

연구 배경

건물 내 재실자의 열 쾌적성이 과거에는 크게 주목받지 못했으나 공조 시스템 기술의 발달과 전반적인 경제 성장 및 소득 증가로 인해 사람들의 생활 수준이 상승함에 따라 점차 중요한 이슈로 부상하였다. 이러한 변화로 인해 건물 설계와 운영에 있어 열 쾌적성은 중요한 요소로 인식되고 있다. 특히 주요 국가들의 에너지 사용 중 상당한 부분이 건물 부문에서 사용되고 있는 것을 고려할 때, 지속 가능한 건물 운영과 에너지 효율 향상 측면에서 재실자의 열 쾌적성은 관심을 기울여야 할 영역이다. 국제에너지기구(International Energy Association)에 따르면 전 세계 에너지 사용량 중 약 36%를 건물 부문에서 차지하고 있으며 이 중 에너지 사용량이 가장 높은 다섯 국가인 미국, 인도, 러시아, 중국, 유럽을 기준으로 건물에서 사용되는 에너지 중 약 38%를 HVAC 시스템의 운전이 차지한다. HVAC 시스템은 온도와 습도, 기류, 청정도를 제어하여 재실자의 열 쾌적성을 만족시켜 생산성과 업무 효율, 건강 등에 중요한 영향을 미친다.

따라서 HVAC 시스템의 셋포인트 제어는 재실자의 열 쾌적성에 의해 결정되며, 이는 재실자의 열 쾌적성이 건물의 에너지 사용량에 큰 영향을 끼친다는 것을 의미한다. 이에 따라 건물의 에너지 효율을 개선하기 위해서는 재실자의 열 쾌적성에 대한 이해와 적절한 제어가 필수적이라 할 수 있다.

이러한 맥락에서 본 연구는 열 쾌적 관련 데이터 베이스를 활용하여 재실자의 열 쾌적성에 대한 분석을 수행함으로써 의미 있는 인사이트를 도출하고자 한다. 그 결과로서 열 쾌적성에 대한 가이드라인을 제공하여 재실자 중심의 설계와 운영에 기여하고자 한다.

ASHRAE Global Thermal Comfort Database II

분석에 사용한 데이터 베이스는 ASHRAE Global Thermal Comfort Database 2이다. 해당 데이터 베이스는 1995년부터 2015년에 이르기까지 총 20여년 동안 전 세계의 여러 연구에서 사용된 열 쾌적성 데이터를 포함하고 있다. 다양한 기상 조건과 건물 유형 별로 실제 측정되고 수집된 데이터를 제공하여 열 쾌적성을 분석하는데 도움을 주는 정보를 제공하며, 데이터 베이스 크기는 107,583 × 70 이다. 그 중 일부는 다음과 같다.

- 기후 및 지역
 - 코펜 기후 분류에 따른 총 16개 기후 포함 (Fig 1)
- 건물 유형
 - 사무실, 학교, 다세대주택, 병원, 노인센터 등
- 재실자 정보
 - 연령, 성별, 의류 단열 수준, 활동에 따른 발열량 등
- 실내 환경
 - 실/내외 온도, 상대습도, 풍속, 복사온도
- 열 쾌적 지표
 - PMV (Predicted mean vote), TS (Thermal sensation)

