

레이더 기반 비접촉식 생체정보 모니터링 플랫폼 실증을 통한 고령층 1인가구 활동 데이터 분석 및 고찰

오현택*, 유종민*, 정창진**, 최준균^o

Data Analysis and Consideration of Radar-Based Contactless Biometrics Monitoring Testbed for Single Elderly Households

Hyeontaek Oh*, Jongmin Yu*, Chang Jin Jung**, Jun Kyun Choi^o

요약

고령층 1인가구용 헬스케어 서비스를 위해서는 사용자의 생체정보를 인식하고, 그에 맞는 서비스를 제공할 수 있는 환경을 구성해야 한다. 모니터링 방식 중 하나로, 사용자 편의성과 가격 등을 고려하여, 레이더 기반의 비접촉식 생체정보 모니터링 방식이 연구되고 있다. 이번 논문에서는 레이더 기반의 비접촉식 생체정보 모니터링 플랫폼 실증사업을 통해 수집된 51명의 참여자 데이터를 바탕으로 그들의 활동 상태를 예측하는 데이터 분석 연구를 진행하였다. 실증 데이터의 분석을 통해 다양한 특징을 확인하였으며, 잘 알려진 기계학습 분석 방법들을 활용하여 데이터를 기반으로 사용자들의 상태 예측을 진행하였다. 또한, 분석 결과를 바탕으로 레이더 기반 비접촉식 생체정보 수집 장치를 통한 고령층 1인가구 모니터링 환경의 의의와 한계를 고찰하였다.

키워드 : 레이더 기반 비접촉식 모니터링, 고령층 1인가구, 헬스케어, 데이터 분석, 기계학습

Key Words : Radar-based Contactless Monitoring, Single Elderly Household, Healthcare, Data Analysis, Machine Learning

ABSTRACT

To provide healthcare services for single elderly households, it is necessary to deploy an environment that can monitor their biometric information. A radar-based contactless biometrics monitoring system is considered one of the possible solutions for monitoring single elderly households due to its convenience and cost-efficiency. This paper performs data analysis of radar-based contactless biometrics monitoring testbed for single elderly households with 51 participants and anticipates their activities based on the measured biometric information. Based on the dataset, the features and the characteristics are analyzed, and well-known machine learning techniques are applied to anticipate the activities of participants. Besides, with the results, it is considered that the significances and the limitations of the radar-based contactless biometric monitoring system for single elderly households.

※ 본 연구는 산업통상자원부와 한국산업기술평화원의 “국제공동기술개발사업”의 지원을 받아 수행된 연구결과임. (과제번호 No. 0011879)

• First Author : Korea Advanced Institute of Science and Technology, Institute for Information Technology Convergence, hyeontaek@kaist.ac.kr, 정회원

° Corresponding Author : Korea Advanced Institute of Science and Technology, School of Electrical Engineering, jkchoi59@kaist.edu, 종신회원

* Korea Advanced Institute of Science and Technology, Institute for Information Technology Convergence, andrew.yu@kaist.ac.kr

** Kuls, Kuls ICT - Healthcare Team, cjjung@kuls.co.kr

논문번호 : 202012-340-C-RE, Received February 15, 2021; Revised March 8, 2021; Accepted March 15, 2021

1. 서 론

4차산업혁명 시대 ICT 기술의 발전과 함께 스마트 홈의 개념은 지속적으로 변하고 있으며, 최근 스마트 기기의 대중화, 기계학습/인공지능 기술의 발전, 그리고 사물인터넷 장치의 보급으로 스마트홈 기반 서비스가 다양하게 보급되고 있다. 특히, 한국디자인진흥원 보고서^[1]에 따르면, 인구변화와 관련하여 앞으로는 점점 고령화와 1인가구화가 심화될 것으로 나타나고 있으며, 스마트홈 서비스가 가장 빨리 적용될 분야 중 하나로 고령층 1인가구를 위한 사물인터넷 및 데이터 분석 기술을 선정하였다.

고령층 1인가구의 증가는 비단 우리나라뿐만 아니라 세계적인 추세로써 OECD 보고서^[2]에서도 전 세계적으로 1인가구가 지속해서 증가할 것으로 예상하였다. 또한, 2019년 통계청의 장래가구추계^[3]에 따르면 2047년 1인가구의 비중은 전체 37.3%까지 높아질 것으로 예상되며, 그중에서도 가장 높은 비율을 차지하는 것이 70대 이상의 고령층으로 예상된다. 이러한 배경을 바탕으로, 고령층 1인가구에서 생길 수 있는 다양한 건강 문제들(고독사 방지, 기저질환 모니터링 등)을 위한 헬스케어 서비스의 필요성이 늘어나고 있다.

고령층 1인가구용 헬스케어 서비스에서 중요한 요소 중 하나는 서비스 사용자들의 상황을 파악하기 위해 그들을 모니터링하고 데이터를 수집하는 것이다. 헬스케어 서비스를 위한 사용자 모니터링 방식은 크게 접촉방식과 비접촉방식으로 나눌 수 있다. 접촉방식^[4,5]은 스마트워치 등과 같은 웨어러블 디바이스를 통해 데이터를 얻는 방식으로, 사용자가 해당 장치를 지속해서 착용한다는 가정하에 데이터를 모니터링하고 관련 서비스를 제공하는 구조이다. 접촉방식의 경우 손목 착용형 웨어러블 단말 (예: 스마트워치, 밴드 등)들을 활용하는 경우가 많으며, 이렇게 얻은 데이터를 활용하여 사용자의 건강 상태를 확인하고 적절한 서비스를 제공하는 방식이다.

