

A Study on the Method of Differentiating Between Elderly Walking and Non-Senior Walking Using Machine Learning Models

Ga Young Kim[†] · Su Hwan Jeong[†] · Soo Hyeon Eom[†] ·
Seong Won Jang[†] · So Yeon Lee[†] · Sangil Choi^{††}

ABSTRACT

Gait analysis is one of the research fields for obtaining various information related to gait by analyzing human ambulation. It has been studied for a long time not only in the medical field but also in various academic areas such as mechanical engineering, electronic engineering, and computer engineering. Efforts have been made to determine whether there is a problem with gait through gait analysis. In this paper, as a pre-step to find out gait abnormalities, it is investigated whether it is possible to differentiate whether experiment participants wear elderly simulation suit or not by applying gait data to machine learning models for the same person. For a total of 45 participants, each gait data was collected before and after wearing the simulation suit, and a total of six machine learning models were used to learn the collected data. As a result of using an artificial neural network model to distinguish whether or not the participants wear the suit, it showed 99% accuracy. What this study suggests is that we explored the possibility of judging the presence or absence of abnormality in gait by using machine learning.

Keywords : Gait Analysis, Machine Learning, Supervised Learning, Neural Network Models

기계학습 모델을 이용한 노인보행과 비노인보행의 구별 방법에 관한 연구

김 가 영[†] · 정 수 환[†] · 엄 수 현[†] · 장 성 원[†] · 이 소 연[†] · 최 상 일^{††}

요 약

보행 분석은 인간의 걸음걸이를 분석하여 보행과 관련된 여러 다양한 정보를 얻기 위한 연구 분야 중 하나로서 의료 분야뿐만 아니라 기계공학, 전자공학 및 컴퓨터공학 등 다양한 학문 분야에서 오랫동안 연구되고 있다. 보행 분석을 통해 걸음걸이에 문제가 있는지를 파악하려는 노력이 꾸준히 이어져 왔다. 본 논문에서는 이러한 보행 이상을 알아보기 위한 전 단계로서 보행 데이터를 활용하여 동일 실험 참가자에 대해 노인 체험복 착용 전후의 걸음걸이를 기계학습 모델에 적용하여 학습시킴으로써 노인 체험복 착용 여부를 구별할 수 있는지를 연구하였다. 총 45명의 실험 참가자들을 대상으로 노인 체험복 착용 전과 후 각각의 보행 데이터를 수집하였고, 총 6개의 기계학습 모델을 이용하여 보행 데이터를 학습시켰다. 신경망 모델을 활용하여 노인 체험복 착용 여부를 판별한 결과 약 99%의 높은 정확도를 보였다. 본 연구에서 시사하는 것은 기계학습을 활용하여 보행의 이상 유무를 판단할 수 있는 가능성을 모색했다는 데 있다.

키워드 : 보행 분석, 기계학습, 지도학습, 신경망 모델

1. 서 론

대한민국은 현재 고령사회(65세 이상 인구 비율이 전체의 14%)를 지나 빠르게 초고령사회(65세 이상 인구 비율이 전체의 20%)로 진입 중이다. 65세 이상 노인의 경우 보행에 문

제가 있는 경우 심혈관 및 근골격계 질환, 치매 발생과 연관 될 가능성이 존재한다는 연구 결과가 발표되었다[1-4]. 보행에 문제가 있음을 발견하기 위해서는 보행 데이터를 웨어러블 기기를 이용하여 수집하고, 이렇게 축적된 많은 양의 보행 데이터를 기계학습 모델을 이용하여 학습 및 분석함으로써 정상 보행과 보행 이상 데이터 사이의 뚜렷한 차이점을 찾아 내려는 노력이 요구된다. 더 나아가 의료기관에 종사하는 의사들의 주관적인 판단에만 의존했던 보행 이상 측정 방법의 대안을 마련하고 더 나아가 검사자의 주관적인 판단에서 벗어나 수치 데이터에 기반을 둔 객관적인 보행 이상 판단 환경이 마련될 필요가 있다.

노인의 보행 이상에 대한 객관적이고 명확한 정의가 마련

※ 이 논문은 2019년도 강릉원주대학교 신입교원 연구비 지원에 의하여 연구되었음.
※ 이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 생애초기연구사업의 지원을 받아 수행된 연구임(No.2020R1G1A1013937).
† 비 회 원 : 강릉원주대학교 컴퓨터공학과 학사
†† 정 회 원 : 강릉원주대학교 컴퓨터공학과 교수
Manuscript Received : June 4, 2021
First Revision : July 16, 2021
Accepted : July 24, 2021
* Corresponding Author : Sangil Choi(schoi@gwnu.ac.kr)



Fig. 1. Elderly Simulation Suit (up) and Components (down)

되어 있지 않은 상황에서 어떤 사람의 걸음걸이를 분석한 후 보행 이상이라고 판단하기는 쉽지 않다[1]. 이에 대한 대안으로 본 연구에서는 서로 다른 2가지 실험 상황을 설정한 후 동일 실험 참가자에 대해 기계학습의 지도학습 모델을 활용하여 보행 데이터를 학습시킴으로써 서로 다른 실험 상황을 구별할 수 있는지를 연구하였다. 실험군과 대조군을 65세 이상의 노인과 노인이 아닌 사람으로 분류한 후 실험을 진행하는 것이 원칙이지만, 코로나19 팬데믹으로 인해 노인들의 보행 데이터를 수집하는 데 어려움이 있어 65세 이상의 노인 보행을 간접적으로 경험할 수 있는 노인 체험복(Fig. 1 참조, *이미지출처: mymedical.co.kr*)을 사용하였다.

보행에 문제가 있는지를 판단하는 명확한 기준이 마련되어 있지 않은 상황에서 서로 다른 보행 패턴을 구별하는 것은 보행 이상 진단 시스템 구축을 위한 첫걸음이 될 수 있다. 동일인에 대해 질병 전과 후에 보행이 어떻게 바뀌었는지를 판단할 수 있을 뿐만 아니라 해당 질병으로부터의 회복 여부를 파악하는 데 있어 객관적인 자료를 제공할 수 있을 것으로 기대한다.

본 논문의 구성은 다음과 같다. 2절에서는 보행 분석과 관련된 최신 연구 동향에 대해서 살펴보고, 3절에서는 보행 실험 및 데이터 수집과 관련된 내용을 다룬다. 4절에서는 3절에서 수집된 보행 데이터의 전처리 과정에 대해 자세히 설명한다. 5절에서는 본 연구를 위한 시스템 구성 요소에 대해 언급하고, 6절에서는 6가지 서로 다른 지도학습 모델을 활용하여 어떻게 학습이 이루어졌는지를 설명하고, 이에 따른 정확도 평가에 대해 논한다. 마지막으로 결론 및 향후 연구 과제를 제시하며 논문을 마무리한다.