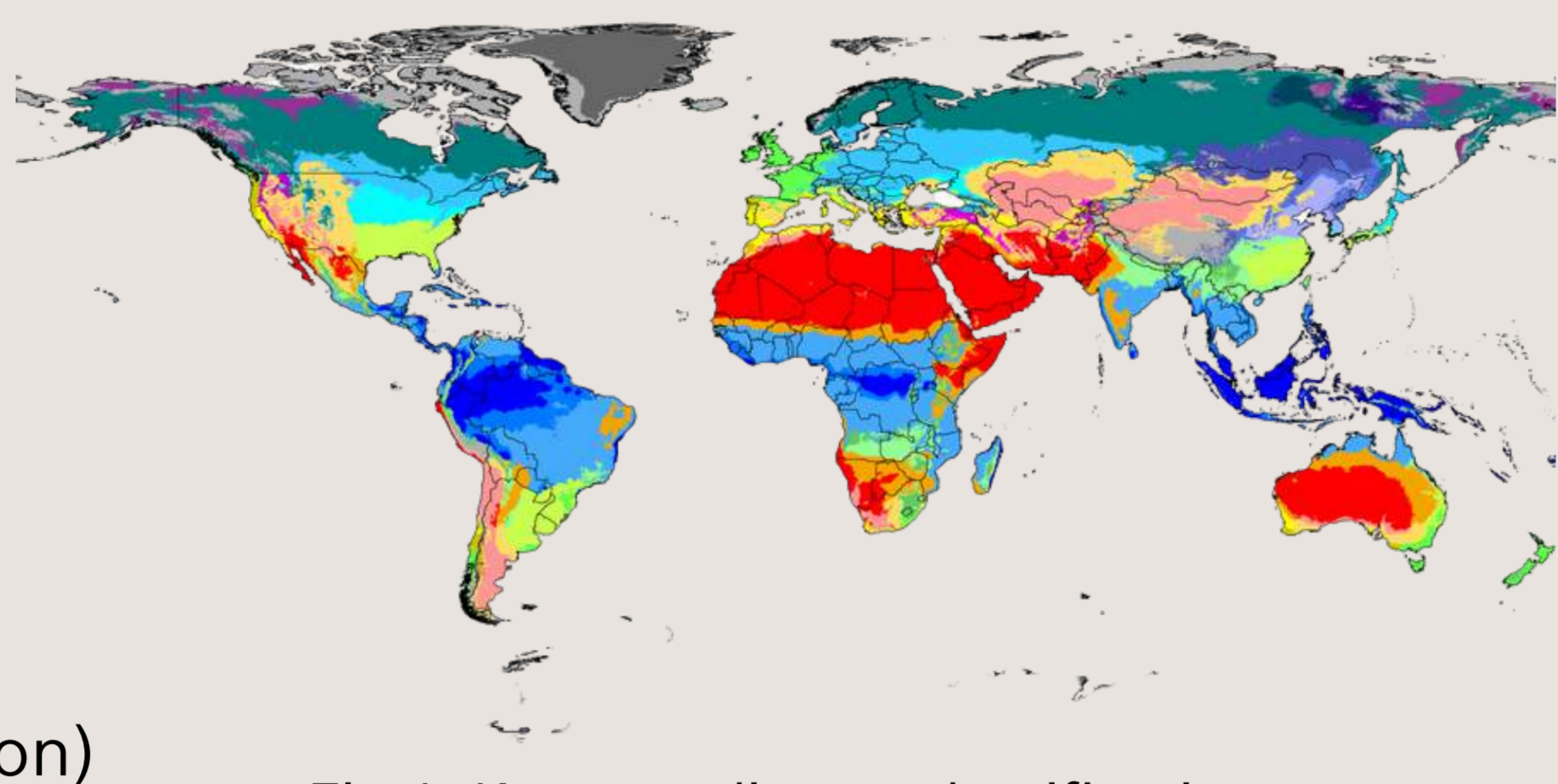


Fig 1. Köppen climate classification

데이터 분석

데이터 베이스의 변수 간 선형 연관성을 직관적으로 확인하기 위해 Pairwise plot과 Pearson 상관계수를 사용하였다. Pairwise plot이란 통계학과 데이터 시각화에서 사용되는 기법으로 다변량 데이터를 두 변수씩 짝을 지어 산점도를 그리는 방식이다. Pearson 상관계수는 두 변수의 상관 관계를 -1에서 +1 사이의 값으로 표현한 수치이며, +1은 완벽한 양의 선형 상관관계, 0은 선형 상관관계가 없음, -1은 완벽한 음의 선형 상관관계를 의미한다.

데이터를 기후별, 건물 용도별, 성별 등 여러 조건으로 분해하여 변수들 사이의 관계와 패턴을 파악하고자 하였다 (Fig 2).

