable outcome to the performance of the particular sport (Ertan, Kentel, Tümer, & Korkusuz, 2003; Martin, Spenst,

* Corresponding author at: Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia.

E-mail address: rabiumuazu@imamslab.com (R.M. Musa).

https://doi.org/10.1016/j.humov.2017.12.008
Received 21 April 2017; Received in revised form 4 December 2017; Accepted 10 December 2017
Available online 14 December 2017
0167-9457/

Drinkwater, & Clarys, 1990). Moreover, Soylu, Ertan, and Korkusuz (2006) emphasise that the nature of archery involve interval aerobic and anaerobic activities and therefore, all the major muscle groups are stimulated during the shooting process. The aforementioned physical fitness variables are shown to be essential for performance in the sport of archery (Spencer et al., 2004).

The employment of machine learning or artificial intelligence has gained popularity in predicting and classifying physiological properties owing to its superiority over conventional means. Artificial Neural Networks (ANN) was used to predict energy expenditure from body-worn accelerometers attached to the right hip, right thigh as well as both wrists (Montoye, Begum, Henning, & Pfeiffer, 2017). The accuracy of the ANN prediction was compared to linear regression as well as linear mixed models. It was established from the study that the ANN model could provide a better energy expenditure prediction accuracy compared to conventional methods.

Random forest classifier also used to predict energy expenditure and also to classify physical activity (e.g., household, stairs, walking and running) by using heart rate (HR) data and accelerometer data (Ellis et al., 2014). The household activities consist of laundry, window washing, dusting, dishes as well as sweeping. Furthermore, the authors employed random forest regression trees to estimate metabolic equivalents (MET). The study demonstrated that both random forest classification, as well as regression forest, was able to reasonably classify the PA type as well as predicting MET, respectively.

Trost, Zheng, and Wong (2014) investigated the capability of a regularised logistic regression model to classify seven distinct activity classes namely lying down, sitting, standing, walking, running, basketball as well as dancing through acceleration signals obtained from the wrist and hip. A threefold cross-validation technique was employed to evaluate the predictability of the machine learning model developed. It was established that the hip model could classify the aforementioned physical activities with a classification accuracy of 91.3%, 95.8%, 95.8%, 96.8%, 88.3%, 81.9%, 64.1% for sitting, standing, walking, running, lying down, basketball and dance, respectively. Conversely, the wrist model exhibits classification accuracy of 93.0%, 91.7%, 95.8%, 86.0%, 78.8%, 74.6%, 69.4% sitting, standing, walking, basketball, running, lying down and dance, respectively.

It is evident from the literature that machine learning has potential applications in the field of human movement. Support vector machine (SVM) have been rather popular among researchers in a variety of fields for classification and regression analyses (Azamathulla & Wu, 2011; De Yong, Bhowmik, & Magnago, 2015; Guyon, Weston, Barnhill, & Vapnik, 2002; Khedher, Ramírez, Górriz, Brahim, & Segovia, 2015; Liu, Wang, Wang, & Li, 2013; Morra et al., 2009; Zheng, Yoon, & Lam, 2014). SVM is a supervised machine learning algorithm proposed by Cortes and Vapnik that is generally used for addressing both classification and regression problems (Cortes & Vapnik, 1995).

Nonetheless, the application of SVM to ascertain high and low potential archers has yet been investigated. This study aimed to determine whether an archer's ability could be classified based on their performance on a selection of physical and motor ability assessments namely hand grip, vertical jump, standing broad jump, static balance, upper muscle strength and core muscle strength. The hierarchical agglomerative cluster analysis (HACA) was used to cluster the archers based on their performances in the selected variables and their archery shooting scores. The archers were clustered by the HACA into two distinct classes, i.e. high potential archers (HPA) and low potential archers (LPA). Furthermore, the efficacy of different support vector machines (SVMs) in classifying HPA as well as LPA is also investigated.

2. Material and methods

2.1. Participants

50 archers were selected to take part in this study. The participants consisted of 37 male and 13 female youth archers between the ages of 13 and 20 (17.0 ± 0.6 years) drawn from different archery programmes in Malaysia. The archers were under a development program, who will be consequently targeted to represent the state and national level archery competitions. Written consent was obtained from all the archers who participated in the study. 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. Experimental approach to the problem

To address the objectives and the problems initially raised in the study, fitness assessments of the performance variables were conducted in accordance with the standard protocol for the fitness evaluation. The procedures for the fitness assessments are explained in detail in the subsequent sections.

2.3. Sit up test

The test was performed corresponding to the suggested technique for physical fitness assessments (Noguchi, Demura, & Takahashi, 2013). The archers laid on their back with their knees bent at approximately right angles (90°) while both feet held flat on the floor. They placed their hands on their chest where they should remain throughout the test. In the assessment process, a supporter held the archers' feet on the ground. The archers sat up until they touched their knees to both elbows; thus, they came back to the floor. The test conducted within 1 min. Once the test ended, and the assistant totalled and recorded the number of the proper sit-ups test completed. The test was carried out one time due to the impact of exhaustion. The test was implemented to evaluate the archers' core muscle strength (CMS).

