

# 배깅과 부스팅 알고리즘을 이용한 핸드볼 결과 예측 비교

김지웅<sup>1</sup>, 박종철<sup>2\*</sup>, 김태규<sup>2</sup>, 이희화<sup>3</sup>, 안지환<sup>4</sup>

<sup>1</sup>상명대학교 스포츠산업학과 연구원, <sup>2</sup>부경대학교 해양스포츠학과 교수,  
<sup>3</sup>상명대학교 스포츠산업학과 교수, <sup>4</sup>성균관대학교 스포츠과학과 박사수료

## Comparison of Handball Result Predictions Using Bagging and Boosting Algorithms

Ji-eung Kim<sup>1</sup>, Jong-chul Park<sup>2\*</sup>, Tae-gyu Kim<sup>2</sup>, Hee-hwa Lee<sup>3</sup>, Jee-Hwan Ahn<sup>4</sup>

<sup>1</sup>Researcher, Department of Sport management, Sangmyung University

<sup>2</sup>Professor, Department of Marine Sports, Pukyong National University

<sup>3</sup>Professor, Department of Sport management, Sangmyung University

<sup>4</sup>Ph.D. Candidate, College of Sport Science, Sungkyunkwan University

**요약** 본 연구는 여자핸드볼 경기에서 발생하는 움직임 정보를 바탕으로 앙상블 기법의 배깅과 부스팅 알고리즘의 예측력을 비교하고, 움직임 정보의 활용가능성을 분석하는데 목적이 있다. 연구의 목적을 달성하기 위하여 15번의 연습 경기에서 관성센서를 활용해 수집한 움직임 정보를 활용한 경기 결과예측을 랜덤포레스트와 Adaboost 알고리즘을 활용해 비교·분석하였다. 연구결과 첫째, 랜덤포레스트 알고리즘의 예측률은  $66.9 \pm 0.1\%$ 로 나타났으며, Adaboost 알고리즘의 예측률은  $65.6 \pm 1.6\%$ 로 나타났다. 둘째, 랜덤포레스트는 승리 결과는 모두 예측하였고, 패배의 결과는 하나도 예측하지 못하였다. 반면, Adaboost 알고리즘은 승리 예측 91.4%, 패배예측 10.4%라고 나타났다. 셋째, 알고리즘의 적합성 여부에서 랜덤포레스트는 과적합의 오류가 없었지만, Adaboost는 과적합의 오류가 나타났다. 본 연구결과를 바탕으로 스포츠경기를 예측할 때 움직임 정보도 활용 가능성을 확인하였으며, 랜덤포레스트 알고리즘이 보다 우수함을 확인하였다.

**주제어** : 머신러닝, 결과예측, 배깅, 부스팅, 움직임

**Abstract** The purpose of this study is to compare the predictive power of the Bagging and Boosting algorithm of ensemble method based on the motion information that occurs in woman handball matches and to analyze the availability of motion information. To this end, this study analyzed the predictive power of the result of 15 practice matches based on inertial motion by analyzing the predictive power of Random Forest and Adaboost algorithms. The results of the study are as follows. First, the prediction rate of the Random Forest algorithm was  $66.9 \pm 0.1\%$ , and the prediction rate of the Adaboost algorithm was  $65.6 \pm 1.6\%$ . Second, Random Forest predicted all of the winning results, but none of the losing results. On the other hand, the Adaboost algorithm shows 91.4% prediction of winning and 10.4% prediction of losing. Third, in the verification of the suitability of the algorithm, the Random Forest had no overfitting error, but Adaboost showed an overfitting error. Based on the results of this study, the availability of motion information is high when predicting sports events, and it was confirmed that the Random Forest algorithm was superior to the Adaboost algorithm.

**Key Words** : Machine learning, outcome prediction, bagging, boosting, movement

\*Corresponding Author : Jong-chul Park(jcpark@pknu.ac.kr)

Received May 10, 2021

Accepted August 20, 2021

Revised June 1, 2021

Published August 28, 2021

## 1. 서론

4차 산업 시대로 변화하면서 현대사회는 빠르게 변화하고 있으며, 변화하는 환경에 따라 빠르게 대응할 수 있는 능력이 승패를 결정짓고 있다[1]. 다양한 스포츠 상황에서 승패에 결정적인 영향을 미치는 요인들에 대한 미래예측은 필수적인 요소로 정확한 예측을 위한 다양한 방법과 방대한 양의 데이터가 요구되고 있다[2]. 경기력과 관련된 양적, 질적 데이터의 양은 폭발적으로 생산되었지만, 현장에서 실시간으로 적용되는데 한계가 있다[3]. 그럼에도 불구하고 스포츠과학자들은 훈련, 연습경기, 시합상황에서 발생하는 기술정보와 선수들의 움직임 정보를 정량화하여 분석하고 있으며, 이를 바탕으로 한 경기 결과 예측연구를 시도하고 있다[4].

