2025 학년도 2학기
융합연구학점제 결과보고서

<table>
<tr><td>팀명</td><td colspan="2">SOMA</td><td rowspan="2">프로젝트유형
(해당사항에 ☑ 표시)</td><td colspan="2" rowspan="2">☑ 학술연구(조사, 분석/실험, 실습)
☐ 발명,발견/사업,기획/창업
☐ 정책제안/사회문제 해결
☐ 문화/예술 프로젝트(전시, 공연, 창작)
☐ 생성형 AI를 활용한 창의적 콘텐츠 제작</td></tr>
<tr><td>프로젝트 제목</td><td colspan="2">외향성 정도에 따른 호감도 예측</td></tr>
<tr><td rowspan="8">팀원정보</td><td>구분</td><td>이름</td><td>학과</td><td>학년</td><td>학번</td></tr>
<tr><td>팀장</td><td>정다은</td><td>인공지능융합전공</td><td>2</td><td>2024311822</td></tr>
<tr><td rowspan="5">팀원</td><td>윤혁준</td><td>문헌정보학과</td><td>2</td><td>2024313423</td></tr>
<tr><td>이시온</td><td>인공지능융합전공</td><td>2</td><td>2024314257</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">지도교수</td><td>성명</td><td colspan="4">이장원</td></tr>
<tr><td>소속</td><td colspan="2">소프트웨어융합대학 실감미디어공학과</td><td>직위</td><td>부교수</td></tr>
<tr><td>연락처</td><td colspan="2">02-760-0557</td><td>e-mail</td><td>leejang@skku.edu</td></tr>
</table>

- 결과보고서는 A4 15장 이내로 하며, 이슈 탐색 및 관련 문제 해결을 위한 구체적인 방안을 포함하여 작성한다.
- 첫 페이지는 결과 보고서를 요약한 초록을 1장 이내로 구성하고, 바로 이어서 보고서 전문을 기재한다.
- 시작품 제작 또는 소비자 대상 테스트 등 결과물이 산출되었을 경우, 붙임 자료(사진, 그림파일 등)로 별도 작성하여 제출한다.
- 결과보고서 구성은 팀별 주제에 맞추어 창의적으로 작성할 수 있다.(단, 기승전결의 구조를 갖추어야 함)

<참고 사항>

- 결과보고서 작성시 글자크기 12pt, 글자체 휴먼명조, 장평 97%, 자간 -2%, 들여쓰기 12pt, 줄간격 180%로 하고, 한글(hwp) 편집용지 A4(국배판 210*297mm)로 한다.
- 제목의 부여 및 글자크기
  - 1단계: Ⅰ, Ⅱ, Ⅲ…(20, 진하게, 휴먼명조, 가운데 정렬)
  - 2단계: 1, 2, 3…(15, 진하게, 휴먼명조, 양쪽 정렬)
  - 3단계: 가, 나, 다…(12, 진하게, 휴먼명조, 양쪽 정렬)
  - 4단계: (1), (2), (3)…(11, 굴림, 양쪽 정렬)
  - 5단계: (가,) (나), (다)…(11, 굴림, 양쪽 정렬)

초록

본 연구는 짧은 시간 동안 대면 상호작용이 이루어지는 스피드 데이트 환경에서, 참가자의 외향성(HEXACO Extraversion) 수준에 따라 호감 형성에 기여하는 비언어적 행동의 결정적 시점이 달라지는지를 규명하는 것을 목적으로 한다. 기존 연구들이 성격을 고정된 속성으로 보거나 단일 시점의 행동 데이터만을 활용한 것과 달리, 본 연구는 행동의 시간적 흐름(Early vs. Late)과 개인의 성향이 호감 형성 예측에 미치는 상호작용 효과를 탐구하였다.

이를 위해 MatchNMingle 데이터셋의 스피드 데이트 세션을 활용하였으며, ResNet3D(R3D), DINOv2, YOLOv11-Pose 등 최신 딥러닝 모델을 도입하여 시공간적(Spatiotemporal) 맥락과 의미론적(Semantic) 특징을 추출하였다. 가설 검증을 위해 영상을 전반부(Early)와 후반부(Late)로 분할하여 구간별 호감 예측 성능(AUC)을 비교하였으며, 참가자를 성별과 외향성 수준에 따라 세분화하여 분석하였다.

연구 결과, 전체 집단을 대상으로 한 분석에서는 구간별 예측 성능의 유의미한 차이가 발견되지 않았으나, 성별과 외향성을 결합한 집단 분석에서는 통계적으로 유의미한 행동 패턴의 차이가 확인되었다. 고외향성 남성 집단은 전반부의 호감 예측 성능이 후반부보다 유의미하게 높게 나타나($p<0.05$) 초반의 탐색적 행동이 호감 형성에 중요함을 시사했다. 반면, 고외향성 여성 집단은 후반부의 예측 성능이 전반부보다 유의미하게 높게 나타나($p<0.05$), 라포 형성 이후의 행동이 호감 판단의 핵심 신호임을 보였다. 또한, DINOv2 모델로 뽑은 시공간적 피처로 학습시킨 MLP 모델이 AUC 0.67 수준의 가장 안정적인 성능을 보였다.

결론적으로 본 연구는 단일 모달리티(Vision-only)와 오버헤드 카메라 시점의 한계에도 불구하고, 성별과 성격 특성에 따라 호감 신호가 발현되는 타이밍이 상이함을 실증하였다. 이는 향후 인간의 사회적 상호작용을 분석하는 AI 모델이 개인의 특성에 따라 시간적 가중치를 유동적으로 적용해야 함을 시사한다.