2.3.1. Push up test

The archers adopted a prostrating posture on the floor with the palms precisely underneath the arms, legs stretched and toes folded under so they are in touch with the floor, (push up position). The archers then press with the arms until they are stretched out and then lower their body until the chin or chest meets the floor. At this moment the head to the toes was straight. All of these actions were carried out by the arms and shoulders. The amount was determined by the number of push-ups while maintaining proper form completed within 1 min. The test was conducted once to avert fatigue. The test is primarily carried out to measure the upper muscle strength (UMS) of the archers.

2.3.2. Standing stork test

Standing stork test was implemented to assess the ability of the archers to maintain stability (S) while standing on one leg. The general aim of the test is to evaluate the capability of the archers to balance on the ball of their foot. To materialise this test, the archers were instructed to remove their shoes and place their hands on their hips, then position the non-supporting foot against the inner part of the supporting leg. The archers raise their heel to balance on the ball of the foot. A stopwatch was used to time the period taken to complete the test in the correct form. The tests stopped when the archers move to any direction, and the non-supporting foot misses contact with the knee, or the heel of the supporting foot touches the floor. The archers alternate both legs and average of the total scores from both legs were utilised for statistical analysis.

2.3.3. Hand grip test

Hand grip (HG) was used to measure the maximum isometric strength of the hand and forearm muscles of the archers. In order to implement this test, hand grip dynamometer was used, and the archers were requested to grasp the dynamometer in hand to be examined, with the arm at right angles and the elbow by the side of the body. The knob of the dynamometer is adjusted according to the size of the individual archer. The base of the hand grip was placed on first metacarpal (heel of palm) while the handle rested in the middle of four fingers. When the archers were ready, they were permitted to squeeze the dynamometer with maximum isometric effort, which is maintained for about 5 s. No other body movement was permitted. The archers were encouraged by the researchers to give a maximum effort. Both hands were alternated.

2.3.4. Standing broad jump test

Standing broad jump (SBJ) was administered to test the explosive leg power of the archers. Long jump landing mat was placed on the flat synthetic surface, and the take-off line was clearly indicated from behind the mat. The archers stand behind the line marked on the ground with feet slightly apart. A two-foot take-off and landing are utilised, with the swinging of the arms and bending of the knees to provide forward drive. The archers were instructed to jump as far as possible and land on both feet without falling backwards. Three attempts were permitted, and best was taken for statistical analysis.

2.3.5. Vertical jump test

A Vertec testing gadget (M-F Athletic Co., Cranston, Rhode Island) was utilised to decide vertical jump (VJ) height (cm), an appropriate and robust measure of lower body power (Baechle & Earle, 2008). In order to realise this test, a prepared tester attuned the height of the colour-coded plastic vanes such that it paralleled to the archer's standing reach height. The vane stack was then raised to a standardised distance (corresponding to the participants' anticipated jump), so the archers would not jump higher or lower than the arrangement of vanes. By executing a countermovement, the archers flexed the ankles, knees, and hips and swung the arms in an upward motion whilst tapping the highest plausible vane with the fingers of their dominant hand. Every archer performed three jumps with 40–60 s rest between every jump. The best of three trials was documented and utilised for statistical analysis.

2.3.6. Archery shooting score test

A simulated shooting competition area was set up, and all the archers' shoot six arrows (one end) over a distance of 50 m. A total of 10 arrow shots were permitted for the archers, the first four were used as a warm-up to prepare the archers to deliver their best performance while the last six were documented for the statistical analysis (Spratford & Campbell, 2017).

2.4. Data analysis

2.4.1. Pre-processing

Normality assessment was carried out by applying the Shapiro-Wilk test, and the archers were observed to be homogeneously distributed. The normality test was performed to ensure that the samples possess relatively similar characteristics. Moreover, a test for identifying outliers was carried out using Grubbs' test. The test examines the hypothesis that the value that is the furthest from the sample mean is an outlier (Grubbs, 1950). The results from the analysis revealed that outliers were identified and removed accordingly.

2.4.2. Clustering: hierarchical agglomerative cluster analysis (HACA)

In this study, hierarchical agglomerative cluster analysis (HACA) was employed to identify the classes of the archers on relative performance components measured using XLSTAT 2014 add-in Software. The Cluster Analysis is an explorative analysis that tries to identify structures within the data. Cluster analysis is also known as distribution analysis or taxonomy analysis. More precisely, it attempts to identify homogenous groups of cases, i.e., observations, participants or respondents (Abdullah et al., 2016). Cluster analysis is applied to classify groups of events if the grouping is not earlier established. The decision is illuminated by a dendrogram, showing the groups and their proximity (Forina, Armanino, & Raggio, 2002). Nevertheless, the Euclidean distance (similarity distance) is presented as Dlink/Dmax, which stands for the quotient between the correlations distances separated by the largest distance (Shrestha & Kazama, 2007).