스포츠에서 예측은 주로 경기결과, 부상예측, 경기력을 평가하거나 다음 경기를 대비하기 위해 연구되고 있으며, 궁극적인 목적은 경기에서 이기는 것이다[5]. 특히, 팀 스포츠는 상호 간 유기적인 움직임이 경기력에 영향을 미치기 때문에 선수들의 움직임 정보는 중요한 정보 중 하나로[6], 움직임 정보는 선수들에게 피드백을 제공하고, 경기력을 향상 시킬 수 있으며, 부상 위험을 줄이고, 훈련 프로그램을 설계하는데 활용되고 있다[7]. 하지만, 개인 종목에 비하여 전술적 이해도 및 다양한 요인들이 복합적으로 영향을 미치기 때문에 정확한 분석과 경기력 예측에 어려움이 있다[8].

일반적으로 스포츠 상황에서 예측을 위하여 사용된 기법 중 데이터마이닝 기법이 주로 활용되고 있으며, 데이터마이닝이란 대규모의 데이터에서 자동으로 통계적 규칙과 패턴을 분석하여 정보를 발견하는 과정을 말한다. 특히, 데이터마이닝 기법은 분류 문제를 해결하기 위해서는 주어진 데이터를 활용하여 목표변수를 잘 예측할 수 있는 분류자를 형성하는 것이 분류기법의 목적이다[9]. 분류기법을 활용하여 경기력을 예측한 연구를 살펴보면 신경망 분석을 활용하여 축구 경기에서 승리와 패배에 관련된 경기기록 요인을 추출하여 축구경기 순위예측 모형연구[10], 2013년 시즌 KBO 프로야구팀과 선수들의 경기데이터로부터 다음 경기의 승패를 예측하기 위해 의사결정나무, 랜덤포레스트, 로지스틱회귀, 신경망 분석, SVM, 판별분석을 이용한 연구들이 있으며[11], 김유철[12]은 E-sports(League of Legends)를 사례로 게임의 기록정보를 활용하여 승부예측 모델을 설계하였다.

국외에서는 농구경기 778경기를 대상으로 데이터마이닝 기법을 활용하여 경기결과를 예측하여 09/10 시즌

의 결과와 비교한 연구[13], 영국 프리미어리그 3시즌의 경기기록을 바탕으로 Bayesian Networks를 활용하여 경기결과를 예측하는 연구[14]등에서 확인할 수 있듯이 팀 스포츠경기를 예측할 때 경기기록만이 활용되었다.

최근 스포츠에서 예측 확률을 높이기 위한 방법으로 단일 알고리즘의 예측보다 앙상블 기법이 많이 활용되고 있다. 기존 축구 경기결과 예측에서 좋은 성능을 보인 통계학적 모델과 머신러닝 모델을 결합한 새로운 앙상블 모델을 제안한 연구[15], 농구경기 결과예측을 위하여 하이브리드 앙상블을 활용한 연구[16], 4개의 월드컵 경기에 대한 예측성능을 확인하기 위하여 하이브리드 랜덤포레스트와 푸아송(poisson) 회귀 모델과 비교한 연구[17]들과 같은 연구들이 있었다.

이처럼 대표적인 앙상블 기법은 단일 알고리즘 보다 예측력이 높게 나타난다는 연구 결과가 있지만, 스포츠 상황에서 앙상블을 활용하여 예측력을 검증하고 비교한 연구들은 매우 부족하다. 또한, 기존의 스포츠 경기를 예측할 때 활용된 경기기록이 아닌 움직임 정보를 활용한 연구 역시 매우 부족하다.

따라서 본 연구는 팀 스포츠 중 공격과 수비상황에서 다양한 움직임이 발생하는 핸드볼 선수들을 대상으로 움직임 정보의 활용 가능성과 앙상블 기법(ensemble method)의 배깅(bagging)과 부스팅(boosting)의 예측력을 비교하는 것이 목적이다.

## 2. 연구방법

### 2.1 연구대상

본 연구는 15번의 국내·외 연습경기에 참가한 24명의 여자 핸드볼 선수들을 대상으로 움직임 정보를 수집하였다. 실험 전 선수들에게 연구의 목적과 실험 절차 등에 대한 설명을 하였고, 연구 참여에 동의한 24명을 연구 대상으로 선정하였다. 연구대상의 세부 정보는 Table 1과 같다.