2. 관련 연구

웨어러블기기를 이용한 최신의 보행 분석 연구는 보행 주

기 내에서 반복적으로 발생하는 패턴을 찾은 후 패턴에 따른 시간 및 공간적 특성을 관찰하고 분석하려는 노력을 기울여 왔다[3-9]. 대부분의 보행 분석 연구에서 보행 주기 시간, 발 디딤 시간 등 보행 주기 내에서의 특정 사건과 관계된 정보를 찾으려고 노력했다. 국내의 한 연구는 발가락 떼기(TO: Toe Off)와 뒷굽 닿기(HS: Heel Strike) 사이에서 발생하는 보행의 동작을 분석하기 위해 압력센서, 가속도 센서, 그리고 각속도 센서를 사용하여 보행 분석 시스템을 개발하였다[10]. 국외의 한 연구는 관성 측정 장치인 Inertial Measurement Units (IMUs) 4개를 이용해 각 다리의 허벅지와 종아리 부위에 부착하여 실험을 진행하였고 보행 시 관절 각도의 궤도와 보행 주기를 얻어 연구를 진행하였다[11]. 두 연구에서 보는 바와 같이 보행의 특징을 정확하게 얻어내기 위해 보행 시 움직이는 부위에 적어도 3~4개의 센서를 부착하여 실험하였다. 하지만 이는 주로 보행 이상을 가진 노인이나 환자들에게 불편을 초래할 수 있다[2,12].

보행 분석 연구에 있어 가장 최근의 연구 동향은 주로 머신러닝 또는 딥러닝 기술을 적용하는 추세이다[13-25]. 센서로부터 수집한 원시 데이터에서 보행 주기를 알아내고 개인 또는 그룹별로 보행 주기의 특성을 분석하여 원하는 결과를 얻기 위해 보행 분석 알고리즘을 직접 설계한 기존의 방법을 넘어서 머신러닝 모델을 활용하여 수집된 데이터를 바탕으로 모델을 학습시킴으로써 보행 분석을 진행하려는 다양한 시도가 진행 중이다. 기존 보행 분석 연구에 사용된 가속도계, 각속도계 및 깔창 센서는 여전히 기계학습 지향 보행 연구에서 보행 데이터 수집 및 분석에 사용 중이다. 또한, 보행 분석 연구에 사용되는 다양한 유형의 머신러닝 모델이 존재한다. 기계학습을 사용한 보행 분석은 가까운 장래에 점점 더 많은 연구자의 관심을 받을 것으로 예상된다.

본 연구에서는 관련 연구를 통해 기존 연구가 가지는 복잡한 분석 단계를 단순화하기 위해 복잡한 데이터 전처리 과정을 최소화하였고, 보행 주기 내에서의 특정 보행 패턴을 찾으려는 노력보다는 원시 보행 데이터를 기반으로 한 기계학습을 진행하였다. 또한 실험 참가자들의 불편을 최소화하기 위하여 가속도 센서를 하나만 사용했다.

3. 보행 데이터 수집

3.1 실험 대상자

보행 데이터 수집을 위해 총 45명의 실험 대상자를 모집하였다. 실험 참가자는 20세에서 29세 연령대의 신체 건강한 20대 청년(남자 20명, 여자 25명)을 대상으로 삼았다. 모든 실험 참여자들은 전체 실험 진행 과정과 주요 유의사항에 대해 충분히 숙지하는 시간을 가졌다. 이와 더불어 실험을 통해 수집되는 개인 정보를 연구에 활용하는 것에 동의하는 절차를 진행하였다. 인간의 보행 데이터 수집을 위한 실험은 기관생명관리위원회의 심의를 거쳐 실험에 대한 승인(GWNUIRB-2020-33)을 취득하였다.

3.2 노인보행과 비노인보행의 정의

지속적인 코로나19 팬데믹으로 인해 65세 이상의 노인들을 대상으로 하는 보행 데이터 수집이 불가능함에 따라 이에 대한 차선책으로 Fig. 1의 노인 체험복을 실험 참여자에게 착용시켜 실험군과 대조군으로 나누어 실험을 진행하였고, 노인보행과 비노인보행을 다음과 같이 정의하였다. **노인보행은 실험 참여자가 노인 체험복을 착용한 후의 보행을, 비노인보행은 실험 참여자가 노인 체험복을 착용하지 않은 보행을** 의미한다.

3.3 노인 체험복

본 연구에서 사용한 노인 체험복은 ㈜명문의료기의 제품으로 초고령사회를 대비한 노인 체험 교육을 목적으로 제작되었다(특허등록: 10-18384490000호). 이 장비의 특징은 첫째, 팔근력 약화, 팔관절 퇴화, 허리근력 약화, 무릎관절 퇴화 체험을 위한 다양한 부속 장비들이 갖춰져 있어 노인들이 일상적으로 겪는 근력 저하 및 관절의 불편함을 간접적으로 경험할 수 있도록 제작되었다. 둘째, 백내장, 녹내장, 황반퇴화 등과 같은 증상을 특수 제작된 안경을 착용함으로써 직·간접적으로 체험하는 것이 가능하다. 마지막으로, 총 15가지의 다양한 이명 소리를 발생시키는 장비와 헤드폰을 통해 노인이 겪는 난청 및 여러 가지 이명을 체험할 수 있다. 따라서, 본 연구에서는 노인 체험복이 65세 이상의 노인의 보행을 시뮬레이션하기에 가장 적합한 장비로써의 역할을 담당해줄 것으로 예상했다.

3.4 보행 데이터 수집 장비 및 애플리케이션

보행 데이터는 3축 가속도 센서가 내장된 삼성 갤럭시 노트 9 (SM-N960F) 스마트폰을 사용하여 수집하였다. 본 연구를 위해서 보행 데이터 수집용 안드로이드 애플리케이션을 직접 제작하였다. 사람의 걸음걸이는 상하, 좌우, 전후 방향의 영향을 받는다. 3축 가속도 센서는 X(좌우), Y(상하), Z(전후), 이렇게 총 3개 축으로부터의 가속도 센서값을 측정하는 것이 가능하다. 따라서, 인간이 걸을 때 발생하는 3축 방향의 가속도 센서의 값을 보행의 원시 데이터로 활용하였다. Fig. 2는 실험 참가자들의 보행 데이터를 실시간으로 수집하도록 설계된 안드로이드애플리케이션의 실행 화면을 보였다.