접촉방식의 경우 생체 데이터를 비교적 정확하게 얻을 수 있는 장점이 있지만, 사용자가 계속 모니터링 장치를 착용해야 하는 불편함이 있다. 특히, 고령층의 경우 다른 연령층에 비해 스마트기기에 관한 거부감이 높고 사용률이 낮아^[6] 지속해서 사용자의 상태를 모니터링할 수 없는 경우가 많다. 이러한 상황으로 인해 헬스케어 서비스 제공자들은 고령층 사용자들을 대상으로 적절한 형태의 서비스를 제공하는 데 어려움을 겪고 있다. 이러한 문제점을 해결하기 위하여, 일상생활 시 거부감없이 자연스럽게 헬스케어 서비스

를 수용할 수 있도록 비접촉방식^[7]의 일상 친화적 데이터 수집 기술이 연구되고 있다.

비접촉식 방식 중 하나로 레이더를 기반으로 한 생체정보 모니터링 방식이 있다. 레이더 방식은 인체의 움직임에 의해 반사되는 레이더의 위상변화를 통해 개인의 심박이나 호흡 등 생체정보를 취득하는 방식으로, 다른 방식에 비해 가격이 저렴하고 비침습적인 측정 방식의 장점을 활용하여 다양한 연구^[8-12]가 진행되고 있다.

이번 연구에서는 고령층 1인가구 돌봄 서비스를 위해 설치된 레이더 기반 비접촉식 생체정보 모니터링 장치들을 통해 수집된 대상자들의 호흡, 심박 및 움직임 데이터 분석을 진행하였다. 이러한 분석을 바탕으로, 그들의 활동 상황을 판단할 때 중요하게 작용하는 요소들을 알아보고, 수집된 데이터를 바탕으로 모니터링 대상이 처한 상황을 예측하는 연구를 진행하였다. 특히, 이번 연구에서는 실제 개발된 장치가 구축된 실증 환경에 참여한 51명의 고령층 1인가구 모니터링 데이터를 분석하였다. 이를 통해, 비접촉식 생체정보 수집 장치를 통한 헬스케어 서비스 환경에서 수집된 데이터들의 특징을 살펴보고, 고려하는 환경을 기반으로 한 헬스케어 서비스의 의의와 한계점을 살펴본다.

II. 선행연구

레이더 기반 비접촉식 생체정보 모니터링 관련 기술은 크게 모니터링 장치에서 수집되는 데이터의 정확도를 높이는 장치 고도화 연구와 얻어진 정보를 정확하고 효과적으로 분석하는 연구로 나뉜다.

레이더 기반 비접촉식 생체정보 모니터링과 관련해서는 주로 호흡 및 심장 박동 데이터 모니터링과 관련된 기술들이 연구되고 있으며, IR-UWB (Impulse-Radio Ultra Wide Band) 레이더 기반의 환경에서 연구가 주로 진행되고 있다. 이러한 연구에는 호흡 및 심장 박동의 연속 모니터링에 관한 연구^[8], 수면 모니터링과 관련된 연구^[9], 수면 효율 측정 연구^[10], 심박 모니터링^[11] 등이 해당한다. 특히, 레이더 기반 모니터링 기술 연구는 측정 대상자가 움직이지 않는 환경에서 실시간으로 정확한 데이터(접촉식 측정 방식과 비교하여)를 얻기 위한 방식이 주로 연구되고 있다.

한편, 얻어진 데이터를 정확하게 분석하는 연구들도 진행되고 있다. 특히, 기계학습과 인공지능 기술의 발달로 해당 기술들을 접목하여 호흡 및 심박 데이터를 분석하는 연구들이 다양하게 진행되고 있다. IR-UWB 레이더 기반 장치를 통해 수집된 호흡 신호

를 로지스틱 회귀 기반으로 분류한 연구^[12], 중환자실에서 얻어진 다양한 생체정보 데이터 세트를 기반으로 다양한 딥러닝 기술의 성능을 분석한 연구^[13], 전이 학습(transfer learning) 기반 심전도 데이터 분석 연구^[14] 등이 있다.

기존 연구들에서는 실험을 위해 통제된 환경을 구성하고 비접촉식 모니터링 시스템을 구축한 후, 사용자들의 생체정보 데이터 수집 및 분석을 진행하였다. 기존 연구들과 다르게 이번 연구에서는 실증 환경에서 사용자들이 자유롭게 일상생활을 보냈으며, 자유도가 높은 환경에서 수집된 데이터를 분석하였다는 차이가 있다.

Ⅲ. 레이더 기반 비접촉식 생체정보 모니터링 실증 환경의 고령층 1인가구 데이터 분석

3.1 데이터 수집 환경

이번 연구를 위해 사용된 비접촉식 생체정보 모니터링 장치는 24GHz 레이더 센서를 기반으로 제작되었으며, 사용된 센서에 관한 자세한 사양은 표 1과 같다. 제작된 센서는 고령층 1인가구를 대상으로 한 건강 상태와 생활안전 위험 감지 모니터링을 위한 실증 사업을 통해 부산지역 고령층 1인가구 약 150여 명의 참가자의 생활공간에 설치되었으며, 실증에서는 그림 1과 같다.