Table 1. Subject Information

total	height (cm)	weight (kg)	career (year)	age (year)
n = 24	173.4±5.3	73.3±5.3	14.9±3.7	26.2±4.4

### 2.2 자료수집

핸드볼 경기 중 총 15회의 연습경기에서 선수들의 움

직업 정보를 획득하였으며, 자료의 획득은 IMU(Inertial Measurement Unit)센서가 내장된 Catapult OptimEye-S5 (Catapult, Melbourne, Australia) 모델을 활용하였다. Fig. 1과 같이 경기시작 30분 전 2-6번 휴추 사이 위치 주머니가 있는 조끼에 센서를 넣어 고정하였으며, 총 12개의 방향 정보를 선행연구를 기반으로 Low(1.5-2.5%), Mid(2.5-3.5%), High(3.5%이상)로 설정하였다[18].

전후반 각각 30분씩 진행되는 경기에서 선수들의 객관적인 비교를 위해 경기의 시작과 종료 시간을 기록하였으며, 선수교체를 기록해 휴식시간에 대한 데이터는 제외하고 표준화(normalization)하여 분석에 활용하였다.



Fig. 1. IMU Sensor Setting

### 2.3 자료분석

핸드볼 선수들이 연습경기에서 나타낸 움직임을 12개 시계방향에 따른 강도별 움직임을 정량화하여 Microsoft Excel(Microsoft, Redmond, WA, USA) 소프트웨어를 이용하여 포지션별, 경기결과(승패)별 데이터를 5분 단위로 분류하였으며 총 1,452개의 움직임 정보를 추출하여

사용하였다. 그중 다중공선성의 VIF(Variance inflation factor)값을 Python 3.7(Python Software Foundation, NED)로 확인하여 10 이하인 36개의 요인만을 사용하였으며, Table 2와 같다.

데이터를 정규화하기 위해 Z값(z-score)으로 변화시키고, 학습데이터와 검증데이터의 비율은 70 : 30으로 설정하였으며, 정확성을 검증하기 위하여 K-fold 교차검증을 10회 실시하여 최종 예측력을 확인하였다.

결과 예측을 위해 사용된 앙상블 학습은 여러개의 알고리즘을 조합하여 결과를 예측하는 방식으로 정확도가 높은 하나의 모델을 쓰는 것보다, 정확도가 낮은 여러개의 모델을 조합하는 방식이 더 좋은 성능을 가질 수 있다는 가정하에 만들어진 기법으로 본 연구에서는 배깅과 부스팅 방법을 사용하였으며, 배깅(bagging) 기법 중 가장 대표적인 랜덤포레스트(Random Forest)와 부스팅(boosting) 기법 중 가장 대표적인 AdaBoost를 활용하였다.

### 2.4 통계분석

hyperparameter는 성능을 최적화하거나 편향(bias)과 분산(variance) 사이의 균형을 맞출 때 알고리즘을 조절하기 위해 사용하며, 학습 전에 미리 사용자에게 의하여 조절한다. 랜덤포레스트와 Adaboost의 예측률을 확인하기 위하여 Table 3과 같이 설정하였으며, gridsearch 패키지 사용하여 자동적으로 최적의 hyper parameter를 튜닝하였다. 그 결과 Table 4와 같이 랜덤포레스트는 depth = 3, leaf\_node = 8, feature = 8이 가장 최적의 값으로 나타났으며, Adaboost는 n\_estimators = 1로 나타났다.

Table 2. Factors classified as multicollinearity

Factor	VIF		
	Low	Med	High
IMA 1 O'Clock	1.89228	1.513499	1.466911
IMA 2 O'Clock	2.55998	1.650441	1.522137
IMA 3 O'Clock	4.149533	1.750158	1.433881
IMA 4 O'Clock	5.700789	1.858046	1.273673
IMA 5 O'Clock	3.709199	1.686021	1.265861
IMA 6 O'Clock	2.166208	1.500203	1.226768
IMA 7 O'Clock	1.951593	1.455561	1.277478
IMA 8 O'Clock	2.535826	1.635877	1.341183
IMA 9 O'Clock	4.933025	1.993429	1.337265
IMA 10 O'Clock	5.516561	2.197122	1.386172
IMA 11 O'Clock	3.969266	1.824022	1.478476
IMA 12 O'Clock	2.421854	1.534634	1.421215