3.5 실험 과정 및 장소

총 2명의 실험 진행자가 전체 실험을 담당했다. 첫 번째 진행자는 실험 참여자에게 실험에 대한 설명과 유의사항에 대해 안내하는 역할을 맡았고, 나머지 진행자는 데이터 수집용 애플리케이션에서 획득한 3축 가속도 센서값과 데이터 획득 시간이 데이터베이스에 실시간으로 잘 저장되는지를 모니터링하는 역할을 담당했다. 데이터 수집용 애플리케이션이나 데이터베이스에 문제가 발생하여 실시간 보행 데이터 수집이 원활하지 않다면 다른 실험 진행자에게 즉시 통보하여 실험을 중단한 후 다시 시작하도록 함으로써 만일의 사태에 대비하였다.

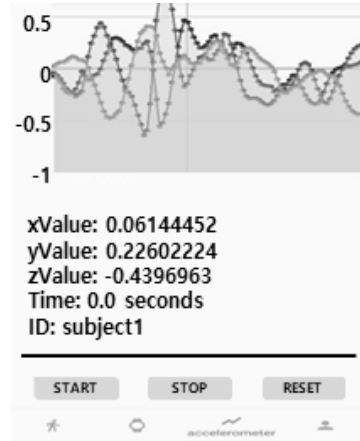


Fig. 2. Android Application for Capturing Gait Data

실험 참가자는 진행자로부터 스마트폰을 건네받아 자신의 바지 앞주머니에 넣은 후 진행자의 안내에 따라 보행을 시작했다. 바지 앞주머니에 있는 스마트폰이 보행 시 움직이게 되면 실험 전에 설정한 3축의 방향이 바뀔 수 있으므로 실험 참여자들에게 몸에 최대한 밀착되는 바지를 입고 실험에 참여하도록 안내했다. 이렇게 함으로써 스마트폰이 실험 참여자들의 바지에서 이리저리 움직이는 것을 사전에 방지했다.

보행 실험은 Fig. 3의 장소에서 이뤄졌다. 실험 코스는 1단계 계단 내려오기, 2단계 평지 걷기, 3단계 내리막길 걷기, 4단계 평지 걷기, 5단계 계단 오르기, 이렇게 총 다섯 단계로 구성하였다. 실험 코스는 현실에서의 걷기 활동을 최대한 반영할 수 있도록 디자인하였고, 따라서 평지에서 걷는 활동이 전체 활동 중에 가장 많은 시간을 차지하도록 했다.

보행 실험에 든 총 실험 시간은 약 1시간으로 설정하였다. 비노인보행 데이터 수집을 위해 Fig. 3에 제시된 실험 코스(①~⑤)를 걷는 데 20분을 설정하였고, 약 20분의 휴식을 취한 후 노인보행 데이터 수집에 20분을 할당하였다. 평균적으로 노인 체험복을 착용하지 않았을 때는 주어진 실험 코스를 총 7번 완주하였고, 착용했을 때는 약 5번 완주하는 데 그쳤다. 이 사실을 통해 노인보행이 실험 참가자들의 보행에 영향



Fig. 3. Experimentation Place and Course

을 미친다는 것을 알 수 있었다. 또한, 남성 참가자보다 여성 참가자들이 노인 체험복을 착용했을 때 더욱 많은 피로감을 호소하였다. 그 이유는 노인 체험복의 구성품이 차지하는 무게로 인해 여성 실험 참가자들의 피로도를 더욱 증가시켰을 것으로 판단했다.

4. 데이터 전처리 및 데이터 셋 구성

4.1 데이터 전처리 프로그램

갤럭시 노트 스마트폰에 내장된 가속도계 센서를 통해 실험자의 보행에 따른 3축 가속도 값과 데이터 측정 시간을 실시간으로 Firebase 데이터베이스로 전송하여 저장했다. Firebase 데이터베이스에서는 기본적으로 데이터를 JSON 파일 형식으로 저장하여 관리한다. JSON 포맷은 정형화된 데이터뿐만 아니라 비정형 데이터를 관리하기에 적합한 데이터구조로써 XML을 대체하여 데이터 전송 등에 많이 이용되고 있다. 이 파일 형식은 "(키, 값)"의 형식으로 데이터를 저장하고 관리한다. 대부분의 프로그래밍언어에서 JSON 포맷을 다룰 수 있는 라이브러리 함수를 제공하고 있다. 본 연구에서는 파이썬 언어를 이용하여 JSON 포맷으로 저장된 3축 가속도 센서 값을 기계학습 모델에서 처리하기 적합한 형태의 CSV 파일 포맷으로 변환하는 데이터 전처리 프로그램을 제작하였다.

데이터 전처리 프로그램은 먼저, JSON 형식의 3축 가속도 데이터를 기계학습 모델로 학습시키기 위해 CSV 파일 형식으로 변환하는 작업을 진행한다. 실험에 참여한 모든 사람의 JSON 형식의 데이터를 입력받아 실험 참가자 ID, X, Y, Z 축의 가속도 데이터, 노인 체험복 착용 여부, 이렇게 총 5개의 항목으로 이루어진 CSV 파일로 변환하여 저장한다.

4.2 데이터 셋 구성

Fig. 3의 실험 코스에서 가장 많은 시간을 소요하는 부분은 ②와 ④의 평지 걷기에 해당한다. 20분의 전체 보행 데이터 중 가장 많은 시간을 소비하는 평지 걷기 데이터를 추출하여 기계학습 모델에 적용하였다. 이를 위해 전처리 프로그램에서 평지 걷기와 연관 있는 가속도 데이터를 찾아서 해당 부분을 따로 저장했다.

보행 실험에 사용된 가속도 센서의 가속도 샘플 주기 (Sampling Rate)는 100Hz로써 초당 100개의 3축 가속도 값을 얻을 수 있다. 이번 연구에서는 10초 동안의 평지 걷기 데이터를 학습에 이용하였다. 따라서, 하나의 데이터 셋은 1,000개의 서로 다른 데이터 샘플로 이루어진다. 평지 걷기에 든 시간은 실험 참가자마다 조금씩 차이는 있었지만 평균적으로 약 40초였다. 그러므로, 한 실험 참가자마다 총 80초 (실험코스 ②와 ④의 평지 걷기에 든 시간)의 평지 걷기 데이터를 가진다. 또한, 모든 실험 참가자들로부터 노인보행과 비노인보행, 이렇게 2가지 보행 데이터를 수집했다.