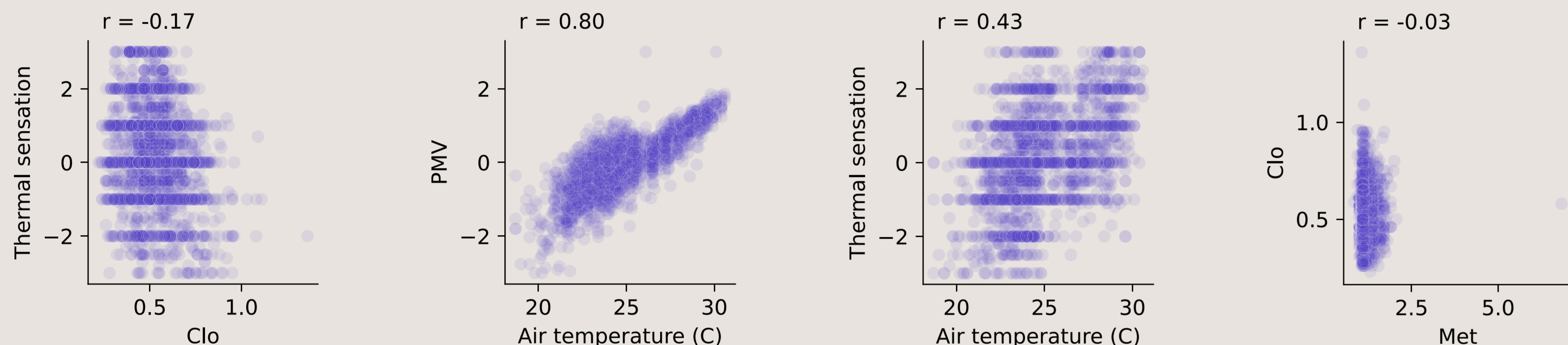


Fig 2. 온난습윤 기후 여름철 사무실 데이터 변수 상관관계 시각화

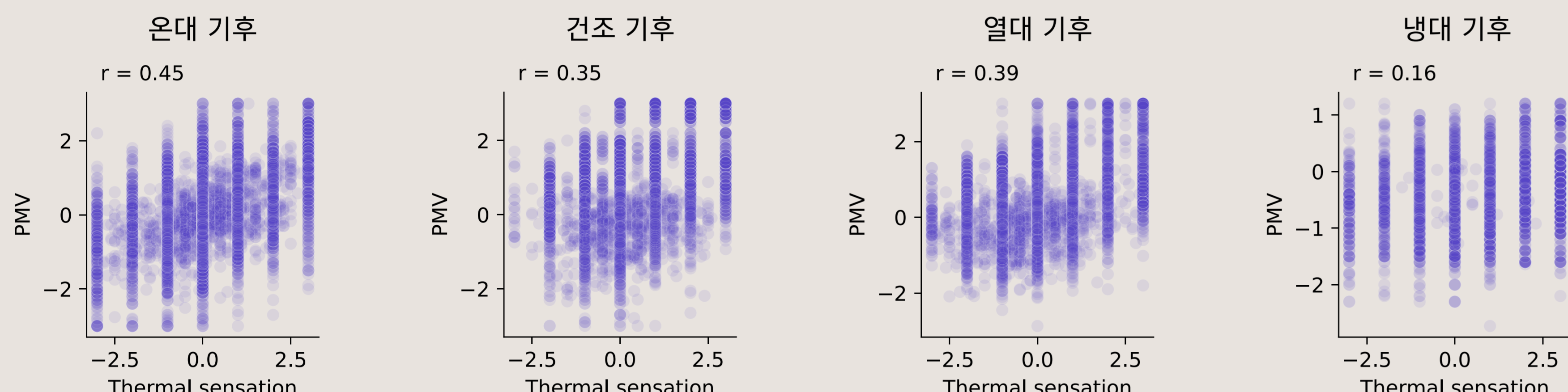


Fig 3. 각 기후대에서의 PMV와 TSV 상관관계 시각화

데이터 분석 결과 PMV가 TSV를 잘 예측하지 못한다는 점을 발견하였다 (Fig 3).

PMV (Predicted mean vote)는 인간의 온열 환경에 대한 쾌적도를 예측하는 지수이다. 이는 열 쾌적 분야의 저명한 연구자인 P. O. Fanger가 제안했으며, 열 쾌적성 평가의 표준으로 널리 사용되고 있다. PMV는 4가지 환경요인(공기온도, 평균 복사온도, 상대습도, 기류속도)과 2가지 개인요인(활동에 따른 발열량, 의류 단열 수준)을 종합적으로 고려한 값이다 (Fig 4).