비접촉식 모니터링 센서는 참가자들의 침실에 설치되었으며 생성된 데이터는 2분 간격으로 LoRa 망을 통해 중앙 서버에 전달되었고, 프라이버시 보호를 위하여 데이터는 2분 동안 생성된 활동의 평균으로 수집되어 데이터베이스에 저장되었다. 저장된 데이터는 시스템 관리자 혹은 담당자들이 확인할 수 있도록 웹 기반 대시보드 형태로 제공되었다. 또한, 데이터 분석

표 1. 구축된 레이더 기반 비접촉식 생체정보 모니터링 장치 사양

Table 1. A specification for deployed radar-based contactless biometrics monitoring device

Category	Specification
Chip	Sharp DC6M4JN3000
Method	Microwaves (24.05 to 24.25 GHz)
Range Resolution	60cm
Range (Max.)	1.5m (Heartbeats, Breathing) 7m (Body motion)
Directionality	Azimuth: 25°, Elevation: 20°
Error rate	±10% (~3m)



그림 1. 고령층 1인가구용 비접촉식 생체정보 모니터링 실증 예시
Fig. 1. An illustration of contactless biometrics monitoring testbed for a single elderly household

을 위해 수집된 데이터를 활용할 수 있는 데이터 분석 모듈 또한 개발되었다. 데이터 분석을 위해 반출되는 데이터는 모두 반출 전 비식별화를 진행하였으며, 분석을 위해 필요 최소한의 데이터 속성들만 이번 연구에 활용되었다. 그림 2는 이번 연구를 위해 구축된 레이더 기반 고령층 1인가구 생체정보 모니터링 시스템의 개요를 나타낸다.

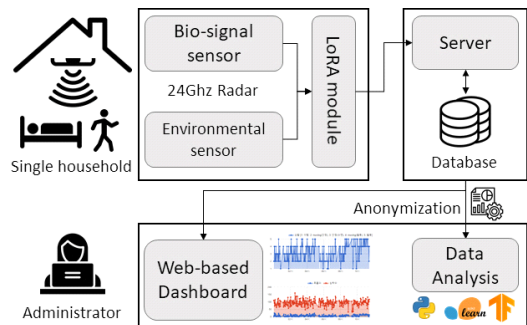


그림 2. 고령층 1인가구용 비접촉식 생체정보 모니터링 실증 개요
Fig. 2. A high-level architecture for contactless biometrics monitoring testbed for single elderly households

3.2 수집 데이터 분석

이번 연구에서는 부산지역 고령층 1인가구 150여 명의 참가자 중, 실증 기간에 문제없이 데이터가 수집된 51명의 참가자의 데이터를 분석하였다. 분석에 활용된 데이터는 총 5가지 종류로 심박수(heart_rate), 호흡수(breath_rate), 활동량(bb_rate), 활동지수(body_motion), 그리고 생체지수(sv_condition)에 관한 정보로 구성되어 있으며, 각 값은 2분 동안 측정된 데이터의 평균으로 계산되어 수집되었다. 그림 3은 실제 수집되어 데이터베이스에 저장된 데이터의 예시를

결과	메시지	rea_date	heart_rate	breath_rate	bb_rate	sv_condition	body_motion
64...		2020-03-16 08:52:05,880	74	12	13	3	15.6
64...		2020-03-16 08:54:13,207	NULL	NULL	NULL	0	NULL
64...		2020-03-16 08:55:56,160	115	9	11	3	9.9
64...		2020-03-16 08:57:56,270	96	11	12	3	13.2
64...		2020-03-16 09:00:03,390	NULL	NULL	NULL	0	NULL
64...		2020-03-16 09:01:50,500	91	12	12	3	14.4
64...		2020-03-16 09:03:48,570	93	12	13	3	15.6
64...		2020-03-16 09:05:44,713	84	10	12	3	12
64...		2020-03-16 09:07:53,287	NULL	NULL	NULL	0	NULL
64...		2020-03-16 09:09:40,993	96	9	11	3	9.9
64...		2020-03-16 09:11:37,107	88	9	11	3	9.9
64...		2020-03-16 09:13:35,273	63	12	11	3	13.2
64...		2020-03-16 09:15:31,320	91	15	11	3	16.5
64...		2020-03-16 09:17:25,453	101	12	11	3	13.2
64...		2020-03-16 09:19:21,557	89	12	11	3	13.2
64...		2020-03-16 09:21:15,670	83	11	12	3	13.2
64...		2020-03-16 09:23:23,420	NULL	NULL	NULL	0	NULL
64...		2020-03-16 09:25:09,933	97	12	11	3	13.2
64...		2020-03-16 09:27:08,083	106	9	13	3	11.7
64...		2020-03-16 09:29:13,450	NULL	NULL	NULL	0	NULL
64...		2020-03-16 09:31:13,470	NULL	NULL	NULL	0	NULL
64...		2020-03-16 09:33:13,483	NULL	NULL	NULL	0	NULL
64...		2020-03-16 09:34:56,540	123	10	13	3	13
64...		2020-03-16 09:37:03,527	NULL	NULL	NULL	0	NULL

그림 3. 실증 환경에서 수집된 사용자 생체정보 데이터 예
Fig. 3. An example of collected biometric information from the testbed

나타낸다.