보행 실험에 걸린 시간은 총 20분이다. 비노인보행의 경우 평지 걷기에 약 80초 정도가 걸렸고, 나머지 활동(계단 오리

Table 1. Dataset Configuration

| Item | Contents |
|---|----------|
| Number of Samples/Dataset | 1000 |
| Time to Set Up the Dataset | 10 secs |
| Time of Flat Walking | 40 secs |
| Number of Flat Walking | 2 |
| Number of repetitions of the Experiment | 5 |

기, 내리막길 걷기 등)에 약 60초가 소요되었다. 노인보행의 경우에는 체험복을 입지 않았을 때보다 실험 코스를 전부 완주하는데 적게는 약 15초에서 많게는 약 30초 정도 시간이 더 걸렸다. 2가지 실험 환경으로부터 얻은 보행 데이터의 수가 서로 같지 않았기 때문에(비노인보행의 경우 평균 7번 반복, 노인보행의 경우 평균 5번 반복) 총 5번의 반복 실험이 이루어진 것으로 설정하였다. Table 1은 기계학습을 위해 획득한 데이터 셋 현황을 정리한 것이다.

Table 1을 기준으로 실험 참가자 45명의 총 데이터 셋을 계산하면 다음과 같다.

$$\begin{aligned} \text{Total Number of Datasets} = & \{ \text{Time of Flat Walking}(40) \\ & \div \text{Time to Set Up the Dataset}(10) \} \\ & \times \text{Number of Flat Walking}(2) \\ & \times \text{Number of repetitions of the Experiment}(5) \\ & \times \text{Number of Participants}(45) \times 2 \end{aligned}$$

위에 제시된 식에 따라 총 데이터 셋의 개수를 계산하면 산술적으로 3,600개가 만들어지는데 한 실험 참가자의 노인 보행 데이터가 Firebase 데이터베이스 연결 문제로 인해 유실되었다. 따라서, 해당 데이터 셋 개수(40개)를 빼고 난 후 총 3,560개의 데이터 셋을 얻었다. 데이터 셋 3,560개는 기계학습에 있어 충분하지 않다는 판단하에 각 데이터 셋마다 평지 걷기의 시작 위치를 무작위로 지정하여 추가로 데이터 셋을 늘리는 작업을 수행하였다. 3,560개의 데이터 셋 모두 10회에 걸쳐 무작위로 평지 걷기의 최초 시작 위치를 변경함으로써 데이터 셋 개수를 증가시켰다. 이렇게 하여 최종적으로 총 35,600개의 데이터 셋을 기계학습에 활용하였다.

5. 시스템 구성 및 학습용 데이터 선정

5.1 시스템 구성

본 연구를 위한 시스템 구성을 Fig. 4에 나타냈다. 보행 분석 시스템은 먼저 Fig. 2의 보행 데이터 수집용 애플리케이션을 통해 3축 가속도 데이터(보행 데이터)와 데이터 측정 시간을 수집하여 Firebase 데이터베이스에 실시간으로 저장한다. 다음으로 데이터 전처리 프로그램이 JSON 형식의 원본 데이터를 기계학습을 위한 CSV 파일 형식으로 변환한 후 평지 걷기 데이터만을 추출하여 저장한 후 기계학습 모델을 학습시

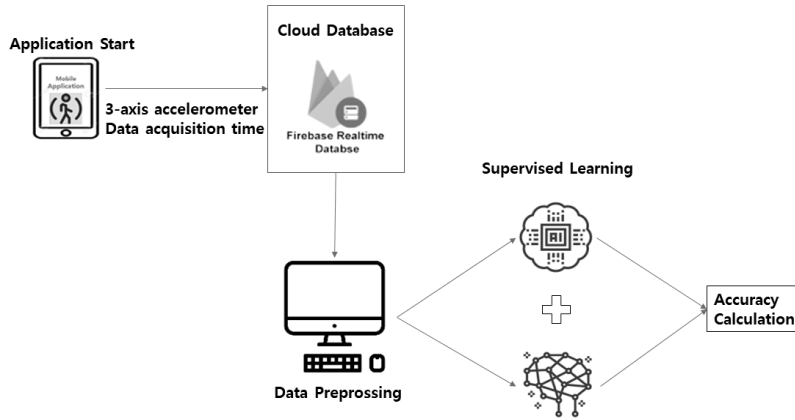


Fig. 4. System Configuration for Gait Analysis

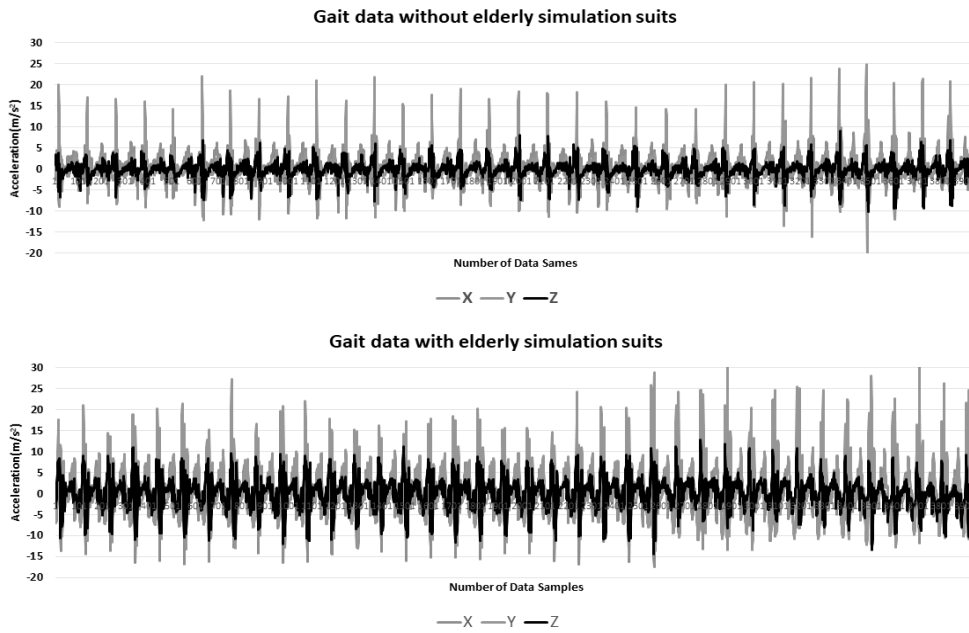


Fig. 5. 3-axis Gait Data With/without Elderly Simulation Suits

키는 데 이용한다. 이렇게 획득한 3축 가속도 데이터값을 그래프로 나타낸 것을 Fig. 5에 보였다.