TSV (Thermal sensation vote)는 실제 사람들이 체감하는 열 쾌적성을 설문조사를 통해 평가한 점수이다. TSV는 참여자들이 직접 점수를 부여하기에 연구자들로 하여금 실제 사람들이 느끼는 쾌적성에 대한 정량적 파악을 가능하게 한다.

PMV는 TSV를 예측하기 위해 개발된 지수이며, 두 지표 모두 -3~3 범위 열 쾌적 정도를 나타낸다 (Fig 5). PMV 모델을 통해 건물 설계 및 시스템 초기 단계에서 실내 환경의 최적화를 도모할 수 있다. 그러나 데이터베이스 분석을 통해 PMV 모델의 예측 성능이 낮다는 것을 확인하였고, 이것은 국제표준화기구(ISO)와 ASHRAE에서 온열 환경 표준으로 PMV 모델이 널리 사용되는 것에 대한 타당성에 의문을 제기할만 하다.

따라서 본 연구는 합리적인 열 쾌적성 지표 예측에 초점을 맞추었다.



Fig 4. 열 쾌적에 관련된 4가지 환경요인과 2가지 개인요인

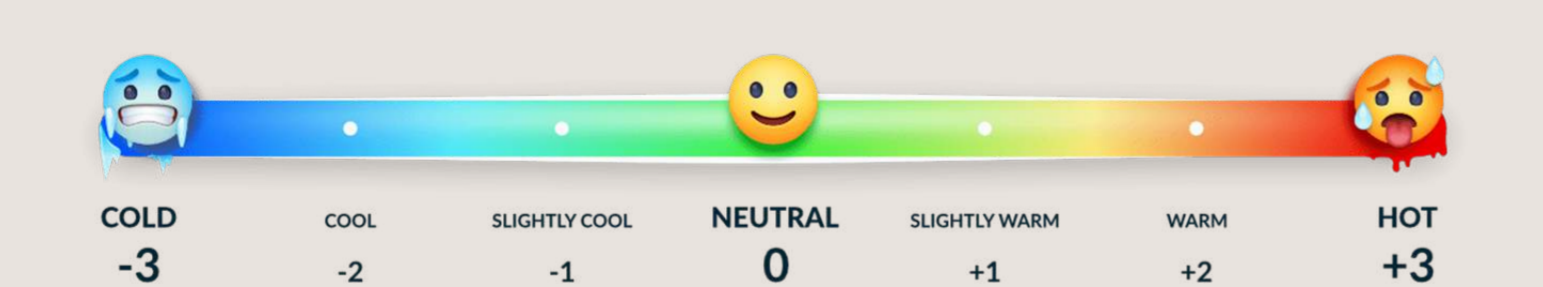
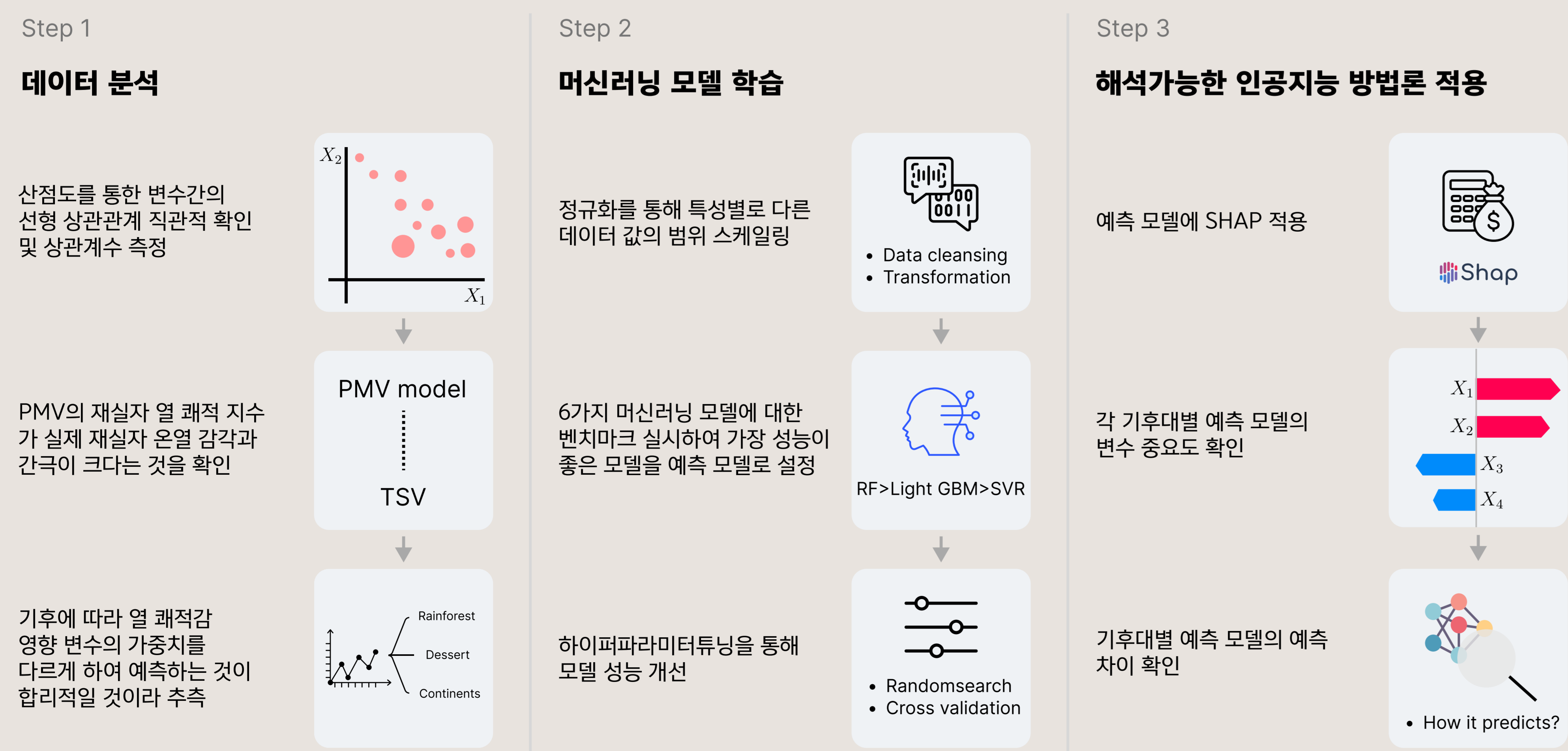