Table 3. hyper parameter tuning

algorithm	hyper parameter
Random Forest	depth : [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] leaf_node : [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] feature : [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
Adaboost	n_estimators : [1, 5, 10, 20, 50, 100]

Table 4. hyperparameter value with best prediction

algorithm	hyper parameter
Random Forest	depth : 3 leaf_node : 8 feature : 8
Adaboost	n_estimators : 1

Table 5는 지도 학습으로 학습된 분류 알고리즘의 성능을 시각화할 수 있는 표로 본 연구에서 재현율(recall)은 실제 승리인 데이터에 대하여 승리로 정확하게 예측한 비율을 보는 지표로 TP를 TP와 FN으로 나눈 값으로 계산되었다.

Table 5. Confusion matrix

	Positive	Negative
Positive	TP	FP
Negative	FN	TN

정확도(accuracy)는 실제로 정확하게 예측한 비율을 설명하는 값인데, TP와 TN의 합을 전체 데이터 합으로 나누어 계산되었다.

F1-score는 예측모델의 성능을 평가할 수 있는 지표로 정밀도와 재현율을 종합적으로 보여주며, 정밀도와 재현율의 조화평균 값으로 계산되었다.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

정밀도(precision)는 모델에서 A로 분류한 데이터 중 실제로 A가 얼마나 있는지 보는 비율이며, 재현율은 실제 A인 데이터에 대하여 모델에서 A로 정확하게 예측한 비율이다.

정확도(accuracy)는 전체 데이터 중에서 실제로 A or B를 정확하게 예측한 비율이며, F1-score는 정밀도와 재현율을 종합적으로 보여주는 지표로 정밀도와 재현율의 조화평균 값으로 계산되었다.

### 3. 연구결과

#### 3.1 랜덤포레스트 예측

Table 6. Random Forest Prediction Results

	precision	recall	f1-score	support
loss	0.00	0.00	0.00	144
win	0.67	1.00	0.80	292
accuracy			0.67	436
macro avg	0.33	0.50	0.40	436
weight avg	0.45	0.67	0.54	436

Table 6은 랜덤포레스트 알고리즘을 활용하여 예측한 결과표이다. 전체 1452개의 데이터 중에서 테스트를 위하여 30%인 436개의 데이터를 활용하였다. 이 중 결과가 win인 데이터 292개, 결과가 loss인 데이터 144개가 활용되었는데, 먼저 결과가 loss인 예측결과를 보면 정밀도(precision), 재현율(recall), f1-score 모두 0.00으로 나타나 0%로 나타났으며, 결과가 win인 예측결과를 보면 precision은 0.67로 나타나 67%, 재현율(recall)은 1.00으로 나타나 100%, f1-score는 0.80으로 나타나 80%로 확인되었으며 알고리즘의 예측 정확성은 정확도는 0.67로 나타나 67%의 예측률로 확인되었다.

Table 7. Random Forest Confusion Matrix

	Predicted label		
	loss	win	
True Label	loss	0	144
	win	0	292

랜덤포레스트 알고리즘을 활용하여 움직임 데이터를 승리와 패배로 분류한 결과는 Table 7에 제시하였다. 먼

저 실제 패배인 데이터를 패배로 예측한 값은 0개, 승리로 예측한 값은 144개로 나타났으며, 실제 승리 데이터를 패배로 예측한 값은 0개, 승리로 예측한 값은 292개로 나타나 모두 승리로만 예측하였다.

Table 8. Random Forest K-fold Cross-validation

1	Class Distribution	[432 874]	accuracy	0.671
2	Class Distribution	[432 874]	accuracy	0.671
3	Class Distribution	[432 875]	accuracy	0.669
4	Class Distribution	[432 875]	accuracy	0.669
5	Class Distribution	[432 875]	accuracy	0.669
6	Class Distribution	[432 875]	accuracy	0.669
7	Class Distribution	[432 875]	accuracy	0.669
8	Class Distribution	[432 875]	accuracy	0.669
9	Class Distribution	[432 875]	accuracy	0.669
10	Class Distribution	[432 875]	accuracy	0.669
fold CV accuracy : 0.669 ± 0.001				

Table 8은 K-fold 교차검증으로 모든 데이터가 최소한 한번은 테스트셋으로 쓰이도록하여 10번의 테스트 결과 평균 정확도는 0.669 ± 0.001로 나타나 66.9±0.1%의 예측률로 확인되었다.