2.4.3. Classification: support vector machines (SVMs)

SVM is utilised to acquire the optimal hyperplane that correctly classifies or divides the data into two distinct classes which in this study, these classes are represented by HPA and LPA. Nonetheless, it is worth to note that SVM may also be extended to cater multiclass problems (Brereton & Lloyd, 2010; Xu, Zomer, & Brereton, 2006). The acquisition of the aforementioned optimal hyperplane is through the identification of the maximal distant from the classes that in turn minimises the risk of misclassification of both the training and validation data set.

The generalised SVM classification function is given by

$$f(x) = \sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b \tag{1}$$

Whereby $K(x_i,x)$ is the kernel function that is used to measure the training vector (x_i,x) and it may be modified to cater linearly non-separable data. Different kernel functions are used in this study to investigate its efficacy in classifying the data correctly. The kernel functions employed are

$$K(x_i,x) = (x_i^T.\, x), \text{ Linear kernel} \tag{2}$$

$$K(x_i,x) = \left(\frac{X_i^T \cdot X + 1}{C}\right)^d, \text{ Polynomial kernel of degree d} \tag{3}$$

$$K(x_i,x) = \exp\left\{-\frac{(X-X_i)^2}{\sigma^2}\right\}, \text{ Radial Basis Function(RBF)kernel} \tag{4}$$

The following objective function is minimised in order to obtain the optimal classification function is as follows:

$$\min \frac{1}{2} w^T w + C\left(\sum_{i=1}^{l} \varepsilon_i\right) \tag{5}$$

where C is the penalty parameter whilst ε is the slack variable. For a more in-depth mathematical treatment of the subject matter, the readers are referred to (Cortes & Vapnik, 1995). In this study, six different kernel functions are investigated namely, linear, quadratic, cubic, fine RBF, medium RBF, and Coarse RBF. It is worth to mention that the quadratic and the cubic kernels fall under the polynomial kernel function. The scale of the fine, medium and coarse RBF is defined by $\frac{\sqrt{P}}{4}$, $\sqrt{P}$, and $4\sqrt{P}$, respectively where P is the number of predictors, i.e. six. The solver used for the training is the sequential minimal optimisation algorithm.

2.5. Model training and testing

The data was normalised between 0 to 1 to allow equal distribution and representation of the selected variables. It is noteworthy to mention that 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 (40 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. Then, the average performance over all the folds is then computed. Furthermore, the remaining data from the overall observation, i.e. 10 was subsequently used to evaluate the predictability of the classifiers to ascertain HPA and LPA. The SVM analysis and assessment were conducted by means of MATLAB 2016a (Mathworks Inc., Natick, USA).

2.6. 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 the Matthew's correlation coefficient (MCC). The ACC is essentially the ratio between the number of correctly classified observations and the total number of observations. The SENS and the SPEC are the true positive (TP) rate or the positive class accuracy as well as the true negative (TN) rate or negative class accuracy, respectively. The PREC computes the number of correct positive predictions over the total number of positive predictions. The ERR, on the other hand, evaluates all misclassifications over the number of total observations.

Conversely, the MCC is a discrete version of the Pearson's correlation coefficient that measures the quality of binary classification, and it has a range of -1 to 1 whereby 1 suggests a completely correct binary classifier and -1 suggests otherwise. It gauges the performance of the classification models. The confusion matrix allows the observation of correctly classified and misclassified observations that transpires between the defined classes. The confusion matrix of a two-class classification problem is shown in Table 1.

Where TP, TN, FP and FN are true positives (the number of positive samples correctly predicted), true negatives (the number of

Table 1
Confusion Matrix.

	Predicted class	
Actual Class	TP	FN
	FP	TN

negative samples predicted correctly), false positives (negative samples predicted as positive) and false negatives (number of positive samples predicted as negative). The formulae of the classifier performance assessment parameters are as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$SENS = \frac{TP}{TP + FN} \tag{7}$$

$$SPEC = \frac{TN}{TN + FP} \tag{8}$$

$$PREC = \frac{TP}{TP + FP} \tag{9}$$

$$ERR = \frac{FP + FN}{TP + TN + FP + FN} \tag{10}$$

$$MCC = \frac{(TP*FP) + (FN*TN)}{(TP + FN) + (TP + TN) + (FP + FN) + (FP + TN)} \tag{11}$$

3. Results

3.1. Clustering

The assignation of the low and high potential archers was determined based on the shooting score as well as the six performance variables (i.e. hand grip, vertical jump, standing broad jump, static balance, upper muscle strength and the core muscle strength). The agglomerative hierarchal cluster analysis was utilised to define the group membership as either HPA or LPA. As demonstrated by the dendrogram in Fig. 1 and the box plot illustrated in Fig. 2, the HPA outperformed the LPA across all the performance variables including the shooting score. This method enables the correct assignation of the defined memberships, as not only the shooting scored is considered, but other performance fitness variables were also taken into account. These variables that are shown in the box plots are therefore considered as essential attributes that distinguish HPA from LPA.