5.2 기계학습용 데이터 선정

Fig. 5는 노인 체험복 착용 전후의 X, Y, Z 3축의 가속도 데이터값을 나타낸다. 이 그림에서 가로축은 데이터 샘플의 개수를, 세로축은 가속도(m/s²) 값을 표현한다. 또한, X 축은 좌우, Y 축은 상하, Z 축은 전후 방향의 가속도를 나타낸다. 노인 체험복을 착용했을 때와 그렇지 않을 때의 평지 걷기 데이터를 비교하면 노인 체험복을 착용했을 때, X, Y, Z 축 모두 값의 변화가 더 큰 것을 확인할 수 있다. 모든 실험 참가자의 데이터를 분석한 후 X, Y, Z 축 가속도 값의 변화량(Δ_x)을 Equation (1)을 이용하여 계산하였다. 가속도 값의 변화량은 X 축 0.549, Y축 1.838, Z축 2.964로써 노인 체험복

착용 전후의 평지 걷기 데이터에서 가장 큰 변화를 보인 것은 전후 방향(Z 축)의 움직임이었다. 따라서, 노인보행과 비노인 보행을 구별하는 데 있어 전후 방향의 움직임을 나타내는 가속도 값을 기계학습에 이용한다면 동일인의 보행이라고 할지라도 노인 체험복 착용여부를 구별할 수 있을 것으로 예상했다. 이러한 이유로 본 연구에서는 전후 방향을 나타내는 가속도 Z 값만을 기계학습에 활용하였다.

$$\Delta_x = \left| \sqrt{\frac{\sum(x_t - \mu)^2}{N}} - \sqrt{\frac{\sum(\hat{x}_t - \hat{\mu})^2}{N}} \right| \tag{1}$$

where
 x_t = The acceleration of x-axis of elderly waking at time t
 μ = The mean of acceleration of elderly walking
 \hat{x}_t = The acceleration of x-axis of non-senior waking at time t
 $\hat{\mu}$ = The mean of acceleration of non-senior walking
 N = The total number of acceleration of x-axis

6. 기계학습을 이용한 학습 및 정확도 평가

기계학습을 위한 개발 환경은 Google의 Colab에서 이뤄졌다. Colab은 구글에서 제공하는 Jupyter Notebook으로 클라우드 환경을 기반으로 Keras, Tensorflow, Pytorch 등의 파이썬 라이브러리를 별도의 설치과정 없이 사용할 수 있다. 사용자들은 머신러닝 학습을 위한 고가의 하드웨어 및 소프트웨어 구입에 대한 부담 없이 자신의 구글 계정만으로 Google Colab으로부터 머신러닝 개발 환경을 제공받을 수 있다.

노인보행과 비노인보행의 구별을 위해 6가지 서로 다른 기계학습 모델을 선정한 후 Z 축 가속도 데이터의 80%를 학습 데이터로 사용하였고, 나머지 20%를 정확도 분석을 위한 검증 데이터로 설정하였다. 데이터 셋에 대한 다양한 테스트를 진행한 결과, 너무 많은 데이터를 학습에 사용하면 과적합 발생 가능성이 크고, 너무 적은 데이터를 이용하면 데이터 부족으로 인해 정확도가 떨어지는 것을 확인했다. 여러 테스트를 통해 전체 데이터 셋에서 80%를 학습에 사용하는 것이 이러한 문제를 해결할 수 있다는 결론을 얻었다.

본 연구에서 사용한 6가지 기계학습 모델에 대한 간략한 설명 및 특징은 아래와 같다. 파이썬 머신러닝 라이브러리인 scikit-learn (1, 2, 3)과 scikit-learn 호환 라이브러리인 sktime (4, 5)에서 총 5가지 모델을 선정하였고, 파이썬을 위한 딥러닝 프레임워크인 케라스(Keras)에서 제공되는 신경망 라이브러리를 사용하여 학습을 진행했다.

1) DummyClassifier

간단한 규칙을 사용하여 예측하는 분류기 모델로써 본 논문의 주제인 노인보행과 비노인보행을 구별하는 분류 문제에 적합한 모델이다.

2) RandomForestClassifier

여러 개의 의사결정 트리를 만들고 이 결정트리로부터 최종 결과를 얻는 모델로써 사용이 단순하며 분류와 회귀 문제에 모두 사용할 수 있다는 장점이 있다.

3) KNeighborsClassifier

가장 가까운 훈련 데이터의 포인트 하나를 최근접 이웃으로 찾아내어 예측에 활용하는 모델로써 노인보행과 비노인보행 훈련 데이터로부터 두 개의 포인트, 즉 노인보행 및 비노인보행의 클래스를 정하고 가까운 것끼리 그룹화하여 분류를 진행한다.

4) TimeSeriesForestClassifier

결정트리의 앙상블 기법을 활용하는 모델로써 결정트리로부터 랜덤 표본을 추출하고 복잡한 선형 계산을 위해 병렬 컴퓨팅 기법을 활용하는 특징이 있다. 또한, 시간에 따른 중요도 곡선을 바탕으로 유용한 시간적 특성을 포착한다.

5) KNeighborTimeSeriesClassifier

시계열 데이터에서 데이터 집합을 처리하는 모델 중 하나로써 가장 가까운 이웃을 찾아 집합으로 묶어내어 분류하는데 사용한다.

6) CRNN(Convolutional Recurrent Neural Networks)

합성곱 신경망(Convolutional Neural Network)과 순환 신경망(Recurrent Neural Network)을 같이 사용하여 딥러닝 모델을 구축한 신경망 모델이다. 합성곱 신경망을 통해 보행 데이터가 지닌 특징(feature)을 추출한 후 시계열 데이터에 대한 학습을 위해 순환 신경망의 대표적인 모델 중 하나인 LSTM을 사용하였다.