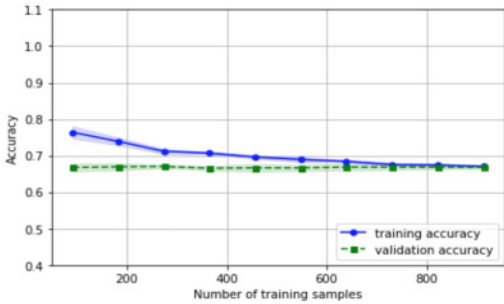


Fig. 2. Random Forest algorithm conformity

랜덤포레스트의 적합성을 판단하기 위한 그림은 Fig. 2와 같다. 학습데이터와 검증데이터의 일치율을 확인한 결과 800개 이상의 데이터에서 표준편차 범위 내에 들어와 적합성을 확인할 수 있었다.

### 3.2 Adaboost 예측

Table 9는 Adaboost 알고리즘을 활용하여 예측한 결과표이다. 전체 1453개의 데이터 중에서 테스트를 위하여 30%인 436개의 데이터를 활용하였다. 이 중 결과가 win인 데이터 292개, 결과가 loss인 데이터 144개가 활용되었는데, 먼저 결과가 loss인 예측결과를 보면 정밀

Table 9. Adaboost Prediction Results

	precision	recall	f1-score	support
loss	0.38	0.10	0.16	144
win	0.67	0.91	0.78	292
accuracy			0.65	436
macro avg	0.52	0.51	0.47	436
weight avg	0.58	0.65	0.57	436

도(precision)은 0.38로 나타나 38%, 재현율(recall)은 0.10으로 나타나 10%, f1-score은 0.16으로 나타나 16%로 나타났으며, 결과가 win인 예측결과를 보면 정밀도(precision)은 0.67로 나타나 67%, 재현율(recall)은 0.91로 나타나 91%, f1-score는 0.78로 나타나 78%로 확인되었으며 알고리즘의 예측 정확성은 정확도는 0.65로 나타나 65%의 예측률로 확인되었다.

Table 10. Adaboost Confusion Matrix

True Label	Predicted label	
	loss	win
	loss	15
win	25	267

Adaboost 알고리즘을 활용하여 움직임 데이터를 승리와 패배로 분류한 결과는 Table 10에 제시하였다. 먼저 실제 패배인 데이터를 패배로 예측한 값은 15개, 승리로 예측한 값은 129개로 나타났으며, 실제 승리 데이터를 패배로 예측한 값은 25개, 승리로 예측한 값은 267개로 나타나 승리와 패배 모두 예측되었다.

Table 11. Adaboost의 K-fold K-fold Cross-validation

1	Class Distribution	[432 874]	accuracy	0.671
2	Class Distribution	[432 874]	accuracy	0.678
3	Class Distribution	[432 875]	accuracy	0.669
4	Class Distribution	[432 875]	accuracy	0.641
5	Class Distribution	[432 875]	accuracy	0.655
6	Class Distribution	[432 875]	accuracy	0.621
7	Class Distribution	[432 875]	accuracy	0.662
8	Class Distribution	[432 875]	accuracy	0.648
9	Class Distribution	[432 875]	accuracy	0.662
10	Class Distribution	[432 875]	accuracy	0.655
fold CV accuracy : 0.656 ± 0.016				

Table 11은 K-fold 교차검증으로 모든 데이터가 최소한 한번은 테스트셋으로 쓰이도록 하여 10번의 검증을

통하여 평균 정확도는  $0.656 \pm 0.016$ 으로 나타나  $65.6 \pm 1.6\%$ 의 예측률로 확인하였다.

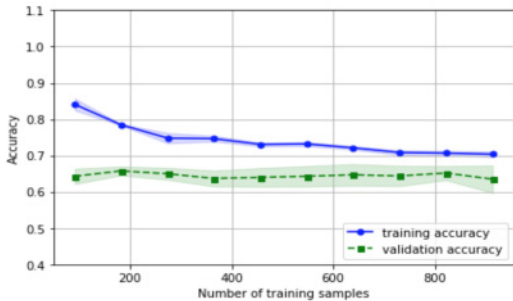


Fig. 3. Adaboost algorithm conformity

Adaboost의 적합성을 판단하기 위한 그림은 Fig. 3과 같다. 학습데이터와 검증데이터의 일치율을 확인한 결과 800개 이상의 데이터에서 표준편차 범위 내에 벗어나 있어서 과적합의 오류가 있는 것으로 확인하였습니다.