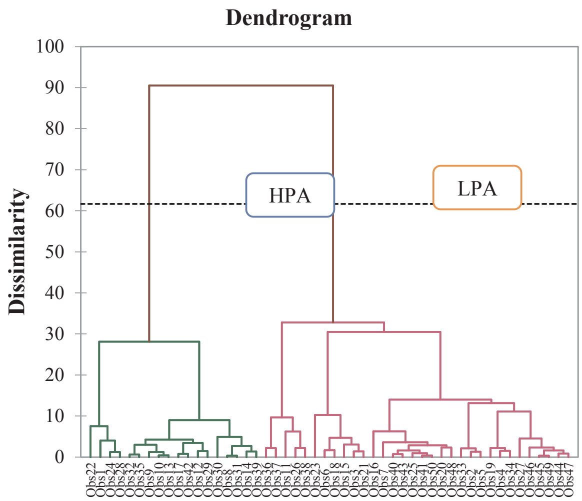

Fig. 1. Dendrogram of the two classes assigned by the Cluster Analysis.

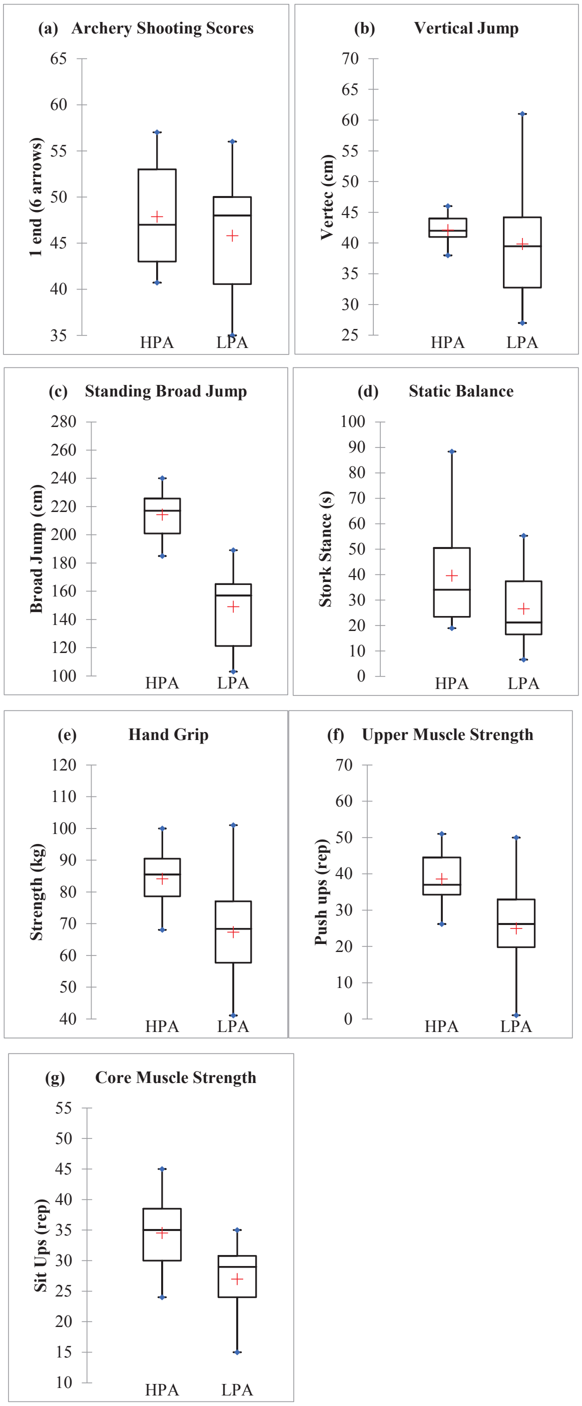

Fig. 2. Comparisons of performance differences of the archers based on seven variables evaluated (a) Archery shooting score; (b) Vertical Jump; (c) Standing Broad Jump; (d) Static Balance; (e) Hand Grip; (f) Upper Muscle Strength; (g) Core Muscle Strength.

3.2. Classification

It could be observed from the tabulated results in Table 2 that the linear, quadratic, cubic, as well as the medium RBF kernel functions based SVM models, are able to produce exceptionally high classification through the evaluation of all the assessed performance variables. The coarse RBF could also provide a reasonably accurate classification with an error rate of 12.5%. However, it is apparent that the fine RBF is unsuitable for predicting the correct classification of HPA and LPA. Through the present study, it is

Table 2
Model evaluation.