Fig 5. PMV, TSV의 인덱스

연구 설계 프로세스



데이터 전처리

데이터 분류

모델 학습을 위해 학습 데이터와 테스트 데이터를 분류하였다. 학습데이터는 PMV 모델의 변수와 유사하게 공기온도, 복사온도, 기류속도, 상대습도, 활동에 따른 발열량, 의류 단열 수준으로 정하였고 테스트 데이터는 TSV로 설정하였다. TSV와 PMV가 모두 -3 ~ 3사이 실수형으로 분포해있다는 점에서 회귀모델이 적합하다고 판단하여 회귀모델을 예측에 사용하였다. 각 기후대 별로 결과를 비교하기 위해, 필요한 데이터를 모두 포함하고 있고 데이터 수가 많은 기후대 3개를 선정하였다. 선정된 기후는 Cfa (온난 습윤 기후), Bwh (더운 사막 기후), Af (열대 우림 기후)이며, 데이터 수는 각각 5161, 4198, 812 개이다.

성능 지표

모델의 성능을 평가하기 위하여 평균 제곱 오차(MSE)와 회귀계수(R²)를 사용하였다. 평균제곱오차란 예측값과 실측값의 차이를 제곱한 값의 평균으로 낮은 값일 수록 예측 모델의 성능이 좋다고 판단할 수 있다. 회귀계수는 0~1사이의 값을 가지며 값이 1에 가까울수록 모델이 데이터를 잘 설명하고 있다는 것을 의미한다.

데이터 정규화

특성별로 데이터 값의 범위가 다르기 때문에 이상치에 강한 Robust scaler 알고리즘을 사용하였다. Robust scaler는 데이터의 중앙값과 사분위수를 사용하여 피쳐들의 값을 조절한다.

파라미터 튜닝은 벤치마크 이후에 가장 성능이 좋은 모델에 한해서 실시할 것이기에 별도의 튜닝은 수행하지 않았다.

모델 벤치마크

총 6가지의 모델에 대해서 벤치마킹을 수행했다. 사용된 모델은 Linear regression, Decision tree, Random forest, Support vector machine, XGBoost, Light GBM 등 총 6가지이다. 벤치마킹한 모델의 성능지표 결과로 평균제곱오차와 회귀계수를 확인하였다 (Fig 6).