## 4. 논의

본 연구는 핸드볼 경기에서 발생하는 움직임 정보를 바탕으로 앙상블 기법의 배깅과 부스팅 알고리즘의 예측력을 비교하고, 움직임 정보의 활용가능성을 분석하는 것이 목적이다. 연구의 목적을 달성하기 위하여 15번의 연습경기에서 관성 움직임에 따른 경기결과 예측력을 랜덤포레스트와 Adaboost알고리즘의 예측력을 분석하여 비교하였다.

이러한 결과와 선행연구를 바탕으로 알고리즘의 성능과 움직임 정보의 활용가능성에 대하여 논의하고자 한다.

### 4.1 알고리즘 성능 비교

배깅(Bagging)의 대표적인 알고리즘인 랜덤포레스트의 예측력은 서포트 벡터 회귀(SVR)와 k-최근접 이웃(kNN) 보다 성능이 우수하다고 하였으며[19], 김성진·안현철[20]은 랜덤포레스트 알고리즘을 활용하여 기업신용등급 예측에 활용하여 예측의 우수함을 확인하였다. 또한, SVM, 배깅, 부스팅 알고리즘의 성능 비교에서 결측치가 존재하는 경우에는 배깅이 성능이 향상되었으며, 관측자료의 수가 많은 경우에는 부스팅의 성능이 뛰어난 것을 밝힌 연구[21]가 있었다. 국외에서는 Chan et al. [22]의 연구에서 환경평가를 위하여 랜덤포레스트 알고리즘과 Adaboost 알고리즘의 분류를 사용하였는데 랜

덤포레스트와 Adaboost의 예측 정확성이 1% 미만의 차이로 거의 동일하다고 하였다. 이러한 선행연구 결과들처럼 본 연구에서 나타난 예측률도 같은 양상으로 나타났다. 다만, 본 연구에서 랜덤포레스트는 Adaboost보다 예측 안정성이 높으며, 과적합의 오류를 피했지만, 패배의 데이터는 하나도 예측을 하지 못하였다. 그러나 Adaboost는 랜덤포레스트보다 예측률이 1.3% 낮고 과적합의 오류를 보였지만, 패배를 일부 예측하는 것으로 나타났다.

### 4.2 움직임 정보의 활용가능성

움직임 정보는 현장에서 지도자들이 스포츠경기에서 경기력을 평가하는데 활용되고 있기 때문에, 경기력을 예측하는 정보로 활용할 가치가 있다[23]. 선행연구들을 보면 인공지능 시스템을 위해서 IMU센서를 활용한 움직임 정보가 동작분석 예측을 가능하게 해줄 수 있다고 하였으며[24], IMU센서를 활용하여 무용수의 움직임을 정량화하고 부상 예측모델의 적용가능성을 이야기하였다[23]. 또한, 근전도와 가속도계의 움직임 정보를 상지움직임 구분 및 예측 가능한 모션 제어시스템을 개발하는데 활용하는 연구[25]처럼, 본 연구에서 활용된 시계방향별 움직임 정보 역시 IMU센서에서 측정된 정보이며, 핸드볼 선수들의 포지션별 움직임에 따른 경기력이 다르다는 연구[18]에서 밝혔듯이 움직임 정보는 경기력을 예측하는데 활용가능성이 크다고 사료된다.

## 5. 결론 및 제언

본 연구는 핸드볼 경기에서 발생하는 움직임 정보를 바탕으로 앙상블 기법의 예측력을 비교하고, 움직임 정보의 활용가능성을 분석하는 것이 목적으로 연구되었다. 이에 본 연구결과를 토대로 도출한 결론은 다음과 같다.

첫째, 랜덤포레스트 알고리즘의 예측력은  $66.9 \pm 0.1\%$ 로 나타났으며, Adaboost 알고리즘의 예측력은  $65.6 \pm 1.6\%$ 로 나타났다.

둘째, 랜덤포레스트는 승리 결과는 모두 예측하였고, 패배의 결과는 하나도 예측하지 못하였다. 반면, Adaboost 알고리즘은 승리 예측을 91.4%, 패배예측을 10.4% 할 수 있는 것으로 나타났다.

셋째, 알고리즘의 적합성 여부에서 랜덤포레스트는 과적합의 오류가 없었지만, Adaboost는 과적합의 오류가 나타났다.

기존의 스포츠 분석 관련 연구들은 한정된 몇 가지의 분석기법만을 사용한 반면, 본 연구는 기존에 사용되지 않았던 배깅과 부스팅 알고리즘 비교하며 적용가능성을 확인할 수 있었다. 또한, 선수들의 움직임 정보도 경기력을 예측하기 위한 중요한 정보로 활용가치가 있다고 할 수 있으며, 추후 스포츠에서 경기결과를 예측하는 연구에서 경기기록 뿐만 아니라 선수들의 움직임 정보도 함께 사용한다면 보다 정확한 예측률을 기대할 수 있을 것이다.