데이터베이스에 저장된 데이터는 Python을 활용해 CSV(Comma-Separated Values)로 변환되었으며, 머신러닝 데이터 분석에서 널리 쓰이는 Pandas, Numpy, Scikit-learn, Tensorflow 라이브러리를 활용하여 정제되고 분석되었다. LoRa망의 통신오류 등으로 인해 NULL 등의 형태로 정확하게 수집되지 못한 데이터를 정제하고 약 550만 개(1인당 평균 11만 개; 약 5개월분)의 데이터를 확보하였다.

각 데이터는 사용자의 현재 상태를 나타내는 상태 값에 대응되며, state_0~5까지 6가지의 상태로 구분되었다. 각 상태는 부재/감지되지 않음(state_0), 사람 인식(state_1), 안정적인 움직임(state_2), 수면(state_3), 활동적인 움직임(state_4), 기타 분류되지 않는 활동(state_5)을 나타낸다. 이러한 분류를 기반으로, 그림 4는 수집된 데이터에 해당하는 각 상태 값의 분포를 나타낸다.

분류된 데이터를 살펴보면, 생체 모니터링 센서가 주로 침실에 설치되었다는 물리적인 제약사항으로 인해 참여자들이 센서 밖에서 활동하는 경우가 많아, 부재(state_0) 상태가 가장 많이 관측되었으며 그다음으로 수면(state_3), 안정적인 움직임(state_2) 등의 순으로 나타났다.

먼저, 수집된 5가지 종류의 데이터 간의 연관도를 분석하기 위하여 각 데이터의 상관관계 분석을 진행하였다. 그림 5는 각 데이터 간의 산점도 행렬(scatter matrix)을 나타낸 것이다. 산점도 행렬의 X축과 Y축

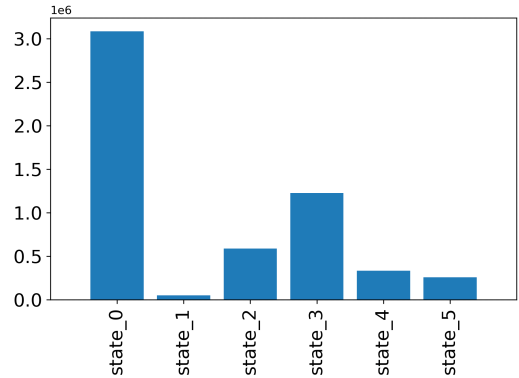


그림 4. 수집된 사용자 상태 정보 데이터 분포도
Fig. 4. A distribution of collected user status from dataset

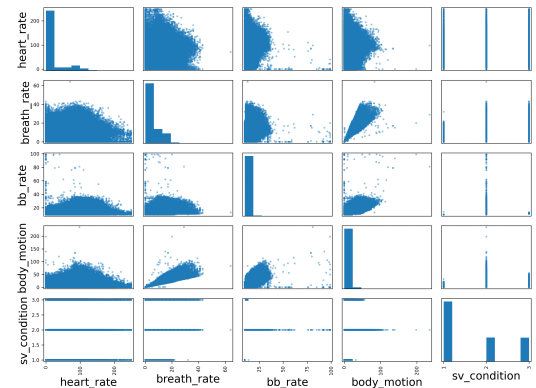


그림 5. 생체정보 연관도 분석을 위한 산점도 행렬
Fig. 5. A scatter matrix for collected biometric information

이 교차하는 부분이 각 속성 간 연관성을 나타내기 위한 산점도 분포를 나타내며, 대각선 요소들은 각 속성의 히스토그램을 나타낸다.

산점도 행렬을 살펴보면 그림 4에서 확인하였던 것과 같이 구축된 실제 환경의 한계상 부재중으로 인식된 데이터 수가 많아, 각 데이터도 0(인식되지 않음)과 가까운 값이 많이 수집된 것을 히스토그램(대각선 요소들)을 통해 알 수 있었다.

강한 상관관계를 보이는 요소 조합은 호흡수(breath_rate)와 활동지수(body_motion) 조합과 활동량(bb_rate)과 활동지수의 조합이었다. 특히, 전자의 경우 강한 양의 상관관계를 가져 활동지수의 증가에 따라 호흡수도 선형적으로 증가하는 것을 확인할 수 있었다. 그 외에도 생체지수(sv_condition)와 활동지수 및 활동량의 관계도 다른 조합에 비해 뚜렷한 것을 알 수 있다.

한편, 사용자의 건강과 밀접한 연관이 있는 심박수(heart_rate)나 호흡수(breath_rate)는 다른 속성들과 연관관계가 적어 사용자의 활동 상황을 판단하는데 크게 영향을 미치지 않으리라고 예상되었다. 이는 두 데이터 모두 정규분포와 비슷한 형태를 따르고 있어 다른 데이터들과 조합하였을 때 특징적인 상관관계를 추출하기 어려운 부분이 있기 때문이다.

이러한 데이터 해석을 바탕으로 입력 데이터의 각 속성에 따른 특성 중요도를 정량적으로 분석하기 위하여 1,000개의 결정트리(Decision Tree)로 구성된 랜덤포레스트(Random Forest) 분류기를 활용하여 각 특성의 중요도를 파악하였다. 그림 6은 각 데이터 속성의 특성 중요도를 나타낸다. 각 속성을 특성 중요도 순으로 나열하면 생체지수(sv_condition; 49.9%), 활동량(bb_rate; 30.4%), 활동지수(body_motion; 13.4%), 호흡수(breath_rate; 5.0%), 심박수(heart_rate; 1.3%) 순으로 나타났으며, 그림 5를 통한 산점도 행렬 분석으로 예측된 결과와 비슷한 순으로 나타난 것을 알 수 있었다.