6.1 scikit-learn과 sktime의 기계학습 모델을 활용한 학습

scikit-learn에서 제공하는 모델은 입력 데이터 셋의 형태가 2차원 리스트로 주어져야 하므로 데이터 전처리를 통해 만들어진 데이터 셋을 2차원의 데이터 프레임 형태로 변환하였고, sktime 모델의 경우에는 입력 데이터 셋이 1차원 리스트이어야 하므로 이에 따라 데이터 셋을 변경하는 작업을 먼저 수행하였다. 다음으로 전체 데이터 셋을 학습 데이터 셋(Train dataset)과 시험 데이터 셋(Test dataset)으로 구분하였다. 특히, 학습 데이터 셋을 구성하는 데 있어 전체 데이터 셋에 있는 데이터 샘플을 무작위로 추출하여 학습 데이터 셋과 시험 데이터 셋으로 만들어주는 *train_test_split* 함수를 이용하여 두 데이터 셋을 구성하였다. 세 번째로 앞에서 언급한 총 다섯 개의 서로 다른 모델에 학습 데이터 셋을 입력으로 주어 모델을 학습시킨 후 시험 데이터 셋을 활용하여 정확도를 계산해주는 *score* 메서드를 활용하여 각 모델의 정확도를 구하였다. 각 모델로부터 최상의 정확도를 얻기 위해 정확도에 영향을 미치는 하이퍼파라미터 값을 조정하는 작업도 수행되었다. 모델 학습의 전체 진행 과정을 Fig. 6에 순서도로 표현하였다.

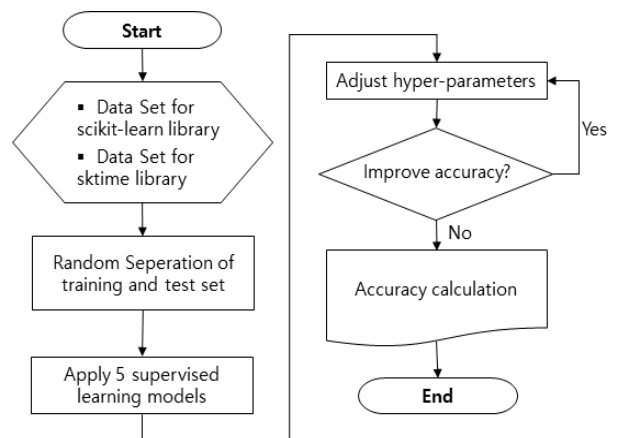


Fig. 6 Model Learning Process

6.2 합성곱 신경망과 LSTM을 이용한 학습

신경망 모델을 위한 입력 데이터의 형태는 정수형이어야 하므로 먼저, 실수형을 정수형으로 변환하는 작업을 수행했다. 모든 데이터 값 x 에 대하여 음의 실수 값은 제곱한 후 해당 값에 제곱근을 적용하였고, 양의 실수 값은 소숫점 이하 자리를 버림하여 정수형으로 바꿨다(Equation (2) 참조).

$$x = \begin{cases} |x| & \text{if } x \geq 0 \\ \sqrt{x^2} & \text{if } x < 0 \end{cases} \quad (2)$$

, where $\forall x \in D$ and $D \subset$ Ambulation data set of Z axis

둘째, 전체 데이터 셋에서 무작위로 추출하여 80%는 학습 데이터 셋(train dataset)으로 20%는 검증 데이터 셋(validation dataset)으로 구분하였다. 세 번째 단계에서는 *Embedding, Convolution, Dropout, Max Pooling, LSTM, Dense* 이렇게 총 6개의 레이어를 쌓아서 모델을 구성하였다. 보행 데이터 전체를 사용함으로써 특징을 직접적으로 추출하지 못했다. 이를 보완하기 위해 Convolution 레이어를 추가하여 특징을 추출한 후 시계열 데이터에 대한 학습을 위해 LSTM을 사용했다. 합성곱 신경망 모델로 데이터의 특징을 추출한 후 LSTM으로 시계열 데이터의 입력 값을 순서대로 처리하여 데이터의 순서에 대한 내용을 학습에 반영했다. 전체 데이터를 짧게 자름으로써 짧은 문장에서 특징을 잘 추출하는 합성곱 신경망 모델이 정확도를 높이는데 이바지했다고 생각한다. [26]에서 제시한 모델을 기반으로 시계열 데이터에 맞게 변형하여 사용하였다.

모델 구성을 마친 후 손실 값의 변화를 살펴면서 5회 이상 검증 데이터 손실 값이 증가하면 과적합이 될 수 있으므로 학습을 조기 종료하도록 하였고, 학습을 거듭하면서 검증 데이터의 정확도가 가장 높았던 순간을 저장하면서 학습을 진행하였다. 마지막으로, 학습의 정확도를 계산함으로써 모든 과정을 마무리했다. 전체 학습 단계를 Fig. 7에 도식화하였다.

Fig. 8은 신경망 모델 학습을 위한 실제 파이썬 코드를 보였다. 아래 코드에서 *Dropout*과 *Max Pooling* 레이어는 모델의 과적합 방지를 위해 사용되었고, 기계학습의 결과는 주어진 데이터 셋이 노인 체험복을 착용했는지 여부를 판단하

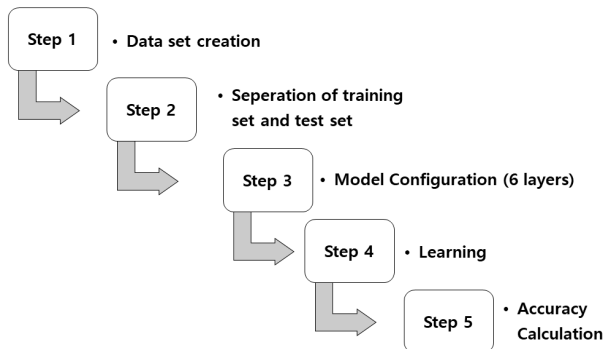


Fig. 7. Deep Learning Process

```

model = Sequential()
model.add(Embedding(500, 64))
model.add(Conv1D(filters=64, kernel_size=3, activation='relu'))
model.add(Conv1D(filters=64, kernel_size=3, activation='relu'))
model.add(Dropout(0.2))
model.add(MaxPooling1D(pool_size=2))
model.add(LSTM(64))
model.add(Dense(1, activation='sigmoid'))
  
```

Fig. 8. Python Code for Model Configuration

는 것이므로 *Dense* 레이어에서 *sigmoid* 활성화 함수를 사용하여 이진 분류를 진행했다.

6.3 기계학습 모델의 정확도 비교

scikit-learn과 sktime 라이브러리에서 선정한 5가지 모델의 정확도를 비교한 그래프를 Fig. 9에 나타냈다. 먼저, Fig. 9에 나타난 정확도 결과는 총 45명의 실험 참가자의 노인 체험복 착용 전후의 평지 걷기 데이터를 학습에 활용한 결과를 나타낸다. 5개의 모델 중 시계열 데이터 학습에 유용하다고 알려진 sktime 라이브러리의 TimeSeriesForestClassifier 모델에서 약 98%의 정확도를 보였다. 이것은 전체 데이터 셋 중 98%의 정확도로 노인보행과 비노인보행을 구별할 수 있다는 것을 의미한다. 다시 말해, 거의 모든 데이터 셋에 대해 노인 체험복 착용 여부를 구별할 수 있었다.