Kernel Functions	ACC (%)	SENS (%)	SPEC (%)	PREC (%)	ERR (%)	MCC
Linear	97.5	100	96.3	92.86	2.5	0.9456
Quadratic	97.5	100	96.3	92.86	2.5	0.9456
Cubic	97.5	100	96.3	92.86	2.5	0.9456
Fine RBF	67.5	0	100	N/A	32.5	N/A
Medium RBF	97.5	100	96.3	92.86	2.5	0.9456
Coarse RBF	87.5	61.54	100	100	12.5	0.72058

evident that SVM in particular, linear, quadratic, cubic and medium RBF based kernel functions may be used for the purpose of talent identification for archers based on the predefined physical fitness and motor abilities. The confusion matrices of the evaluated SVMs are depicted in Fig. 3.

3.3. Sensitivity analysis

A sensitivity test was performed to evaluate the contribution of the individual variables to the performance of the classification models. In this stage, one variable is omitted from the total number of variables and the model is trained to assess its classification accuracy as tabulated in Table 3. It is evident that a better classification accuracy is attained upon removing the HG variable from the model. However, it is worth noting that, with the exception of HG, other variables examined in the present study are shown to have a relatively equal contribution towards the performance of the classifiers.

It could be seen from Table 4 that the linear, quadratic and cubic classifier models are able to predict accurately all the fresh data supplied to the system. However, one misclassification is observed for the medium RBF classifier. Nonetheless, it is still acceptable as the classification accuracy of the model is 97.5 %. It is apparent that the misclassifications of that the fine and coarse RBF models against fresh data are rather pronounced. This is expected owing to the classification accuracy demonstrated earlier.

Table 5 lists the classification ability of the classifier models with the exclusion of the hand grip variable. It could be seen that a better classification is observed for the medium RBF classifier unlike its performance with all variables as indicated via its classification accuracy of 100% in Table 3. It could also be seen that there is an improvement in the classification ability of the fine RBF SVM classifier, as the misclassification has been reduced to four instead of five.

4. Discussion

The finding of the present study has indicated that the selected performance variables established are able to predict or classify well the performance of the archers, i.e. HPA and LPA through the shooting score of the respective archers. The shooting scores enable us to cluster the athletes in respect to measured fitness and motor ability tested namely hand grip, vertical jump, standing broad jump, static balance, upper muscle strength, and the core muscle strength via HACA as illustrated in Fig. 1. Furthermore, the box plot depicted in Fig. 2 signifies the justification that the mean performances of HPA are greater than of LPA across all the aforementioned performance variables evaluated.

The performance of the SVM models are then investigated in terms of its efficacy in classifying the categories of the archers correctly, and it was demonstrated that the linear, quadratic, cubic and medium RBF based kernel functions SVMs are excellent models in classifying the LPA and HPA. Conversely, it was also established that the fine RBF based SVM model provided the least classification accuracy (67.5%). Furthermore, it is also observed that the aforementioned model misclassified all the HPA from the fresh data supplied to the model. This observation is rather obvious as the SENS which indicates the positive class accuracy is shown to be zero as demonstrated in Table 2. In addition, upon carrying out the sensitivity test, it was demonstrated that the hand grip does not significantly influence the predictive models, in fact, it reduces the classification accuracy. Moreover, the remaining performance variables contributed relatively equally to the classification accuracy of the evaluated models. In other words, the remaining performance variables are vital in the sport of archery as based on solely these parameters, a reasonably accurate classification was established. The importance of the selected variables has been reported in the literature as a prerequisite for a better archery performance evaluation (Keast & Elliott, 1990; Landers et al., 1991).

Landers et al. (1991) maintained that archery is a sport that demands stamina and muscular strength to cope with the sport's physical demand since its competition involved continuous shooting and movement forth and back. Therefore, muscular strength and endurance would offer substantial advantages to the archers to perform at their peak. Keast and Elliott (1990) on the other hand indicated that archery requires an upper body and core muscle strength which actuates the main muscle groups. The ability of the main muscle groups to respond to the demand placed upon them in the process of the sport assists the archers to shoot the arrow to the target effectively. It has also been opined that power, balance and hand grip capabilities support the archers to endure the duration of the sport (Calvert, Banister, Savage, & Bach, 1976; Elferink-Gemser, Visscher, Lemmink, & Mulder, 2007).

Musa et al. (2016) inferred that leg power and balancing ability or otherwise called stability is one of the elements that influence stages of the precision of an arrow in aiming its target. The superior the stability of the archer will cause a narrower angle of the parabolic trajectory of the arrow (Suppiah et al., 2017). This will lessen any interference from the influence of the external elements which can alter the direction of the arrow (Spratford & Campbell, 2017). The impact is that, when pulling and holding the bow, the