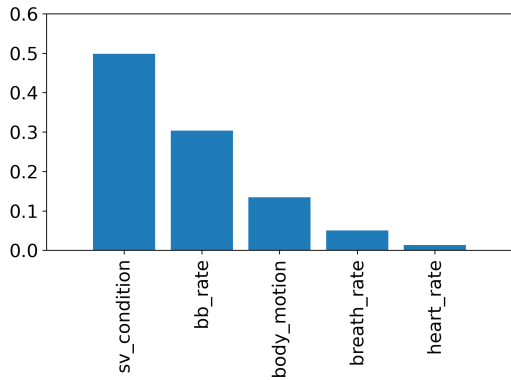


그림 6. 각 데이터 속성의 특성 중요도
Fig. 6. An illustration of feature importance for each attribute

3.3 지도학습 기반 상황 예측 기법 비교

다양한 지도학습 기반 분류기를 적용한 상황 예측 방식을 비교하기 위하여 주어진 데이터의 Min-Max 정규화를 진행하였으며, 이를 통해 모든 데이터 값은 0~1 사이의 값으로 변환되었다. 이번 지도학습 문제는 답이 여러 클래스인 다중클래스 분류(Multi-class classification)이기 때문에, 각 분류기의 성능을 평가하기 위하여 정확도(Accuracy)뿐만 아니라 F1-점수(F1-score)도 같이 평가되었다. F1-점수는 정밀도(Precision)와 재현율(Recall)의 조화평균으로 정의된

척도로, 각 답에 해당하는 데이터의 수가 불균형할 때 도 분류기 모델의 성능을 평가할 수 있는 척도이다.

지도학습 기반의 분류기 성능 분석을 위해서 총 7가지의 분류기를 선정하였다. 로지스틱 회귀(Logistic Regression), 결정트리(Decision Tree), 랜덤포레스트(Random Forest), k-최근접 이웃(k-Nearest Neighbors; kNN), 선형(linear) 및 커널(kernel) 기반 서포트 벡터 머신(Support Vector Machine; SVM), 다층 퍼셉트론(Multi-Layer Perceptron; MLP). 다층 퍼셉트론을 제외한 각 분류기는 Scikit-learn 라이브러리를 활용하여 구현 및 평가되었으며, 각 분류기의 하이퍼파라미터는 정확도 관점에서 최고의 성능을 도출하는 결과를 선정하였다. 한편, 다층 퍼셉트론의 경우 Tensorflow 라이브러리를 활용하여 구현/평가되었다. 입력 데이터의 차원(5차원)을 고려하여 이번 논문에서는 3계층(각 계층은 64-32-8순으로 유닛을 가짐) 구조의 ReLU 활성화 함수를 활용한 구조를 제안하였으며, 3계층 처리 후 마지막으로 softmax 계층을 통해 분류 결과를 도출하였다. 배치 크기(batch size)는 256으로 선정하였다.

지도학습 기반 상황 예측 실험을 진행하기 위해서 상태 값의 분포를 유지한 형태로 전체 데이터를 랜덤하게 섞은 다음 전체의 약 60%를 훈련 세트(training set), 약 10%를 검증 세트(validation set), 약 30%를 평가 세트(test set)로 구성하였다. 이러한 환경을 바탕으로 상황 예측 실험을 진행하였으며, 표 2는 각 분류기의 정확도와 F1-점수를 나타낸다.

분류기별 분류 정확도와 F1-점수를 살펴보면 제안된 3계층 구조의 다층 퍼셉트론이 96.82%과 96.42%를 달성하여 가장 높은 성능을 보였으며, 그다음으로 랜덤포레스트와 결정트리가 높은 성능을 보였다. 랜덤포레스트는 다수의 결정트리 결과를 조합하여 결과를 내는 방식이므로, 결정트리와 성능이 거의 같게 나타

표 2. 지도학습 기반 분류기 성능 비교표 (원본 데이터)
Table 2. A performance comparison of supervised learning based classifiers (original dataset)

Classifier	Accuracy	F1-score
Logistic Regression	0.9620	0.9620
Decision Tree	0.9671	0.9633
Random Forest	0.9672	0.9634
kNN	0.9649	0.9610
SVM (linear)	0.8323	0.7894
SVM (kernel)	0.9251	0.9342
MLP (64-32-8 units)	0.9682	0.9642

났다. 로지스틱 회귀, k-최근접 이웃 또한 비슷한 성능을 나타냈다. 입력 데이터의 차원이 비교적 낮은 특성상 비선형 분류기들의 성능은 거의 비슷하게 나타난 것으로 보인다. 한편, 선형 기반의 분류기인 선형 SVM은 상대적으로 낮은 성능을 보였다. 이는 데이터들이 겹치는 부분이 있어 선형 기반의 분류기들이 정확하게 상태 값들(state_0~5)을 분류하기 어려운 환경이라는 것으로 해석할 수 있다. 이번 분류기 분석을 통해 사용자들이 자유롭게 활동을 하는 환경에서도 그들의 활동 상황을 모니터링하고, 모니터링된 데이터를 통해 높은 정확도로 그들의 상황을 예측할 수 있음을 확인할 수 있었다.

3.4 고찰

이번 절에서는 현실 세계에 대규모로 구축된 비접촉식 생체정보 모니터링 장치로부터 수집된 데이터들을 다룸에 있어 발생하는 고려사항들을 살펴보고 이번 연구의 의의와 한계를 고찰한다.