Figs. 10과 11은 신경망 모델의 학습 반복 횟수에 따른 정확도와 손실 값의 변화를 보여준다. 이 두 그래프에 따르면 학습이 반복됨에 따라 학습 데이터 셋과 검증 데이터 셋의 정확도는 1에 가까워짐을 알 수 있고, 손실 값은 점점 줄어서 0에 가까워지고 있다. 즉, 모델을 반복적으로 훈련시키기에 따라 이진 분류의 정확도가 계속적으로 상승하여 모델 학습이 종료된 후에는 노인 체험복 착용 여부를 정확하게 판단할 수 있게 되었다. 신경망 모델의 정확도는 약 99%를 나타냈다. 특히, Fig. 11의 손실 값 그래프는 신경망 모델의 최대 단점으로 인식되고 있는 과적합 문제가 발생하지 않았다는 것을 의미한다. 또한, 학습 반복 횟수 20 이상부터 정확도와 손실 값의 변화가 발생하지 않고 일정하게 유지된다. 이것은 과적합 문제가 발생하지 않고 모델이 안정적으로 학습되었음을 의미한다.

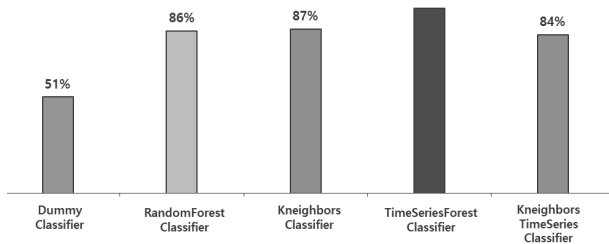


Fig. 9. Accuracy of 5 Machine Learning Models

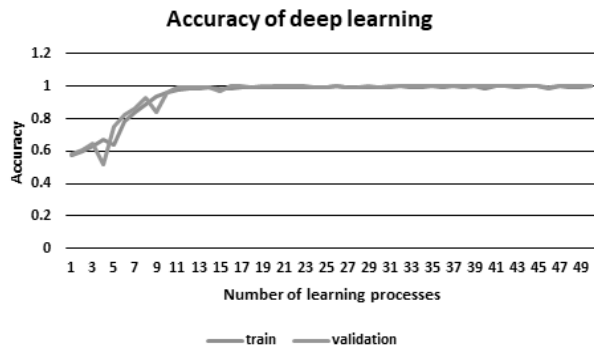


Fig. 10. Transition of Accuracy of RNN Model

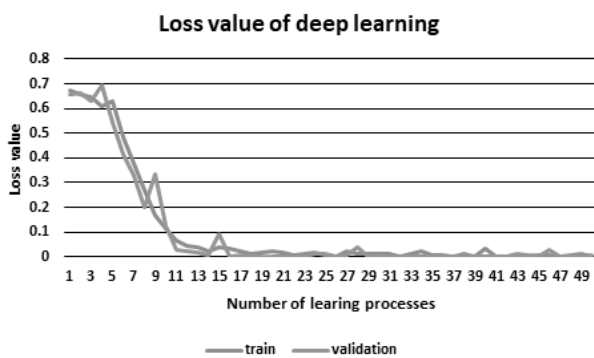


Fig. 11. Transition of Loss Value of RNN Model

7. 결 론

스마트폰에 내장된 3축 가속도 센서를 활용하여 보행 데이터를 수집하였고, 이렇게 수집된 보행 데이터에 대해 기계학습을 위한 전처리 과정을 거친 후 6가지 기계학습 모델을 활용하여 학습함으로써 노인보행과 비노인보행을 구별할 수 있는지에 관한 연구를 진행했다. 시계열 데이터 분석에 강점이 있는 TimeSeriesForestClassifier 모델이 판단한 결과의 정확도는 약 98%, 신경망 모델의 정확도는 약 99%로써 매우 정확하게 노인보행과 비노인보행을 구별할 수 있었다. 기존의 보행 분석은 관련 연구에서 언급한 바와 같이 여러 복잡한 단계를 거쳐 인간의 보행을 분석한 반면 본 연구에서는 복잡한 단계를 과감히 생략하고 수집된 원시 보행 데이터에 나타난 특징을 그대로 유지한 채로 기계학습을 진행했고 높은 정확도의 결과를 얻을 수 있었다. 따라서 기계학습 모델을 활용한 새로운 보행 분석 방법을 모색하는 계기가 되었다.

본 연구가 갖는 의의는 노인보행과 비노인보행을 구별하는 것이 가능하다면 근골격계 및 심혈관계 질환, 치매, 알츠하이머 등의 질병을 앓고 있는 노인들과 이러한 질환이 없는 노인들의 보행 또한 구별이 가능할 것이고, 더 나아가 지속적인 보행 데이터의 수집 및 분석을 통해 위에 언급된 질병의 가능성을 미리 예측하는 시스템을 개발할 수 있을 것으로 기대한다.