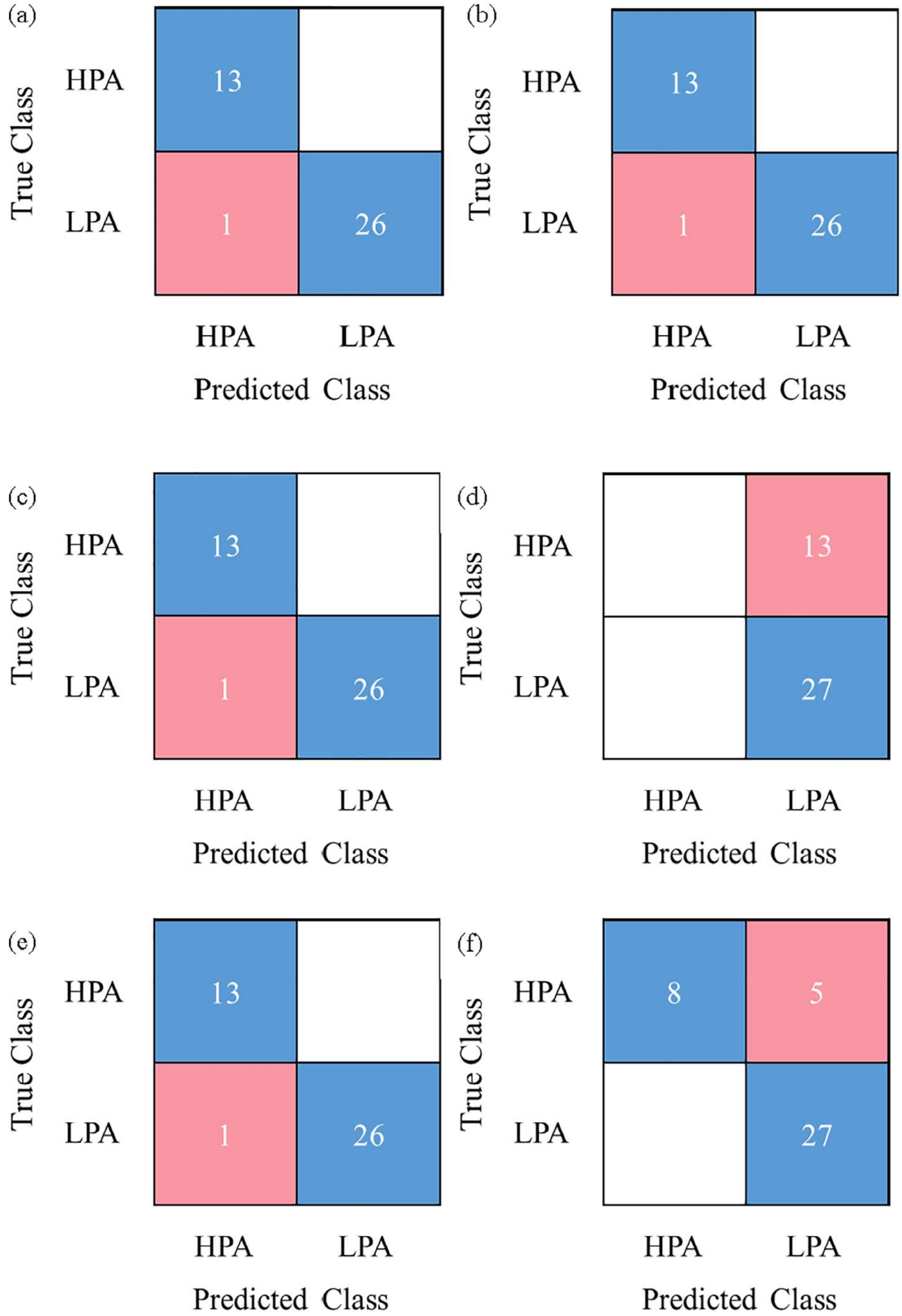

Fig. 3. Confusion matrix. (a) Linear SVM; (b) Quadratic SVM; (c) Cubic SVM; (d) Fine RBF SVM; (e) Medium RBF SVM; (f) Coarse RBF SVM.

athlete requires a good stability. Therefore, based on this explanation it can be understood that there is a correlation between muscular strength, leg power and balance to the achievement of high performance in archery. The ability of the archer to harness the aforementioned performance variables play a significant role in deciding their performance outcome. Moreover, Bompa and Haff (2009) suggest that in power-driven perspective, the accuracy in archery subsequent release of the arrow from the bow is often influenced by environment, such as the wind blow. The higher the wind blows, the larger fluctuation of arrow's direction. Hence, to attain a desirable pulling of the bow string, there is a need for a greater arm strength, as well as the corresponding ability to maintain stability.

Table 3
Variable's contribution towards classification accuracy.

Excluded variables	VJ	SBJ	S	HG	UMS	CMS
Kernel Functions	Classification Accuracy (%)					
Linear	97.5	92.5	97.5	100	92.5	92.5
Quadratic	97.5	92.5	97.5	100	97.5	97.5
Cubic	97.5	92.5	97.5	100	97.5	95
Fine RBF	67.5	72.5	67.5	67.5	67.5	67.5
Medium RBF	100	92.5	100	100	92.5	87.5
Coarse RBF	87.5	67.5	87.5	85	87.5	80

Table 4
Classification ability of the trained classifiers with all variables against fresh data.