## REFERENCES

- [1] K. B. Goo & S. B. Kim. (2018). Issues of Sports Circle in the Era of 4th Industrial Revolution. *The Korean Society for the Philosophy of Sport, Dance & Martial Arts*, 26(2), 7-17.
- [2] W. J. Kim, Y. S. Choi & D. H. Yoo. (2018). Development of Win-Loss Prediction Models and Strategies for Improving Winning Rate of the Korean Professional Baseball Teams Using Data Mining Techniques. *Korean Journal of Sport Management*, 23(3), 88-104.  
DOI : 10.31308/KSSM.23.3.6
- [3] J. H. Cho. (2012). Utilization and Prospect of Sport Big Data. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 14(3), 1-11.  
DOI : 10.21797/ksme.2012.14.3.001
- [4] J. W. Lim & H. J. Kim. (2008). Setting Sprint Zone for performance analysis in Field-Hockey: Using Global Positioning System (GPS). *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 10(1), 69-79.  
DOI : 10.21797/ksme.2008.10.1.006
- [5] I. S. Yeo & C. H. Park. (1998). Ethics of competition and victory in sport: *The Korean Society for the Philosophy of Sport, Dance & Martial Arts*, 6(1), 67-88.  
UCI(KEPA) : I410-ECN-0101-2018-069-001973573
- [6] D. Jennings, S. Cormack, A. J. Coutts, L. Boyd & R. J. Aughey. (2010). The validity and reliability of GPS units for measuring distance in team sport specific running patterns. *International journal of sports physiology and performance*, 5(3), 328-341.  
DOI:10.1123/ijsp.5.3.328
- [7] D. W. Wundersitz, P. B. Gastin, S. Robertson, P. C. Davey & K. J. Netto. (2015). Validation of a trunk-mounted accelerometer to measure peak impacts during team sport movements. *International journal of sports medicine*, 36(09), 742-746.  
DOI: 10.1055/s-0035-1547265
- [8] S. K. Min. (2014). Sports Science: On-site support using the Global Positioning System (GPS) in field hockey competitions compared to the Asian Games. *Sports science*, 128, 48-53.
- [9] Y. H. Jung, S. H. Eo, H. S. Moon & H. J. Cho. (2010). A Study for Improving the Performance of Data Mining Using Ensemble Techniques. *Communications for Statistical Applications and Methods*, 17(4), 561-574.  
UCI : G704-000420.2010.17.4.006
- [10] W. H. Lee. (2006). *development of soccer ranking prediction model using neural network analysis*. Unpublished doctoral dissertation, Myongji University, Seoul.
- [11] Y. H. Oh, H. Kim, J. S. Yun & J. S. Lee. (2014). Using Data Mining Techniques to Predict Win-Loss in Korean Professional Baseball Games. *Journal of the Korean Institute of Industrial Engineers*, 40(1), 8-17.  
DOI : 10.7232/JKIE.2014.40.1.008
- [12] Y. C. Kim. (2018). *E-sports game prediction using machine learning : focusing on League of Legends*. Unpublished master's thesis, Chung-Ang University, Seoul.
- [13] D. Miljković, L. Gajić, A. Kovačević & Z. Konjović. (2010, September). The use of data mining for basketball matches outcomes prediction. In *IEEE 8th International Symposium on Intelligent Systems and Informatics*. (pp. 309-312). IEEE.  
DOI: 10.1109/SISY.2010.5647440
- [14] N. Razali, A. Mustapha, F. A. Yatim & R. Ab Aziz. (2017, August). Predicting football matches results using Bayesian networks for English Premier League (EPL). In *Iop conference series: Materials science and engineering*, 226(1), p.012099.  
DOI: 10.1088/1757-899X/226/1/012099
- [15] J. H. Yi & S. W. Lee. (2020). Prediction of English Premier League Game Using an Ensemble Technique. *KIPS Transactions on Software and Data Engineering*, 9(5), 161-168.
- [16] W. Cai, D. Yu, Z. Wu, X. Du & T. Zhou. (2019). A hybrid ensemble learning framework for basketball outcomes prediction. *Physical A: Statistical Mechanics and its Applications*, 528, 121461.  
DOI: 10.1016/j.physa.2019.121461
- [17] A. Groll, C. Ley, G. Schauburger & H. Van Eetvelde. (2019). A hybrid random forest to predict soccer matches in international tournaments. *Journal of Quantitative Analysis in Sports*, 15(4), 271-287.  
DOI: 10.1515/jqas-2018-0060
- [18] J. C. Park, K. S. Yoon & J. E. Kim. (2015). Movement Analysis of Women's Handball Players by Position using Inertial Measurement Units. *Journal of the Korea Convergence Society*, 11(4), 343-350.  
DOI : 10.15207/JKCS.2020.11.4.343
- [19] D. R. Jang & M. J. Park. (2020). A Study on the Art Price Prediction Model Using the Random Forests. *Journal of Applied Reliability*, 20(1), 34-42.  
DOI : 10.33162/JAR.2020.3.20.1.34