3.4.1 데이터 불균형 문제

앞서 그림 4에서 살펴봤듯이 원본 데이터에 수집된 각 상태 값들(state_0~5)의 양은 불균형하게 구성되어 있다(즉, 환경적 특성상 state_0의 데이터가 다른 데이터들에 비해 많다). 이러한 불균형 때문에 분류기들은 많이 입력된 데이터에 대한 정답을 추론하여 정확도를 높이는 형태로 과적합(overfit)될 가능성이 있다.

각 상태 값이 균등하게 있을 때의 각 분류기의 성능 변화를 확인하기 위하여, 모든 상태 값을 균등하게 하는 업샘플링(부족한 데이터를 늘리는 방법)과 다운샘플링(많은 데이터를 줄이는 방법)을 적용하여 성능을 비교하였다. 업샘플링의 경우 데이터가 가장 많은 상태 값인 state_0의 데이터 수를 기준으로 나머지 상태 값들의 데이터를 늘렸으며(총 약 1,800만 개), 다운샘플링의 경우 데이터가 가장 적은 상태 값인 state_1의 데이터 수를 기준으로 나머지 상태 값들의 데이터를 줄였다(총 약 30만 개). 데이터 업/다운샘플링에는 Scikit-learn의 resample 함수를 활용하였다.

업샘플링 및 다운샘플링 데이터를 원본 데이터의 실험과 비슷하게 각각 약 60%를 훈련 세트, 약 10%를 검증 세트, 약 30%를 평가 세트로 구성한 후, 상대적으로 빠른 속도로 실행되는 4가지 분류기를 선정하여 정확도와 F1-점수를 비교하였다. 표 3을 보면 업샘플링 데이터와 다운샘플링 데이터에 대한 분류기들의 정확도와 F1-점수 성능 모두 원본 데이터에 대한 성능보다 낮게 나타난 것으로 확인되었다.

표 3. 지도학습 기반 분류기 성능 비교표 (업/다운샘플링 데이터)
Table 3. A performance comparison of supervised learning based classifiers (up/down sampling datasets)

	Classifier	Accuracy	F1-score
Up	Decision Tree	0.8317	0.8300
	Random Forest	0.8317	0.8289
	SVM (linear)	0.7436	0.6972
	MLP (64-32-8 units)	0.8344	0.8326
Down	Decision Tree	0.8313	0.8296
	Random Forest	0.8311	0.8282
	SVM (linear)	0.7544	0.7234
	MLP (64-32-8 units)	0.8331	0.8311

3.4.2 데이터 분포 문제

이러한 원인을 살펴보기 위해, 차원 축소 기법을 통해 각 상태 값(state_0~5)을 정확히 분류할 수 있는 결정 경계를 찾을 수 있는지 분석하였다. 이번 연구에서는 t-SNE(t-distributed Stochastic Neighbor Embedding) 기법^[15]을 활용하여 원본 데이터의 차원을 축소하였다. t-SNE는 기본적으로 비슷한 샘플을 가까이, 비슷하지 않은 샘플을 멀리 떨어지도록 하면서 차원 축소를 진행하며, 데이터 시각화에 널리 쓰이는 기법의 하나다. 차원 축소는 데이터의 특성을 살펴보기 위한 것으로 이번 연구에서는 전체 데이터에서 각 상태 값(state_0~5)의 분포를 균등하게 유지하면서 0.1% (약 5,000여 개) 정도의 소수 샘플을 뽑아 차원 축소의 결과를 분석하였다.

t-SNE를 활용하여 데이터를 2차원 평면에 축소한 결과는 그림 7과 같다. 그림 7를 살펴보면 t-SNE의 차원 축소 결과 데이터 비율이 가장 높은 2가지 상태 값들(state_0과 state_3)의 데이터는 특정 분포가 확실하여 차원 축소를 했을 때, 결정 경계가 뚜렷하여 쉽

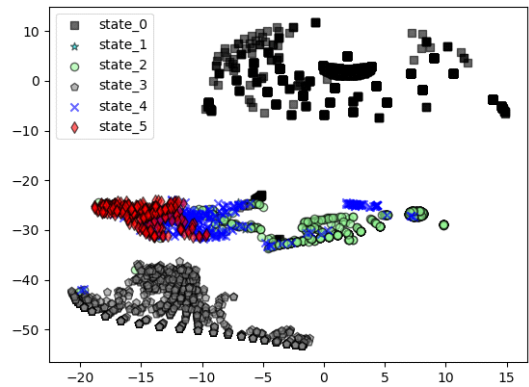


그림 7. t-SNE 기반 차원 축소 결과 (2차원)
Fig. 7. t-SNE based dimension reduction (2-D)

게 구분할 수 있는 것을 알 수 있었다. 원본 데이터 전체 약 550만 개 중 state_0과 state_3의 비율은 합쳐서 약 78%의 비율을 차지하기 때문에 이 두 상태를 맞추으로써 기본적인 예측 정확도의 향상이 이뤄진 것으로 볼 수 있다. 하지만, 나머지 상태 값들(state_1, 2, 4, 5)에 관해서는 데이터 특성이 확실하게 구분되지 못한 것을 확인할 수 있다.