벤치마킹 결과 전반적으로 랜덤 포레스트의 성능이 높은 것으로 나타났다. 최종적으로 서로 다른 3개의 기후대를 예측하는 머신러닝 예측 모델로 랜덤 포레스트를 사용하는 것으로 결정하였다. 하이퍼 파라미터 튜닝은 파이썬의 Scikit learn에서 제공하는 randomizedsearchCV를 사용하였다. 개선된 성능을 가진 모델을 새롭게 적용하여 서로 다른 기후대에서의 열 쾌적성을 예측하였다.

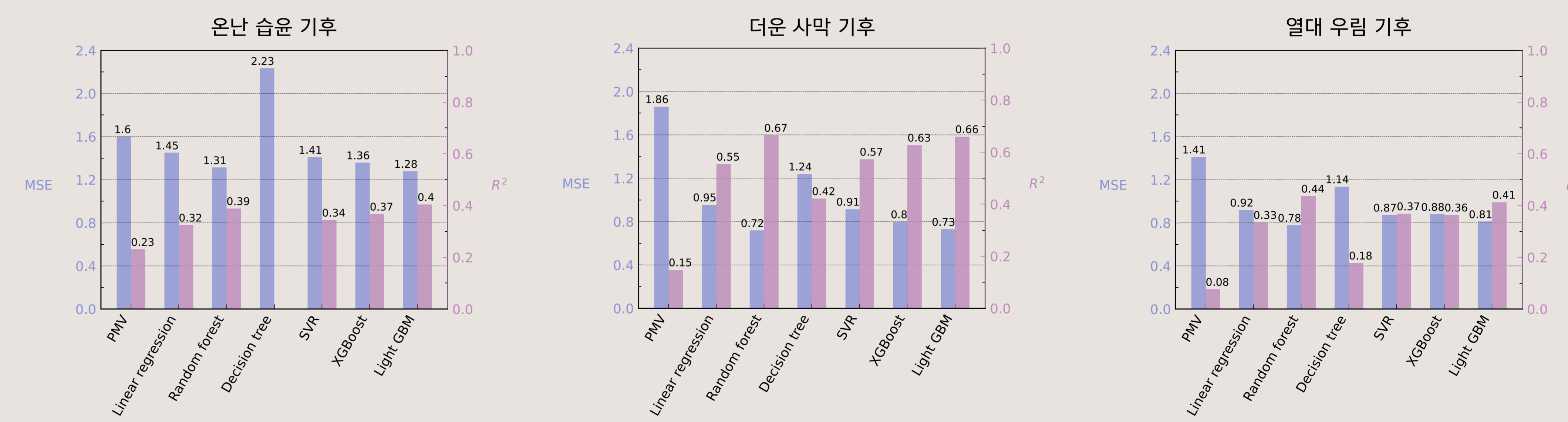


Fig 6. 기후대 별 모델 성능 비교

해석 가능한 인공지능 적용

하이퍼파라미터 튜닝까지 마친 모델로 온난습윤, 더운사막, 열대우림 기후의 TSV에 대한 각각의 예측모델을 만들었다. 제작한 예측 모델에 해석가능한 인공지능 방법론 중 SHAP를 적용하였다. SHAP는 'SHapley Additive exPlanations'의 약자로 XAI 기법 중 하나이다. SHAP를 예측 모델에 적용하면 각 피쳐가 얼마나 예측 결과에 영향을 미쳤는지를 알 수 있다.

여기서 Feature value가 빨간색이라는 것은 피쳐의 값이 크다는 것이며, 파란색은 피쳐의 값이 작음을 의미한다. 또한 SHAP value가 양수면 TSV를 높이는 쪽으로 기여하는 것이고, 음수면 TSV를 낮추는 방향으로 기여하는 것이다. Mean absolute value는 각 변수의 기여도를 절댓값을 취한 것으로, 변수가 예측에 미치는 절대적인 영향을 나타낸다. SHAP 적용 결과는 Fig 7 - 9 와 같다.