#Nonverbal Behavior #Extraversion
#Temporal Dynamics #Interpersonal Attraction

Ⅰ. 서론

1.1 연구의 배경 및 필요성

사람들은 짧은 시간안에 상대에 대한 호감과 관심을 빠르게 판단한다. 예를 들어 연애 리얼리티 프로그램을 보면, 시청자들은 출연자들의 대화 뿐만 아니라 표정, 시선,

몸짓, 거리감, 반응 속도같은 비언어적 신호로 누가 누구에게 마음이 있는지 없는지를 놀랄 만큼 정확하게 추측한다. 사실 이미 여러 연구들에서 사람들은 상대의 대화를 전혀 듣지 않아도 비언어적 신호만으로 상대에게 호감이 있는지 없는지를 추론할 수 있다는 것을 보여줬다(Liu, 2023). 실제로 미국의 커뮤니케이션 교수 Judee K. Burgoon은 여러 연구들을 종합하여, 사회적 의미의 약 60~65%가 말의 내용이 아니라 비언어적 행동에서 비롯된다고 보고했다(Krämer, 2008). Burgoon 교수는 40년 넘게 비언어적 행동, 대인 커뮤니케이션, 사회적 신호 처리 등을 연구해 왔으며, 그녀의 이론은 지금도 인간 행동 분석 분야의 대표적 기준으로 인용되고 있다. 특히 그녀는 표정, 시선, 몸짓, 자세, 공간적 거리같은 비언어적 단서가 말로 표현된 내용보다 더 빠르게 처리되며, 사람들의 첫인상 형성과 호감 판단에도 강력한 영향을 미친다고 강조한다. 이렇게 사람의 비언어적 신호는 언어적 신호만큼 중요하며, 때로는 언어적 신호보다 사회적 의미 형성에 더 강력한 영향을 미칠 수 있다. 특히 비언어적 신호는 상대적으로 사람이 의도적으로 조작하기 어렵기 때문에 사람의 진짜 감정과 태도를 더 정확하게 반영할 수 있다고 한다 (Ekman, Friesen).

최근에는 이러한 인간의 ‘사회적 신호 이해 능력’을 AI가 모방하려는 시도가 활발해졌다. Socially Intelligent AI는 주요 연구 분야로 부상하고 있고, Google, Meta, OpenAI 등 주요 기업들은 표정, 시선, 목소리 억양, 제스처, 신체 움직임과 같은 신호를 멀티모달 방식으로 결합해 사회적 맥락을 이해하는 모델을 개발하고 있다. 하지만 AI가 인간의 사회적 의미를 정확히 파악하기에는 여전히 어려움이 많다. 비언어적 신호는 사람마다 다르고, 문화마다 다르고, 상황마다 달라서 기계는 매번 신호를 읽는데 어려움을 겪는다. 다른 말로 AI는 같은 행동이 상황에 따라 다른 의미를 갖는다는 것을 이해하기 어렵다.

이러한 흐름 속에서 스피드 데이팅 환경은 비언어적 신호가 사회적 판단(소개팅 상대방을 “다시 만나고 싶은가?”)으로 이어지는 과정을 관찰할 수 있는 최적의 실험이다. 본 연구에서 사용하는 MatchNMingle 데이터는 짧은 시간 동안의 대면 상호작용 영상을 제공하고, 그 후 즉각적인 호감 선택은 라벨로 제공하기 때문에, 비언어적 행동이 실제 선택으로 연결되는지 탐구하는 데 높은 학술적 가치를 지닌다.

1.2 선행 연구 소개 및 고찰

Veenstra & Hung(2011)은 스피드 데이트 환경에서 정면 얼굴이나 음성 정보 없이, 천장(Top-down) 카메라의 비디오 단서만으로 호감과 매력도를 예측하는 연구를 수행하였다. 남녀 16명(각 8명)이 수행한 총 64회의 데이트에서 위치와 움직임 정보만을 추출해 SVM과 kNN 모델을 학습시킨 결과, 연락처 교환 의사와 매력도 예측에서 약 70% 수준의 정확도를 달성하였다. 이 연구는 표정이나 시선 같은 정면 정보 없이도 거시적인 신체 행동만으로 사회적 의미 파악이 가능함을 실증했다는 점에서 의의가 있다. 그러나 단일

세션 내의 행동만을 분석했을 뿐, 참가자의 성격 차이나 사전 맥락과 같은 개인적 요인을 고려하지 않았다는 한계가 존재한다.

Vargas-Quiros et al.(2023)은 동일한 MatchNMingle 데이터셋(399개 세션)을 활용하여, 웨어러블 가속도 센서 신호가 호감도와 맺는 관계를 분석하였다. 3축 가속도계에서 추출한 움직임 강도, 주파수 및 상대방과의 동조(convergence) 지표를 분석한 결과, 여성은 호감이 높을수록 움직임이 감소하는 등 성별에 따른 패턴 차이가 확인되었으며, 특히 두 사람의 움직임이 수렴하는 정도가 호감 예측의 가장 강력한 신호임을 밝혀냈다. 하지만 가속도 센서는 행동의 양(Quantity)은 측정할 수 있어도, 고개 끄덕임이나 자세와 같은 행동의 '의미적 정보'는 포착하지 못해 풍부한 사회적 신호를 놓친다는 기술적 한계를 지닌다.

Cabrera-Quiros et al.(2019)은 Mingle 세션의 웨어러블, 오디오, 비디오 데이터를 멀티모달로 결합해 성격(Big-5)을 추정하였다. 행동을 웨어러블 움직임(W), 발화 상태(S), 근접성(P) 등 5가지 모달리티로 세분화하여 분석한 결과, 단일 모달리티보다 결합 시 성능이 향상됨을 입증했다. 특히 외향성(X)의 경우, 회귀 분석에서 '말하는 동안의 움직임(WS)'이 가장 낮은 예측 오차(MSE 0.23)를 보여 핵심 예측 변수임이 확인되었다. 그러나 HEXACO 점수의 중앙값을 기준으로 데이터를 이진 분류하는 과정에서 경계선에 있는 참가자들이 강제로 분할되어 정확도가 저하되는 문제가 있었으며, 핸드크래프트 특징에 의존하여 복잡한 시공간적 맥락을 정교하게 포착하는 데에는 한계를 보였다.