나머지 상태 값들(state_1, 2, 4, 5)의 분포를 더 세밀하게 살펴보기 위하여, 같은 데이터에서 상태 값 state_0, state_3 및 state_1(가장 적은 데이터)과 관련된 데이터를 제거하고 남은 데이터(약 2,000여 개)를 대상으로 다시 한번 t-SNE 차원 축소를 적용하였다. 이번에는 상태 값들을 더 심층적으로 분석하기 위하여 원본에서 3차원 공간으로 차원을 축소하였으며, 그 결과는 그림 8과 같다. 그림 8을 살펴보면 3차원 공간에서의 각 상태 값들의 배치를 파악할 수 있으며, 상태 값 state_2, state_4, state_5의 경우 결정 경계에서 서로서로 얽혀 있음을 확인할 수 있었다. 이런 얽힘과 데이터 자체의 노이즈(noise)가 있어 원본 데이터를 기반으로 분류한 분류기들은 약 96%의 성능에 수렴(표 2)하였으며, 업/다운샘플링을 거친 데이터를 기반으로 분류하였을 때는 약 83%의 성능에 수렴(표 3)함을 확인할 수 있었다.

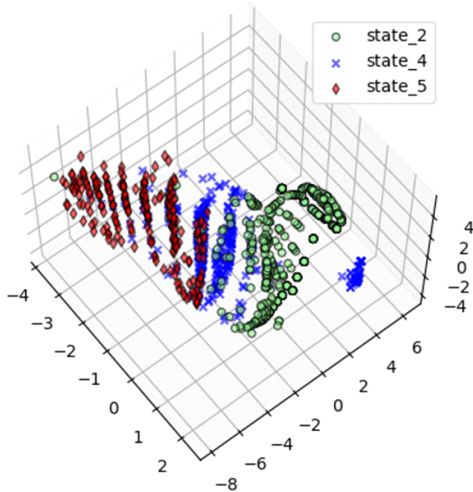


그림 8. t-SNE 기반 차원 축소 결과 (3차원)
Fig. 8. t-SNE based dimension reduction (3-D)

3.4.3 연구의 의의와 한계점

이번 연구에서는 실증 환경에 구축된 약 150여 명의 실증 참여자들에게 각각 설치된 약 150여 개의 레이다 기반 비접촉식 모니터링 센서를 통해 수집된 데

이터를 기반으로 진행되었다. 실증 참여자들은 제약사항 없이 자유롭게 생활하기 때문에, 수집된 데이터는 통제된 환경에서 정확하게 측정된 것이 아니라 현실 세계의 여러 상황이 그대로 반영되었다. 이는 기존의 실험실 환경에서 진행된 다른 연구들과 비교했을 때, 대규모의 실제 데이터 분석을 통해 참여자들의 행동에 관하여 유의미한 예측 정확도를 얻어내었다는 것에 의의가 있다. 하지만, 현실 세계에 구축된 시스템을 그대로 활용하여 연구가 진행되었으므로 그에 따른 다양한 한계점도 존재한다.

먼저, 실제 구축된 장치에서 사용된 LoRa망의 불안정성과 장치의 오류, 대역폭의 한계 등으로 데이터가 온전하게 수집되지 못하거나 일부 기간만 수집된 참여자들도 많이 있었다. 이러한 부분들은 하드웨어를 동반한 실증 환경이 가지는 한계로 추후 비접촉식 센서를 기반으로 한 모니터링 시스템 구축을 대규모로 진행할 때는 서비스 품질을 유지할 수 있는 안정적인 통신 환경이 고려되어야 할 것으로 보인다.

더불어, 기존의 통제된 환경에서 실시간으로 실험 참여자들의 심박과 호흡에 관한 데이터를 수집하는 연구들과는 다르게, 참여자들의 프라이버시 문제와 LoRa망의 성능의 한계로 서버에 수집된 데이터가 장치에서 2분 동안 측정된 데이터의 평균으로 나타남으로써, 다른 연구들에 비해 정밀도가 낮은 한계가 있었다. 특히, 사용자들의 건강 상태와 직접적으로 관련되어 있는 심박수에 관한 데이터의 정밀도가 낮아 추가적인 분석에 한계가 있었다.

IV. 결론

이번 논문에서는 고령층 1인가구 비접촉식 생체정보 모니터링 시스템 기반 헬스케어 서비스 실증 환경에서 수집된 150여 명 참여자의 데이터 중 51명의 데이터를 분석하였다. 통제되지 않은 환경에서 수집된 참여자들의 심박, 호흡 및 운동량 데이터 분석을 통해 고령층 1인가구 사용자의 생체정보와 활동량을 바탕으로 세 가지 카테고리(자비 비율, 활동 및 수면)에 관한 참여자들의 상황을 다층 퍼셉트론을 활용하여 96.8%의 높은 정확도로 예측할 수 있는 것으로 나타났다. 또한, 이러한 예측 결과를 독거노인 돌봄 서비스에 적용할 가능성을 확인하였다. 특히, 이번 연구에서는 통제되지 않은 현실 환경에서 수집된 데이터를 통해서도 참여자들의 상황을 예측할 수 있었다는 것에 의의가 있다. 한편, 하드웨어를 활용한 실증 환경의 특성상 다양한 한계와 문제점도 나타났다. 먼저,

실증 장비는 주로 참여자들의 침실에만 설치되어 참여자들이 다른 공간에서 활동할 때는 모니터링이 불가능하다는 한계가 있었다. 또한, 하드웨어 오류와 LoRa 네트워크 성능의 한계로 인해 수집된 데이터의 정밀도가 떨어지는 것도 이번 실증 연구의 한계로 나타났다. 이번 실증 연구를 기반으로 분석된 다양한 의의 및 한계를 통해 더 향상된 고령층 1인가구 헬스케어 서비스들을 개발할 때 레이더 기반 비접촉식 생체 신호 모니터링 환경을 활용할 수 있을 것으로 기대한다.