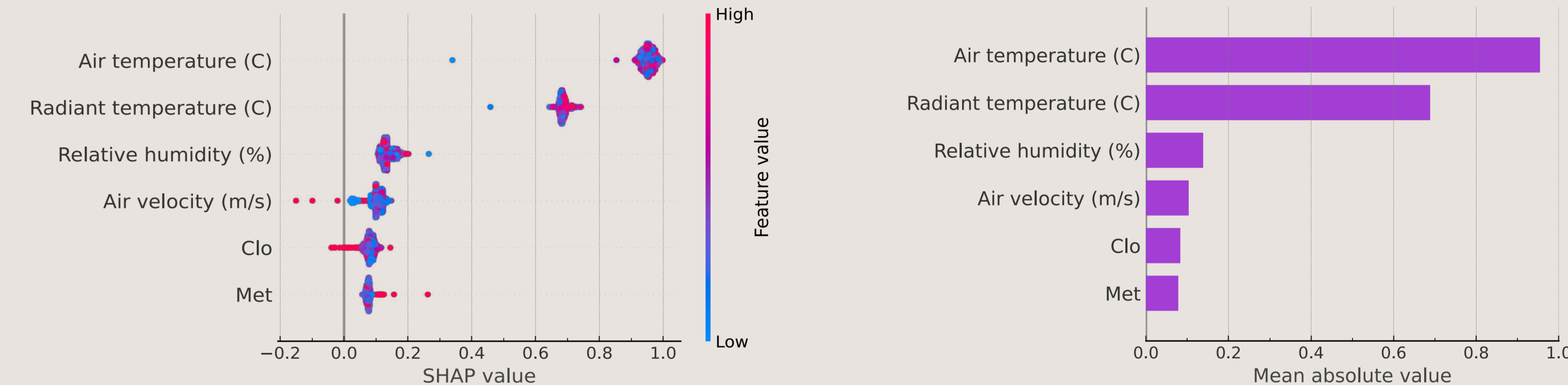


Fig 7. Cfa (온난 습윤 기후) SHAP 적용 결과

- 공기온도와 복사온도, 상대습도는 변수의 크기와 상관없이 TSV 예측에 양의 영향력을 가진다.
- 기류속도의 값이 클수록 급수적으로 TSV 예측에 음의 영향력을 가진다.

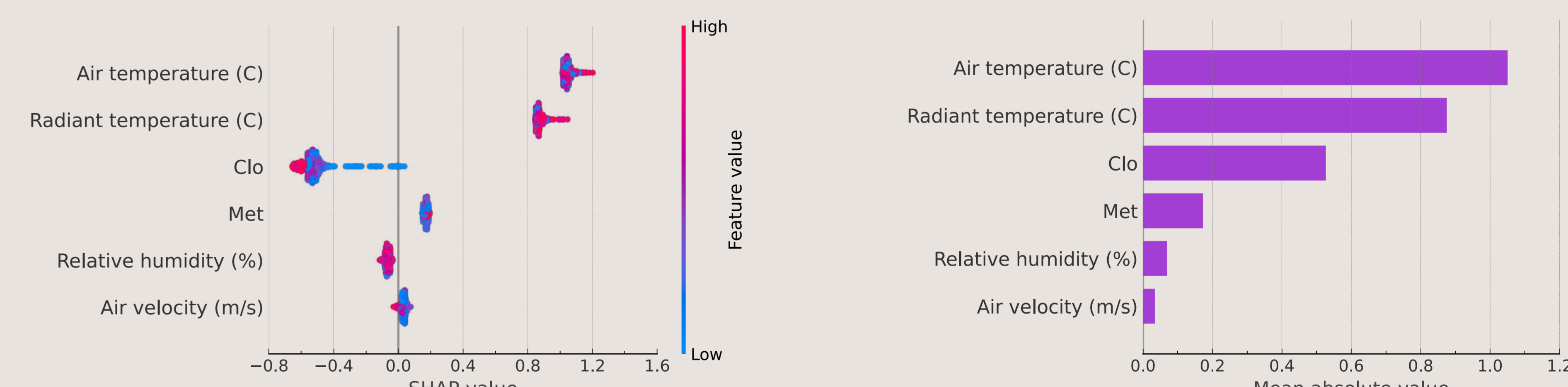


Fig 8. Bwh (더운 사막 기후) SHAP 적용 결과

- 공기온도와 복사온도 다음으로 의류 단열 수준의 영향력이 높다. 이는 낮에 매우 덥고 밤에는 상대적으로 추운 사막 기후 조건에서는 체온 유지가 더 까다롭기 때문에 올바른 의류 선택이 중요한 요소로 간주되는 것으로 보인다.
- 건조한 기후이기 때문에 상대습도의 영향력이 낮다.