마지막으로 Azuma et al.(2025)은 행동의 종류뿐만 아니라 그 행동이 '언제' 나타나는지가 매칭 성공에 중요한 영향을 미친다는 점을 대규모 데이터셋(MMSD)을 통해 분석하였다. 10분간의 데이트를 1분 단위(M1~M10)로 분할하여 영상, 음향, 언어 특징을 학습시킨 결과, 특정 구간(M3, M4, M10)의 데이터만으로도 전체 세션을 사용한 것과 유사한 성능을 보임을 확인하였다. 이는 행동의 등장 시점이 중요한 예측 변수임을 시사한다. 다만, 해당 연구는 참가자의 성격 척도를 모델에 포함하지 않아, 개인의 성향이 이러한 행동 타이밍에 미치는 상호작용 효과를 규명하지 못했다는 아쉬움이 있다.

1.3 연구의 목적

본 연구는 스피드 데이팅이라는 제한된 시간과 공간 속에서, 개인의 성격 특성인 외향성이 행동의 발현 타이밍과 결합하여 상대방의 호감 예측에 어떠한 영향을 미치는지 규명하는 것을 목적으로 한다. 기존 연구들은 행동의 총량이나 평균적인 움직임에 집중했으나, 본 연구는 **"같은 행동이라도 언제 하느냐가 중요하다"** 는 점에 주목한다. 구체적으로, 외향성이 높은 사람은 상호작용 초반(Early)의 행동이, 외향성이 낮은 사람은 라포가 형성된 후반(Late)의 행동이 호감 예측에 더 유효한 단서가 될 것이라는 가설을 검증하고자 한다. 이를 위해 Overhead View 영상에서 추출한 비언어적 특징을 시점별로 분할 분석하고, 성격 요인과의 상호작용 효과를 딥러닝 모델을 통해 실증적으로 분석한다.

Ⅱ. 방법 및 결과

2.1 연구 대상 및 절차

가. 데이터셋 소개

MatchNMingle 데이터셋은 사회적 상호작용을 분석할 수 있도록 설계된 멀티모달 인간 행동 데이터셋이다. 본 데이터셋에는 크게 설문, 상호작용이 담긴 영상(오디오는 음성이 겹쳐 사용하기 어려움), 웨어러블 센서, 참가자들의 정면 사진, 그리고 Manual Annotations가 포함된다. 본 연구에서는 이 중 HEXACO 성격 검사의 외향성 지표, 참가자들이 스피드 데이트 후 적은 Date Response, 스피드 데이트 영상, 그리고 참가자들의 정면 사진(스피드 데이트 영상 속 인물 매칭 목적)을 분석에 활용하였다.

영상은 총 9개의 overhead camera를 통해 수집이 되었으며, 참가자들은 테이블을 순환하며 매번 새로운 사람과 약 3분 동안 스피드 데이트를 진행하였다. 실험은 총 3일간 진행되었으며, 첫째 날에는 남녀 각 16명, 둘째, 셋째 날에는 남녀 각 15명의 참가자가 참여하였다. 각 3분 데이트가 종료될 때마다 참가자는 즉시 상대방에 대한 짧은 설문(Date Response)을 작성하였으며, 이는 상대에 대한 호감과 향후 상호작용 의향을 평가하는 문항들로 구성된다. 예를 들어 “이 사람을 다시 만나고 싶은지“ (이진형 문항)과 “이 사람을 얼마나 다시 보고 싶은지“, “친구로서의 가능성“, “단기적 성적 매력“, “장기적 로맨틱 매력“을 0~7점의 척도로 평가하는 문항들이 포함된다. 본 연구에서는 “이 사람을 다시 만나고 싶은지“의 이진형 문항만 호감도 레이블로 활용할 것이다. 또한 각 실험 내내 참가자들은 모두 웨어러블 가속도 센서를 몸에 차고 있다.

본 데이터셋은 스피드 데이트 세션과 Mingle 세션으로 구성된다. Mingle 세션은 스피드 데이트 이후 다수의 참가자가 한 공간에서 자유롭게 상호작용하는 장면을 촬영한 영상으로 이루어져 있다. 그러나 본 연구는 1:1 상호작용에서 나타나는 비언어적 신호에 초점을 두고 있기 때문에, 전체 데이터셋 중 스피드 데이트 세션만을 분석 대상으로 선정하였다.

나. 연구 절차

본 연구의 목적은 스피드 데이트 장면에서 외향성이라는 개인 특성이 행동 시점(초반 vs. 후반)에 따라 호감(Yes) 예측력에 어떤 차이를 만들어내는지를 규명하는 것이다. 즉, 본 연구가 다루는 질문은 다음과 같다: HEXACO 외향성(X) 점수가 높을수록 초반 행동이, 낮을수록 후반 행동이 상대방이 나에게 호감을 느낄지(yes 선택) 예측하는 데 더 중요하게 작용하는가? 이 연구는 “외향적인 사람이 더 호감인가?” 를 묻는 것이 아니다. 대신, 외향

성이라는 특성 자체가 행동의 ‘언제’가 중요한지를 결정하는지를 확인하는 것이다. 따라서 연구의 초점은 Early Behavior + Extraversion 기반 모델의 예측력 vs Late Behavior + Extraversion 기반 모델의 예측력의 비교 분석이다.

2.2 측정 도구

본 연구에서는 참가자의 외향성 정도와 상호작용 결과를 정량화하고, 영상 데이터로부터 유의미한 비언어적 신호를 포착하기 위해 다음과 같은 측정 도구와 특징 추출 방식을 적용하였다.