References

- [1] T. Kim, et al., *2030 Future Vision Code 4.0*(2018), Retrieved Feb. 8, 2020, from <http://m.designdb.com/index.html?menu=1061&bbsno=30019&boardno=587&ztag=rO0ABXQANDxjYWxsIHR5cGU9ImJvYXJkIiBubz0iNTg3IiBza2luPSJwaG90b19iYnMiPjwvY2FsbD4%3D&siteno=15&act=view>
- [2] OECD, *The Future of Families to 2030* (2011), Retrieved Feb. 8, 2020, from <https://doi.org/10.1787/9789264168367-en>.
- [3] Statistics Korea, *Population Projections 2017-2047*(2019), Retrieved Feb. 8, 2020, from http://kostat.go.kr/portal/korea/kor_nw/1/2/6/index.board?bmode=read&aSeq=377540&sTarget=title
- [4] Y. G. Chen and S. W. Lee, "A study on the trend of development of wearable healthcare devices for the elderly," *J. Korean Soc. Design Culture*, vol. 26, no. 1, pp. 245-260, Mar. 2020.
- [5] H. Mshali, T. Lemlouma, M. Moloney, and D. Magoni, "A survey on health monitoring systems for health smart homes," *Int. J. Ind. Ergonomics*, vol. 66, pp. 26-56, Feb. 2018.
- [6] J. Kim, "How much do we know about the use of smartphones in the silver generation?: Determinants of the digital divide within the silver generation," *Inf. Soc. & Media*, vol. 21, no. 3, pp. 33-64, Dec. 2020.
- [7] M. Bahache, J. P. Lemayian, W. Wang, and J. M. Hamamreh, "An inclusive survey of contactless wireless sensing: A technology used for remotely monitoring vital signs has the potential to combating COVID-19," *RS Open J. Innovative Commun. Techno.*, vol. 1, no. 2, Dec. 2020.
- [8] Y. Shin and J. Park, "A study on the non-contact continuous monitoring of breathing and heart rate using UWB radar technology," in *Proc. Symp. KIICE*, pp. 34-37, Busan, Oct. 2019.
- [9] I. Yun, H.-M. Choi, and H.-W. Oh, "Design and implementation of embedded system for sleep monitoring based on UWB," *J. KICS*, vol. 45, no. 6, pp. 1103-1109, Jun. 2020.
- [10] J. W. Choi, Y. N. Lee, S. H. Cho, Y.-H. Lim, and S. H. Cho, "Sleep efficiency measurement algorithm using an IR-UWB radar sensor," *J. KICS*, vol. 42, no. 1, pp. 214-217, Jan. 2017.
- [11] Y. Lee, J.-Y. Park, Y.-W. Choi, H.-K. Park, S.-H. Cho, S. H. Cho, and Y.-H. Lim, "A novel non-contact heart rate monitor using impulse-radio ultra-wideband (IR-UWB) radar technology," *Scientific Reports*, vol. 8, no. 1, pp. 13053-13062, Aug. 2018.
- [12] C. Park and D. Lee, "Respiration classification based on logistic regression algorithm," *J. KICS*, vol. 46, no. 1, pp. 162-169, Jan. 2021.
- [13] S. Purushotham, C. Meng, Z. Che, and Y. Liu, "Benchmarking deep learning models on large healthcare datasets," *J. Biomed. Informatics*, vol. 83, pp. 112-134, Jul. 2018.
- [14] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation," *2018 IEEE ICHI*, pp. 443-444, New York, U.S.A., Jun. 2018.
- [15] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579-2605, Nov. 2008.

오 현 택 (Hyeontaek Oh)



2012년 2월 : 한국과학기술원 전
산학 학사
2020년 2월 : 한국과학기술원 전
기 및 전자공학 박사
2020년 2월~현재 : 한국과학기술
원 IT융합연구소 팀장

<관심분야> 개인정보 활용, 고신뢰 ICT, 클라우드
[ORCID:0000-0003-0722-0762]

정 창 진 (Chang Jin Jung)



1992년 2월 : 중앙대학교 경영학
학사
2000년 8월 : 중앙대학교 경영학
석사
2018년 2월~현재 : 주식회사 콜
스 헬스케어사업본부장

<관심분야> 디지털 헬스케어, 사물인터넷, 클라우드

유 종 민 (Jongmin Yu)



2013년 8월 : 충남대학교 컴퓨터
공학과 졸업 (학사)
2020년 8월 : 광주과학기술원 전
자전기컴퓨터공학과 박사
2020년 9월~현재 : 한국과학기술
원 IT융합연구소, 연수연구원

<관심분야> 기계학습, 인공지능, 패턴인식

최 준 균 (Jun Kyun Choi)



1998년 2월 : 한국과학기술원 전
자 공학 박사
1997년 2월 : 한국전자통신연구
원 책임연구원
2009년 3월~현재 : 한국과학기술
원 전기 및 전자공학부 정교수

<관심분야> 사물인터넷, 웹, 신뢰네트워크, 에너지