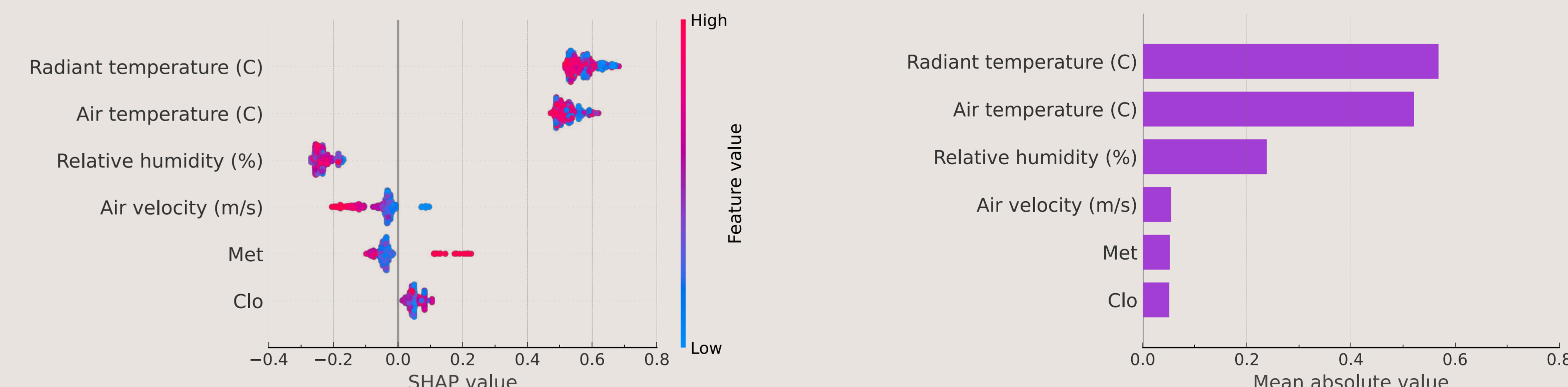


Fig 9. Af (열대 우림 기후) SHAP 적용 결과

- 습한 기후이기 때문에 복사온도와 공기온도 다음으로 상대습도의 영향력이 높다.
- 활동에 따른 발열량의 값이 증가하면 급수적으로 TSV 예측에 양의 영향력을 가진다. 이는 습한 기후이기 때문에 활동에 의해 땀이 발생해도 증발을 방해하기 때문인 것으로 보인다

결론

- 머신러닝을 사용하여 기후대별 열 쾌적감 예측 분석
 - 온난 습윤 기후의 Decision tree 모델을 제외하고는 모든 기후대에서 머신러닝 모델들이 PMV 모델보다 나은 성능을 보이는 것을 확인하였다.
 - 특히 세 기후대의 MSE 값을 평균했을때 PMV 모델은 1.6인데 비해 랜덤포레스트 모델은 0.9로 MSE 값이 약 0.7 정도 감소했다.
- 해석 가능한 인공지능 기법을 사용하여 머신러닝 모델의 의사 결정 과정 분석
 - 실내 공기온도와 복사 온도의 기여도가 모든 기후대에서 크게 나타났다.
 - 더운 사막 기후대에는 의류 단열 수준의 기여도가 상대적으로 높고 상대습도의 기여도가 낮은 것으로 확인하였다.
 - 반면 열대우림 기후에서는 상대습도의 기여도가 공기온도, 복사온도 다음으로 높게 나타났다.

이러한 연구 결과를 기후대에 따라 중요한 변수가 달라지며 중요도의 차이는 특정 기후에서 특정 요인이 열 쾌적성에 영향을 크게 미칠 수 있음을 시사한다. 따라서 열 쾌적 예측에 있어 입력 변수에 대한 보정 계수를 상수로 설정한 PMV 모델로 예측하는 것보다 기후대별로 입력 변수의 가중치를 새롭게 고려하여 예측하는 것이 더욱 효과적임을 확인하였다.