가. HEXACO 외향성 척도

참가자들의 HEXACO 성격 검사의 지표 중 X, 외향성(Extraversion) 점수를 활용하였다. 원본 데이터는 1점(전혀 그렇지 않다)에서 5점(매우 그렇다)의 리커트 척도(Likert Scale)로 구성되어 있어, 이를 0~1 범위로 Min-Max 정규화를 진행하여 피처로 사용하였다. 이를 통해 서로 다른 스케일을 가진 변수들이 혼재될 경우에 발생할 수 있는 가중치 편향을 방지하고자 하였다.

나. MatchNMingle 데이트 설문 “Would you like to see your partner again?” (1/0 라벨)

스피드 데이트에서 상호 호감 여부를 판단하기 위해 MatchNMingle 데이터셋에서 제공하는 Date Response(설문) 데이터를 예측값으로 활용하였다. 6개의 설문조사 질문 중 “Would you like to see your partner again? (상대방을 다시 만날 의향이 있습니까?)“에 대한 이진 분류된 응답을 GT(Ground Truth)로 사용하였다. (재만남 의사가 없는 경우 ’0(No)’, 있는 경우 ’1(Yes)’)

다. 딥러닝 모델

전처리된 영상 데이터로부터 다차원적인 비언어적 정보를 포착하기 위해, 세 가지 상이한 특성의 딥러닝 모델을 활용하여 임베딩 벡터를 추출하였다.

(1) R3D (ResNet 3D)

영상의 시공간적 특징을 동시에 학습하는 모델로, 참가자의 동적인 움직임과 행동 패턴을 포착하기 위해 사용하였다.

(2) DINO (Self-distillation with no labels)

Vision Transformer(ViT) 기반의 자기지도학습 모델로, 레이블 없이도 이미지 내의 의미론적 맥락과 중요 객체에 대한 어텐션 정보를 추출하는 데 활용하였다.

(3) Pose Estimation(YOLOv11)

영상 내 인물의 관절 포인트(Keypoints)를 추적하여, 신체 움직임 관련 정보를 수치화하기

위해 사용하였다.

라. 행동 타이밍 지표 (Early vs. Late Feature)

비언어적 신호가 상호작용의 시간적 흐름에 따라 어떠한 영향력을 갖는지 분석하기 위해, 약 3분의 데이트 영상을 시간 축을 기준으로 분할하여 Early, Late Feature를 비교 분석하였다. 50:50, 20:20 2가지 방법으로 구간을 분할하였다.

(1) 50:50 분할

전체 데이트 영상을 정확히 절반으로 나누어, 전반부(Early)와 후반부(Late)의 예측 확률값을 통해 어느 시점이 결과 판단에 더 크게 기여하는지 확인하였다.

(2) 20:20 분할 (양극단 비교)

첫인상(Primacy Effect)과 최신 효과(Recency Effect)를 극명하게 대조하기 위해, 영상의 맨 앞 20%와 맨 뒤 20% 구간에서의 모델 예측 확률을 비교하였다.

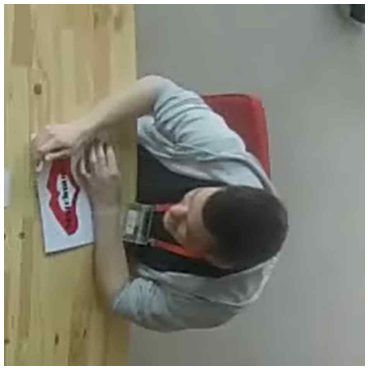
2.3 분석 방법

가. 영상 전처리 과정

영상 전처리 과정은 크게 공간적/시간적 분할, 개체별 추출, 데이터 정제, 레이블링의 단계로 진행되었다. 원본 영상에서 ROI(Region of Interest) 설정을 통해 커플 테이블 영역을 크롭하고, 실제 데이트가 이루어지는 약 3분 구간을 추출하였다. 이후 각 참가자를 224x224 픽셀로 개별 크롭 및 리사이즈 하였으며, 신체가 잘리거나 결측치가 있는 데이터를 제외하여, 총 943개의 유효 영상 데이터를 확보하였다. 확보된 943개의 영상 데이터에 대해 메타데이터를 부여하고, 파일명을 다음과 같은 형식으로 통일하였다.

- 형식: day_본인id_상대id_성별
- day: 실험 일자(day1, day2, day3)
- id*: 각 day별 참가자에게 부여된 고유 번호
- 성별: 남성은 ‘A’, 여성은 ‘B’로 표기
- 예시: day1_01_24_B (1일차 1번 여성이 24번 파트너와 상호작용한 영상)

* 각 참가자는 id가 적혀있는 목걸이를 착용하고 있었으며, 영상 속 숫자(id)가 잘 보이지 않은 경우에는 Participation_day1/2/3 파일에서 참가자 사진+id 목록을 참조하여 영상 속 인물의 본인 id와 상대방의 id를 매칭하였다.

[그림1] 전처리된 영상 예시

나. 데이터 전처리

(1) 데이터 유의성 검정

분석에 앞서 변수 간 상관관계를 검정하였다. Welch 독립표본 t-검정 결과, 상대에게 호감을 선택받은 집단(Yes)이 그렇지 않은 집단(No)보다 외향성(X) 점수가 유의하게 높았다($p<0.05$)*. 로지스틱 회귀분석에서도 외향성 점수가 높을수록 호감 선택 확률이 증가함(OR=2.123)을 확인하였다. 이를 통해 외향성 변수는 호감도 예측에 영향을 끼치는 변수임을 확인하였다. 반면, 외향성 점수가 낮은 참가자일수록 데이트가 반복되는 후반부 순서에 이르러서야 본연의 매력이 발현되어 긍정적 선택(label '1')을 받을 가능성이 높을 것이라 여겨 변수 후보로 둔 '데이트 순서'는 호감도 예측과 통계적 유의성이 없어 최종 피처에서 제외하였다.