Predicted Classification						
Actual Classification	Linear	Quadratic	Cubic	Fine RBF	Medium RBF	Coarse RBF
LPA	LPA	LPA	LPA	LPA	LPA	LPA
LPA	LPA	LPA	LPA	LPA	**HPA**	LPA
LPA	LPA	LPA	LPA	LPA	LPA	LPA
LPA	LPA	LPA	LPA	LPA	LPA	LPA
LPA	LPA	LPA	LPA	LPA	LPA	LPA
HPA	HPA	HPA	HPA	**LPA**	HPA	HPA
HPA	HPA	HPA	HPA	**LPA**	HPA	**LPA**
HPA	HPA	HPA	HPA	**LPA**	HPA	**LPA**
HPA	HPA	HPA	HPA	**LPA**	HPA	HPA
HPA	HPA	HPA	HPA	**LPA**	HPA	HPA

The bold values indicate misclassification of the predictive models from the actual classification.

Table 5
Classification ability of the trained classifiers by excluding HG against fresh data.

Predicted Classification						
Actual Classification	Linear	Quadratic	Cubic	Fine RBF	Medium RBF	Coarse RBF
LPA	LPA	LPA	LPA	LPA	LPA	LPA
LPA	LPA	LPA	LPA	LPA	LPA	LPA
LPA	LPA	LPA	LPA	LPA	LPA	LPA
LPA	LPA	LPA	LPA	LPA	LPA	LPA
LPA	LPA	LPA	LPA	LPA	LPA	LPA
HPA	HPA	HPA	HPA	**LPA**	HPA	HPA
HPA	HPA	HPA	HPA	**LPA**	HPA	**LPA**
HPA	HPA	HPA	HPA	HPA	HPA	**LPA**
HPA	HPA	HPA	HPA	**LPA**	HPA	HPA
HPA	HPA	HPA	HPA	**LPA**	HPA	HPA

The bold values indicate misclassification of the predictive models from the actual classification.

5. Conclusion

The current study has successfully demonstrated that vital fitness and motor ability variables, i.e. hand grip, vertical jump, standing broad jump, static balance, upper muscle strength and the core muscle strength does influence the determination the performance class of the archers. The study also has revealed that the application of machine learning algorithms, in particular, the variation of SVMs is able to accurately predict the class of the archers based on the chosen performance variables. The following classifiers, i.e. linear, quadratic, cubic and medium RBF SVM has shown exceptional classification accuracy and precision throughout the exercise. The use of such machine learning methods is non-trivial as it allows coaches to correctly identify high potential athletes in the sport of archery by considering few essential physiological measurements which would eventually save cost and effort for talent identification programme.

The present study could serve as a pioneer for quantifying and discriminating high and low-performance athletes through a set of pre-defined relevant performance variables in the sport of archery utilising a set of SVM variation algorithms. However, the limited population size, as well as the demographical nature of the study, could not permit the generalisation of the findings to other levels of archery participation. The authors recommend that the proposed methods could be extended to some specific games and sports for the identification of talent and classifying performance abilities in athletes.

Acknowledgement

The researchers wish to thank the National Sports Institute of Malaysia for providing the Grant for the current study (ISNRG: 8/2014-12/2014).

References

Abdullah, M. R., Maliki, A. B. H. M., Musa, R. M., Kosni, N. A., Juahir, H., & Haque, M. (2016). Multi-hierarchical pattern recognition of athlete's relative performance as a criterion for predicting potential athletes. *Journal of Young Pharmacists, 8*(4), 463–470. http://dx.doi.org/10.5530/jyp.2016.4.24.

Azamathulla, H. M., & Wu, F.-C. (2011). Support vector machine approach for longitudinal dispersion coefficients in natural streams. *Applied Soft Computing Journal, 11*(2), http://dx.doi.org/10.1016/j.asoc.2010.11.026.

Baechle, T. R., & Earle, R. W. (2008). National strength & conditioning association (US). *Essentials of strength training and conditioning* (pp. 395–396). Champaign, IL: Human Kinetics.

Bompa, T. O., & Haff, G. G. (2009). *Periodization: Theory and methodology of training.* Human Kinetics Publishers.

Brereton, R. G., & Lloyd, G. R. (2010). Support vector machines for classification and regression. *Analyst, 135*(2), http://dx.doi.org/10.1039/b918972f.

Calvert, T. W., Banister, E. W., Savage, M. V., & Bach, T. (1976). A systems model of the effects of training on physical performance. *IEEE Transactions on Systems, Man, and Cybernetics, 2*, 94–102.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

De Yong, D., Bhowmik, S., & Magnago, F. (2015). An effective power quality classifier using wavelet transform and support vector machines. *Expert Systems with Applications, 42*(15–16), 6075–6081. http://dx.doi.org/10.1016/j.eswa.2015.04.002.

Elferink-Gemser, M. T., Visscher, C., Lemmink, K. A. P. M., & Mulder, T. (2007). Multidimensional performance characteristics and standard of performance in talented youth field hockey players: a longitudinal study. *Journal of Sports Sciences, 25*(4), 481–489.