[20] S. J. Kim & H. C. Ahn. (2016). Application of Random Forests to Corporate Credit Rating Prediction. *Kyungsung University Institute for Industrial Development*, 32(1), 187-211.  
DOI : 10.22793/indinn.2016.32.1.006

[21] Y. S. Lee, H. J. Oh & M. K. Kim. (2005). An Empirical Comparison of Bagging, Boosting and Support Vector Machine Classifiers in Data Mining. *The Korean Journal of applied Statistics*, 18(2), 343-354.  
UCI : G704-000408.2005.18.2.003

[22] J. C. W. Chan & D. Paelinckx. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112(6), 2999-3011.  
DOI: 10.1016/j.rse.2008.02.011

[23] H. H. Lee & J. E. Kim (2019). Application of Electronic Performance Tracking System (EPTS) for Dance Evaluation of Dancers and Prediction Model of Injury Prevention. *The Korean Journal of Sport*, 17(4), 1597-1607.

[24] S. J. Kang. (2019). Preliminary Study on the Prediction of Human Kinematic Motion Analysis Using Artificial Intelligence. *Korean Society for Precision Engineering*, 19-19.  
DOI:10.1007/978-981-13-9129-3\_1

[25] H. Y. Jung, Y. J. Na, C. W. Wang & S. D. Min. (2014). Multi-objective system using EMG and accelerometer. *Journal of Electrical Engineering & Technology. Summer Academic Conference*, 1,502 - 1,503.  
UCI(KEPA) : I410-ECN-0101-2016-560-001853938

**김 지 응(Ji-eung Kim) [정회원]**



- 2017년 8월 : 상명대학교 스포츠정보 기술융합(체육학 석사)
- 2020년 8월 : 상명대학교 체육학과(체육학 박사)
- 관심분야 : 스포츠 ICT, 경기분석
- E-Mail : amdykje@naver.com

**박 종 철(Jong-chul Park) [정회원]**



- 2004년 8월 : 상명대학교 체육학과(교육학석사)
- 2012년 2월 : 상명대학교 체육학과(체육학박사)
- 2014년 8월 ~ 2020년 9월 : 한국스포츠정책과학원 선임연구위원
- 현, 2020년 10월 ~ 현재 : 부경대학교

해양스포츠학과 조교수

- 관심분야 : 운동역학, 경기분석
- E-Mail : jcpark@pknu.ac.kr

**김 태 규(Tae-gyu Kim) [정회원]**



- 2003년 2월 : 부경대학교 해양스포츠학과(체육학사)
- 2008년 8월 : 한국체육대학교 건강관리학과(체육석사)
- 2012년 2월 : 한국체육대학교 체육학과(이학박사)
- 현, 부경대학교 해양스포츠학과 교수
- 관심분야 : 스포츠의학, 체육측정평가
- E-Mail : ktk7718@gmail.com

**이 희 화(Hee-hwa Lee) [정회원]**



- 2007년 9월 : 상명대학교 레저스포츠학과(석사)
- 2011년 2월 : 상명대학교 체육학과(체육학 박사)
- 2014년 3월 ~ 현재 : 상명대학교 스포츠산업학과 교수
- 관심분야 : 스포츠마케팅, 디지털콘텐츠, 경기분석

- E-Mail : leehh@smu.ac.kr

**안 지 환(Jee-Hwan Ahn) [정회원]**



- 1992년 2월 : 서울대학교 언론정보학과 졸업
- 2010년 2월 : 연세대학교 언론홍보대학원 (광고홍보 석사)
- 2014년 9월 ~ : 성균관대학교 스포츠과학과 박사과정 수료
- 현, 대한핸드볼협회 운영본부장/SK

HTF 팀장

- 관심분야 : 스포츠마케팅, 스포츠매니지먼트
- E-Mail : snowboard@nate.com