따라서 특징 추출 전 최종 피처 데이터셋은 다음과 같이 구성하였다. 칼럼은 video_name, subject_id, partner_id, label, sex, X(HEXACO)로 총 6개로 구성하였다.

label은 설문조사 응답 결과인 0('No'), 1('Yes')로, subject_id와 partner_id는 총 3개의 숫자를 결합하여 만들었다.

- 성별(1/2) + day(1/2/3) + id(day별 개인 번호)
- 예시: 여성 + day1 + 1번 -> 2101

또한 성별에 따라 다른 결과를 보인 선행연구에 기반하여 sex 피처를 남성을 A, 여성을 B로 범주화하여 추가하였다.

이후 모델 학습 및 성능 평가에 사용할 train set과 test set을 구축하였다. 데이터셋 분할 시 클래스 불균형을 방지하기 위해 stratify=y 옵션을 활용하여 층화 추출(Stratification)을 적용했으며, 이는 원본 데이터셋의 클래스 비율(0/1)이 훈련 데이터(Train

* p-value는 귀무가설(두 구간의 AUC 차이가 없다)이 참일 때, 지금과 같은 수준의 차이가 관측될 확률을 의미한다. $p < 0.05$이면 이러한 차이가 우연히 발생했을 가능성이 5% 미만이므로 통계적으로 유의한 차이가 존재한다고 해석한다. 따라서 본 연구에서 보고된 p-value는 Early-Late 예측력 차이가 실제로 존재하는지를 판단하는 근거로 사용된다.

set)와 테스트 데이터(Test set)에 동일하게 유지되도록 구성하였다.

(2) 특징 추출 모델 선정 및 최적화

본 연구는 비언어적 정보 포착을 위해 R3D, DINOv2, Pose Estimation 세 가지 접근법을 비교 실험하였다.

(가) R3D (ResNet 3D)

시공간 정보 학습을 위해 Kinetic-400으로 사전 학습된 R3D 모델의 512차원 특징을 추출하여, 입력 길이(T16~T54)에 따라 Logistic Regression(LR), Random Forest(RF), MLP를 비교 실험하였다. 초기에는 T16 조건의 RF(F1 0.63)가 가장 우수하여 이를 기준으로 파라미터 최적화를 수행하였다. LR은 C=[0.01~10], RF는 트리 수[300, 500]와 깊이[12, 24], MLP는 은닉층[(256,128,64), (128,64,32)]과 학습률 등을 조정하여 성능을 탐색하였다. 이후 신뢰도 향상을 위해 Sigmoid Calibration을 적용했으나 F1(0.69) 상승 대비 AUC(0.50~0.52) 하락이 발생했고, PCA(32, 64, 128차원) 실험에서도 32차원 기준 AUC 0.57로 상승했으나 F1이 하락하는 Trade-off가 관찰되었다. 이에 단순 특징 추출의 한계를 극복하고자 모델 전체를 학습하는 End-to-End Fine-tuning 방식을 도입하였다. 배치 2, 학습률 1e-4, AdamW로 8 epoch 동안 미세 조정한 결과, AUC 0.571, F1 0.684를 기록하며 가장 균형 잡힌 일반화 성능을 확보하였다.

(나) DINOv2 (Vision Transformer)

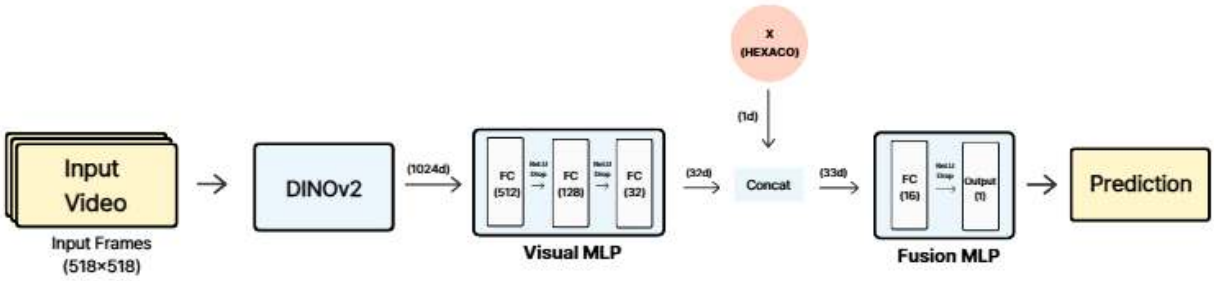
이미지의 의미론적 맥락 포착을 위해 DINOv2-Large(1024차원) 모델을 활용하였으며, 사전 학습 성능을 극대화하고자 224x224 영상을 518x518로 업스케일링하여 입력하였다. 초기에는 1024차원 특징과 외향성 점수를 단순 결합해 학습시켰으나(최고 AUC 0.676), 불안정한 학습 양상을 보였다. 이를 개선하기 위해 프레임 확장(T32), Pooling 방식 변경(Mean+Max), Giant 모델 도입 등을 시도하였으나, 정보량 증가에 따른 노이즈 및 과적합 문제로 명확한 성능 우위를 보이지 못했다. 이에 최종적으로 **Visual MLP(1024→512→128→32)**를 도입하여 시각 정보를 핵심만 압축해 안정화한 뒤, 외향성 정보(1차원)와 결합(Concatenate)하는 Late-fusion 구조를 채택하였다. 이 33차원 벡터를 최종 분류기(Fusion MLP)로 학습시킨 결과, Epoch 11에서 AUC 0.671을 기록하며 가장 안정적인 일반화 성능을 확보하였다.