Ellis, K., Kerr, J., Godbole, S., Lanckriet, G., Wing, D., & Marshall, S. (2014). A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiological Measurement, 35*(11), 2191–2203. http://dx.doi.org/10.1088/0967-3334/35/11/2191.

Ertan, H., Kentel, B., Tümer, S. T., & Korkusuz, F. (2003). Activation patterns in forearm muscles during archery shooting. *Human Movement Science, 22*(1), 37–45. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12623179.

Forina, M., Armanino, C., & Raggio, V. (2002). Clustering with dendrograms on interpretation variables. *Analytica Chimica Acta, 454*(1), 13–19. http://dx.doi.org/10.1016/S0003-2670(01)01517-3.

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. *The Annals of Mathematical Statistics,* 27–58.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46*(1), 389–422. http://dx.doi.org/10.1023/A:1012487302797.

Keast, D., & Elliott, B. (1990). Fine body movements and the cardiac cycle in archery. *Journal of Sports Sciences, 8*(3), 203–213.

Khedher, L., Ramírez, J., Górriz, J. M., Brahim, A., & Segovia, F. (2015). Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing, 151*, 139–150. http://dx.doi.org/10.1016/j.neucom.2014.09.072.

Landers, D. M., Petruzzello, S. J., Salazar, W., Crews, D. J., Kubitz, K. A., Gannon, T. L., et al. (1991). The influence of electrocortical biofeedback on performance in pre-elite archers. *Medicine & Science in Sports & Exercise*.

Liu, M., Wang, M., Wang, J., & Li, D. (2013). Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical, 177*, 970–980. http://dx.doi.org/10.1016/j.snb.2012.11.071.

Martin, A. D., Spenst, L. F., Drinkwater, D. T., & Clarys, J. P. (1990). Anthropometric estimation of muscle mass in men. *Medicine and Science in Sports and Exercise, 22*(5), 729–733.

Montoye, A. H. K., Begum, M., Henning, Z., & Pfeiffer, K. A. (2017). Comparison of linear and non-linear models for predicting energy expenditure from raw accelerometer data. *Physiological Measurement, 38*(2), 343–357. http://dx.doi.org/10.1088/1361-6579/38/2/343.

Morra, J. H., Tu, Z., Apostolova, L. G., Green, A. E., Avedissian, C., Madsen, S. K., & Ellipsis Thompson, P. M. (2009). Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. *NeuroImage, 45*(1), S3–S15. http://dx.doi.org/10.1016/j.neuroimage.2008.10.043.

Musa, R. M., Abdullah, M. R., Maliki, A. B. H. M., Kosni, N. A., & Haque, M. (2016). The application of principal components analysis to recognize essential physical fitness components among youth development archers of Terengganu, Malaysia. *Indian Journal of Science and Technology, 9*(44), 1–6.

Noguchi, T., Demura, S., & Takahashi, K. (2013). Relationships between sit-ups and abdominal flexion strength tests and the thickness of each abdominal muscle. *Advances in Physical Education, 3*(2), 84.

Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software, 22*(4), 464–475. http://dx.doi.org/10.1016/J.ENVSOFT.2006.02.001.

Soylu, A. R., Ertan, H., & Korkusuz, F. (2006). Archery performance level and repeatability of event-related EMG. *Human Movement Science, 25*(6), 767–774.

Spencer, M., Lawrence, S., Rechichi, C., Bishop, D., Dawson, B., & Goodman, C. (2004). Time–motion analysis of elite field hockey, with special reference to repeated-sprint activity. *Journal of Sports Sciences, 22*(9), 843–850.

Spratford, W., & Campbell, R. (2017). Postural stability, clicker reaction time and bow draw force predict performance in elite recurve archery. *European Journal of Sport Science, 17*(5), 539–545. http://dx.doi.org/10.1080/17461391.2017.1285963.

Suppiah, P. K., Wei, K. K., Musa, R. M., Abdullah, M. R., Low, F. L., & Maliki, A. B. H. M. (2017). The effectiveness of a core muscles stability program in reducing the postural sway of adolescent archers: A panacea for a better archery performance. *International Journal of Physiotherapy* http://doi.org/10.15621/ijphy/2017/v4i5/159425.

Trost, S. G., Zheng, Y., & Wong, W.-K. (2014). Machine learning for activity recognition: Hip versus wrist data. *Physiological Measurement, 35*(11), 2183–2189. http://dx.doi.org/10.1088/0967-3334/35/11/2183.

Xu, Y., Zomer, S., & Brereton, R. G. (2006). Support vector machines: A recent method for classification in chemometrics. *Critical Reviews in Analytical Chemistry, 36*(3–4), http://dx.doi.org/10.1080/10408340600969486.

Zheng, B., Yoon, S. W., & Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications, 41*(4), 1476–1482. http://dx.doi.org/10.1016/j.eswa.2013.08.044.