(다) Pose Estimation

영상의 추상적 특징(Feature)이 아닌 신체 움직임 그 자체를 정량화하기 위해 YOLOv11-pose 모델을 도입하였다. R3D 실험과의 일관성을 유지하고자 입력 시퀀스(T)를 16, 32, 48, 54 프레임으로 설정하였으며, 각 프레임에서 검출된 17개 관절 포인트의 좌표와 신뢰도(x, y, conf)를 1차원 벡터로 변환하여 입력값으로 사용하였다. 해당 데이터를 기반으

로 Logistic Regression, Random Forest, MLP 모델을 비교 실험한 결과, Random Forest가 가장 안정적인 성능을 보였다. 특히 T=16과 T=48 조건에서 각각 AUC 0.62 (F1 0.667), AUC 0.61 (F1 0.667)을 기록하며 가장 우수한 수치를 나타냈으나, 딥러닝 기반의 특징 추출 방식(R3D, DINO)에 비해서는 전반적으로 낮은 예측력을 보였다.

결과적으로, 본 연구의 가설 검증(타이밍 분석)에는 가장 안정적이고 높은 성능을 보인 DINOv2-Large 기반의 Late-fusion MLP 모델을 메인으로 활용하였다.



[그림2] DINOv2 기반 Late-fusion 모델 구조도

2.4 분석 결과

```
early20 shape: (236, 1031)
late20  shape: (236, 1031)

Merged shape: (236, 5)
   video_name  label  X(hexaco)  prob_early20  prob_late20
0  d2_36_08_A      0   0.562500      0.000004     0.088655
1  d2_09_28_B      1   0.625000      0.999319     0.999567
2  d2_07_30_B      0   0.500000      0.024573     0.920213
3  d3_05_35_A      0   0.671875      0.999250     0.934695
4  d3_08_21_A      0   0.500000      0.008305     0.009971

Median X: 0.625

High group size: 122
Low  group size: 114

===== AUC (20% 구간) =====
X-high: early20 AUC = 0.6097 | late20 AUC = 0.6771
X-low : early20 AUC = 0.6972 | late20 AUC = 0.5836

===== BOOTSTRAP AUC DIFF (20% 구간) =====
X-high: diff = -0.0674, 95% CI = [-0.1675, 0.0283], p = 0.5080
X-low : diff = 0.1136, 95% CI = [0.0222, 0.2094], p = 0.5075
```

[그림3] 20:20 분할 가설 검증 결과

본 연구의 가설은 다음과 같다: “외향성이 높은(X-high) 개인은 데이트 초반에 더 강한 호감 신호를 보이고, 외향성이 낮은(X-low) 개인은 데이트 후반 신호가 더 예측적일 것이다.” 이를 검증하기 위해 각 스피드 데이팅 영상을 DINOv2-Large 모델로 처리하여 초반 20%와 후반 20% 구간에서 각각 16프레임을 균일 샘플링하여 시각적 feature(1024D)를 추출하였다. 이 두 구간에 대해 동일한 MLP 분류기를 학습하여, 초반-only 모델과 후반-only 모델의 예측 확률을 test set에 산출하였다. 이후 test set에서 HEXACO Extraversion의 중앙값(0.625)을 기준으로 참가자를 X-high / X-low로 이분화하였다. 각 그룹 내에서 early20 AUC와 late20 AUC를 비교하였으며, 차이의 통계적 유의성은 부트스트랩(bootstrap, n = 2000)을 통한 신뢰구간 및 p-value로 평가하였다.

X-high 그룹에서는 가설과 반대로 후반(0.6771)이 초반(0.6097)보다 높게 나타났으며, X-low 그룹에서는 초반(0.6972)이 후반(0.5836)보다 높게 나타나는 패턴을 보였다. 가설과는 반대지만 표면적으로는 두 그룹 모두 "초반 vs 후반" 차이가 존재하는 것처럼 보인다. 하지만 두 구간의 AUC 차이를 diff = AUC(early20)- AUC(late20)로 정의하여 부트스트랩 검정을 수행한 결과 X-high, X-low 두 그룹 모두 early-late 차이가 통계적으로 유의하지 않았다.

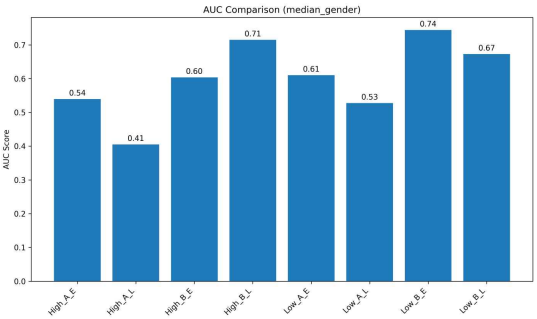
```
====== AUC TABLE ======
     group  early_auc  late_auc
0  X-high   0.593794  0.604417
1   X-low   0.663272  0.589815

===== X 상/하위 20% BOOTSTRAP AUC DIFF =====
X-high20: diff = 0.0062, 95% CI = [-0.1137, 0.1275], p = 0.9165
X-low20 : diff = 0.0658, 95% CI = [-0.0742, 0.2054], p = 0.5220
```

[그림4] 50:50 분할 가설 검증 결과

이전 분석에서는 영상의 앞, 뒤 20%만을 사용하여 가설을 검증하였다. 그러나 더 넓은 시간적 구간을 반영하기 위해 추가적으로 early 50% (전반부)와 late 50% (후반부)에서 각각 16프레임을 균일 샘플링하고, DINOv2-Large 모델로 시각 feature(1024D)를 추출한 뒤 동일한 MLP 분류기를 적용하였다. 이렇게 얻은 prob_early, prob_late 값을 이용해 HEXACO Extraversion의 중앙값 기준(X-high / X-low)으로 나누어 그룹별 AUC를 비교하였다. 또한 early-late 차이의 통계적 유의성은 bootstrap 방식(n = 2000)으로 평가하였다. X-high 그룹에서는 후반부 AUC가 소폭 높았으나 차이는 매우 작기에 두 모델 간 예측 성능은 사실상 유사하다고 볼 수 있다. X-low 그룹에서는 가설과 반대로 early AUC (0.6633)이 late AUC (0.5898)보다 더 높았다. 하지만 bootstrap 유의성 검정 결과, X-high와 X-low 두 그룹 모두 early-late 차이가 통계적으로 유의하지 않았다.

	A	B	C	D	E	F	G	H	I	J	K
1		N	Early_AUC	Late_AUC	Early_CI	Late_CI	Expected	Actual	Match	Significant	p_value
2	High_A	45	0.539525692	0.40513834	(np.float64	(np.float64	Early > Late	Early > Late	TRUE	TRUE	0.0113372366
3	High_B	77	0.603703704	0.714814815	(np.float64	(np.float64	Early > Late	Late > Early	FALSE	TRUE	0.0237043236
4	Low_A	69	0.61025641	0.527350427	(np.float64	(np.float64	Late > Early	Early > Late	FALSE	FALSE	0.272438226
5	Low_B	45	0.744047619	0.672619048	(np.float64	(np.float64	Late > Early	Early > Late	FALSE	FALSE	0.223019371

[그림5] 성별 추가 가설 검증 결과



[그림6] 각 AUC 값 바 그래프

초반, 후반 50%로 나누어 얻은 prob_early, prob_late 값에 대해 참가자를 성별(A: 남성, B: 여성)과 HEXACO 외향성(High/Low) 수준에 따라 네 그룹으로 나누어 AUC를 비교하였고, 구간별 차이의 통계적 유의성을 평가하였다.

분석 결과, 고외향성 여성 집단(High_B)에서는 후반부(Late) AUC가 0.715로 전반부(Early) AUC 0.604보다 뚜렷하게 높았으며, 유의성 검정 결과 p-value는 0.023으로 통계적으로 유의미한 차이를 보였다. 이는 외향적인 여성의 경우 라포가 형성된 후반부에 호감 신

호가 더 명확해짐을 시사한다. 반면, 저외향성 여성 집단(Low_B)에서는 가설과 반대로 전반부(Early) AUC가 0.744로 후반부(0.673)보다 높게 나타나 상반된 경향을 보였다. 하지만 유의성 검정 결과 p-value는 0.223으로, 이 그룹의 Early-Late 성능 차이는 통계적으로 유의하지 않았다.

고외향성 남성(High_A)의 경우, 가설과 일치하게 전반부(Early) AUC가 0.540으로 후반부(Late) AUC 0.405보다 높게 나타났으며, 유의성 검정 결과 p-value는 0.011로 통계적으로 유의미한 차이를 보였다. 이는 외향적인 남성의 경우 초반의 적극적인 탐색 행동이나 첫인상이 호감 판단에 중요한 영향을 미친다는 점을 시사한다. 저외향성 남성(Low_A) 또한 가설(Late > Early)과 달리 전반부 AUC(0.610)가 후반부(0.527)보다 높았으나, 유의성 검정 결과 p-value는 0.272로 통계적으로 유의하지 않았다.

Ⅲ. 논의 및 결론

3.1 연구의 목적 및 의미

본 연구는 스피드 데이팅 상황에서 참가자의 외향성이 상대의 호감 (Yes/No)을 예측하는 과정에서 어떤 방식으로 적용하는지를 탐색하는 것을 목적으로 하였다. 특히 단일한 전체 행동 정보가 아니라, 스피드 데이트의 ‘초반(early)’과 ‘후반(late)’ 구간에서 나타나는 행동의 상대적 중요도가 외향성 수준에 따라 달라질 수 있다는 가능성에 주목하였다. 본 연구가 세운 가설은 다음과 같았다. 외향성이 높은 참가자(X-high)의 경우, 자신의 사회적 활력이나 주도성이 대화 초반부터 드러나기 때문에 초반 행동 신호가 상대방의 호감 예측에 더 큰 역할을 할 것이다. 반면 외향성이 낮은 참가자(X-low)의 경우, 초기에는 상대적으로 표현이 적거나 어색함이 나타나, 대화 후반부에 안정된 태도와 상호작용 품질이 예측력에 더 기여할 것이다. 따라서 본 연구는 ‘외향성 × 시간(timing)’의 상호작용 효과를 검증함으로써 스피드 데이팅 상황에서 개인 성향에 따라 호감 형성 과정에 기여하는 순간이 다를 수 있다는 점을 밝히고자 했다.

3.2 결과의 시사점

본 연구의 분석 결과, 외향성 수준에 따라 호감 예측의 결정적 구간이 달라지는 양상은 특정 집단에서 뚜렷하게 관찰되었다.첫째, 고외향성(High) 집단 내 성별에 따른 타이밍 차이가 확인되었다. 외향성이 높은 남성(High_A)은 가설과 일치하게 초반(Early) 신호의 예측력이 유의하게 높았다. 이는 남성의 경우 초반의 적극적인 탐색이나 주도적인 태도가 호감 형성에 결정적임을 시사한다. 반면, 외향성이 높은 여성(High_B)은 후반(Late) 구간의 예측력이 더 높게 나타났다. 이는 여성이 라포 형성 이후에 보여주는 안정적인 반응이 호감

을 판단하는 더 강력한 신호로 작용함을 의미한다.

둘째, 저외향성(Low) 집단의 신호 모호성이 확인되었다. 저외향성 집단에서는 Early-Late 구간 간의 예측 성능 차이가 유의하지 않았으며, 전반적인 AUC 수치 또한 고외향성 집단 대비 낮은 경향을 보였다. 이는 내향적인 성향의 참가자 자체가 비언어적 표현의 강도가 낮고 미세하여, 오버헤드(Overhead) 카메라 시점의 시각 정보만으로는 구간별 차이를 포착하기 어려웠던 것으로 해석된다.

셋째, 모델의 유효성이다. 본 연구에서 제안한 DINOv2 기반 Late-fusion 모델은 AUC 0.67 수준의 성능을 기록하였으며, 이는 가속도 센서 기반의 선행 연구(Vargas-Quiros et al., 2023)와 동등하거나 상회하는 수준이다. 이는 웨어러블 장비 없이 비접촉 영상 데이터만으로도 유의미한 호감 예측이 가능함을 보여준다.

또한 본 연구의 결과는 단순히 남녀 간의 매칭 예측을 넘어, 인공지능이 인간의 사회적 신호를 해석하는 패러다임에 중요한 시사점을 제공한다. 첫째, ‘개인 맞춤형 감성 컴퓨팅(Personalized Affective Computing)’의 필요성을 제시한다. 기존의 감정 인식 AI는 모든 인간의 행동을 동일한 가중치로 분석하는 획일적 접근을 취해왔다. 그러나 본 연구는 개인의 성격과 성별에 따라 사회적 신호의 ‘골든타임(Golden Time)’이 다름을 실증하였다. 이는 향후 소셜 로봇이나 AI 에이전트가 사용자의 성향을 먼저 파악하고, 그에 맞춰 적절한 타이밍에 반응 가중치를 두는 ‘초개인화된 상호작용’ 기술로 발전해야 함을 시사한다.

둘째, 인간의 사회적 기술 향상을 위한 객관적 지표로서의 활용 가능성이다. 본 연구에서 밝혀진 성향별 호감 발현 타이밍은, 사회적 의사소통에 어려움을 겪는 이들을 위한 AI 기반 커뮤니케이션 코칭 시스템 개발에 구체적인 가이드라인을 제공할 수 있다. 이는 기술이 단순한 분석 도구를 넘어, 인간의 관계 형성을 돕고 사회적 고립을 완화하는 긍정적인 매개체로 확장될 수 있음을 보여준다.

3.3 한계 및 제언

본 연구는 스피드 데이팅 상황에서 외향성과 비언어적 행동의 타이밍이 갖는 예측적 의미를 탐색했다는 점에서 의의가 있으나, 몇 가지 한계가 존재한다.

첫째, MatchNMingle 데이터셋의 오버헤드(Overhead) 카메라는 탑뷰에 가까운 시점으로 촬영되어 있어, 참가자들의 얼굴 표정, 시선, 미세한 감정 반응과 같은 정서적 신호를 충분히 포착하기 어렵다. 즉, 실제 데이트 상황에서 중요한 역할을 하는 섬세한 비언어적 단서를 활용하지 못했다는 점에서 시각 정보의 표현력이 제한적이었다.

둘째, 본 연구는 음성을 활용할 수 없어 단일 모달리티(vision-only) 환경에서 진행되었다. 스피드 데이트에서는 말투, 억양, 웃음, 말의 길이, 발화 속도 등 음성, 언어적 신호가

호감 형성에 결정적인 역할을 하는데, 이를 포함하지 못한 것은 모델의 예측력을 제한하는 주요 요인 중 하나였다.

셋째, 데이터셋의 표본 수가 충분히 크지 않다는 점도 한계로 작용했다. 특히 HEXACO 외향성 지표에 따라 상/하위 그룹으로 나누거나 시간 구간(early vs late)을 세분화할 경우, 각 그룹의 샘플 수가 급격히 줄어들어 통계적 검정의 안정성이 낮아지고, AUC 차이가 작게 나타난 경우 p-value가 크게 나오는 결과로 이어졌다.

이러한 한계를 고려할 때, 후속 연구에서는 멀티모달리티(vision + audio)를 사용하고, 얼굴과 몸이 더 명확히 보이는 영상을 활용하는 것이 필요하다. 또한 더 큰 규모의 데이터셋을 확보하거나, 시간 구간을 더 미세하게 나누어 분석할 수 있을 만큼 충분한 표본을 수집하는 것이 바람직하다. 이를 통해 외향성과 행동 타이밍 간의 실제 상호작용을 더욱 정교하게 규명할 수 있을 것이다.

Ⅳ. 참고 문헌

Azuma, N., Shikama, D., Ogushi, A., Onishi, T., Ishii, R., & Miyata, A. (2025). What timing and behavior patterns determine speed dating success in Japan? In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25) (Article 594, 1–6). Association for Computing Machinery. https://doi.org/10.1145/3706599.3720028

Cabrera-Quiros, L., Gedik, E., & Hung, H. (2022). Multimodal self-assessed personality estimation during crowded mingle scenarios using wearables devices and cameras. IEEE Transactions on Affective Computing, 13(1), 46–59. https://doi.org/10.1109/TAFFC.2019.2930605

Krämer, N. C. (2008). Nonverbal communication. In Human behavior in military contexts (pp. 150–188). National Research Council. https://doi.org/10.17226/12023

Liu, M. (2023). Nonverbal communication conveys more meaning than verbal communication [Bachelor's thesis, Università di Bologna]. Alma Mater Studiorum — Università di Bologna Institutional Repository. https://amslaurea.unibo.it/id/eprint/31762/1/Mengfan%20Liu%20tesi%202022_2023%3B.pdf

Vargas-Quiros, J., Kapcak, Ö., Hung, H., & Cabrera-Quiros, L. (2023). Individual and joint body movement assessed by wearable sensing as a predictor of attraction in speed dates. IEEE Transactions on Affective Computing, 14(3), 2168–2179. https://doi.org/10.1109/TAFFC.2021.3138349

Veenstra, A., & Hung, H. (2011). Do they like me? Using video cues to predict desires during speed-dates. In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops) (pp. 838–845). IEEE. https://doi.org/10.1109/ICCVW.2011.6130339