References

- [1] J.-H. Park, "Characteristics of gait in the elderly: Normal vs. abnormal," *Journal of the Korean Neurological Association*, Vol.35, No.4, pp.1-4, 2017.
- [2] D. K. Lee, "Gait Disorders," *Korean Journal of Clinical Geriatrics*, Vol.12, No.4, pp.141-148, 2011.
- [3] G. S. Heo, S. H. Yang, S. R. Lee, J. G. Lee, and C.-Y. Lee, "A study on particular abnormal gait using accelerometer and gyro sensor," *Journal of the Korean Society for Precision Engineering*, Vol.29 No.11, pp.1199-1206, 2012.
- [4] R. M. Ardle, S. D. Din, B. Galna, A. Thomas, and L. Rochester, "Differentiating dementia disease subtypes with gait analysis: Feasibility of wearable sensors?," *Gait & Posture*, Vol.76, pp.372-376, 2020.
- [5] L. G. M. Ader, B. R. Greene, K. McManus, and B. Caulfield, "Reliability of inertial sensor based spatiotemporal gait parameters for short walking bouts in community dwelling older adults," *Gait & Posture*, Vol.85, pp.1-6, 2021.
- [6] T. Steinmetzer, S. Wilberg, I. Bönninger, and C. M. Travieso, "Analyzing gait symmetry with automatically synchronized wearable sensors in daily life," *Microprocessors and Microsystems*, Vol.77, pp.1-10, 2020.
- [7] S. Clemens, K. J. Kim, R. Gailey, N. Kirk-Sanchez, A. Kristal, and I. Gaunaud, "Inertial sensor-based measures of gait symmetry and repeatability in people with unilateral lower limb amputation," *Clinical Biomechanics*, Vol.72, pp.102-107, 2020.
- [8] S. Mekruksavanich and A. Jitpattanakul, "Classification of gait pattern with wearable sensing data," in *Proceedings of the 4th International Conference on Digital Arts, Media and Technology*, Nan, Thailand, pp.137-141, 2019.
- [9] A. Tanigawa, S. Morino, T. Aoyama, and M. Takahashi, "Gait analysis of pregnant patients with lumbopelvic pain using inertial sensor," *Gait & Posture*, Vol.65, pp.176-181, 2018.
- [10] J. Li, Z. Wang, X. Shi, S. Qiu, H. Zhao, and M. Guo, "Quantitative analysis of abnormal and normal gait based on inertial sensors," in *Proceedings of the 22nd IEEE International Conference on Computer Supported Cooperative Work in Design*, Nanjing, China, 2018.
- [11] S. S. Fathima and W. Banu, "Abnormal walk identification for systems using gait patterns," *Biomedical Research*, pp.112-117, 2016.
- [12] K.-W. Park, "Gait Disturbances in Elderly Life," *Journal of the Korean Neurological Association*, Vol.35, No.4, pp.10-15, 2017.
- [13] X. Wang, D. Ristic-Durrant, M. Spranger, and A. Gräser, "Gait assessment system based on novel gait variability measures," in *Proceedings of the 15th IEEE International Conference on Rehabilitation Robotics (ICORR)*, London, UK, pp.467-472, 2017.

- [14] C. P. Burgos, L. Gärtner, M. A. G. Ballester, J. Noailly, F. Stöcker, M. Schönfelder, T. Adams, and S. Tassani, "In-ear accelerometer-based sensor for gait classification," *IEEE Sensors Journal*, Vol.20, No.21 pp.12895-12902, 2020.
- [15] R. LeMoyné and T. Mastroianni, "Network centric therapy for machine learning classification of hemiplegic gait through conformal wearable and wireless inertial sensors," in *Proceedings of the 8th IEEE International Conference on E-Health and Bioengineering (EHB)*, 2020.
- [16] R. LeMoyné and T. Mastroianni, "Conformal wearable and wireless inertial sensor system for machine learning classification of hemiplegic reduced arm swing," in *Proceedings of the 8th IEEE International Conference on E-Health and Bioengineering (EHB)*, 2020.
- [17] H. Zhang, Y. Guo, and D. Zanotto, "Accurate ambulatory gait analysis in walking and running using machine learning models," *IEEE Transaction on Neural Systems and Rehabilitation Engineering*, Vol.28, No.1, pp.191-202, 2020.
- [18] J. C. Pérez-Ibarra, A. A. G. Siqueira, and H. I. Krebs, "Identification of gait events in healthy and parkinson's disease subjects using inertial sensors: A supervised learning approach," *IEEE Sensors Journal*, Vol.20, No.24, pp.14984-14993, 2020.
- [19] B. Shi, S. C. Yen, A. Tay, D. M.L. Tan, N. S. Y. Chia, and W. L. Au, "Convolutional neural network for freezing of gait detection leveraging the continuous wavelet transform on lower extremities wearable sensors data," in *Proceedings of the 42nd IEEE Conference of Engineering in Medicine and Biology Society (EMBC)*, pp.5410-5415, 2020.
- [20] S. Strada, J. Paris, F. Piccoli, D. P. Tucci, P. Casali, and S. Savaresi, "Machine learning recognition of gait identity via shoe embedded accelerometer," in *Proceedings of the International Conference on Internet of Things (iThings)*, pp.852-857, 2020.
- [21] M. Nagashima, S. G. Cho, M. Ding, G. A. G. Ricardez, J. Takamatsu, and T. Ogasawara, "Prediction of plantar forces during gait using wearable sensors and deep neural networks," in *Proceedings of the 41st IEEE Conference of the Engineering in Medicine and Biology Society (EMBC)*, pp.3629-3632, Berlin, Germany, 2019.
- [22] A. Rubio-Solis, G. Panoutsos, C. Beltran-Perez, and U. Martinez-Hernandez, "A multilayer interval type-2 fuzzy extreme learning machine for the recognition of walking activities and gait events using wearable sensors," *Neurocomputing*, Vol.389, pp.42-55, 2019.
- [23] S. Majumder, T. Mondal, and M. J. Deen, "A simple, low-cost and efficient gait analyzer for wearable healthcare applications," *IEEE Sensors Journal*, Vol.19, No.6, pp.2320-2329, 2019.
- [24] U. Martinez-Hernandez and A. A. Dehghani-Sanij, "Adaptive bayesian inference system for recognition of walking activities and prediction of gait events using wearable sensors," *Neural Networks*, Vol.102, pp.107-119, 2018.
- [25] J. Hannink, T. Kautz, C. F. Pasluosta, J. Barth, S. Schülein, K-G. Gßbmann, J. Klucken, and B. M. Eskofier, "Mobile stride length estimation with deep convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, Vol.22, No.2, pp.354-362, 2018.
- [26] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.39, No.11, pp.2298-2304, 2017.



김 가 영

<https://orcid.org/0000-0003-2824-7163>
 e-mail : xdfc1745@gmail.com
 2021년 강릉원주대학교 컴퓨터공학과 (학사)
 2021년~현 재 인텔리언스 개발팀 AI 연구원

관심분야 : 머신러닝, 딥러닝



정 수 환

<https://orcid.org/0000-0002-3396-2455>
 e-mail : wjdtngkhsdle@naver.com
 2021년 강릉원주대학교 컴퓨터공학과 (학사)
 2021년~현 재 영원아이앤에스 인프라사업부 대리

관심분야 : 머신러닝, 딥러닝



엄 수 현

<https://orcid.org/0000-0002-9463-0198>
 e-mail : eomsh0529@naver.com
 2021년 강릉원주대학교 컴퓨터공학과 (학사)
 2021년~현 재 영원아이앤에스 인프라사업부 사원

관심분야 : 사물인터넷, 빅데이터 컴퓨팅



장 성 원

<https://orcid.org/0000-0001-6439-2689>
e-mail : eliotjang2@gmail.com
2021년 강릉원주대학교 컴퓨터공학과
(학사)
2021년~현 재 육군 제9585부대 2대대
통신과 소대장

관심분야: 컴퓨터네트워크, 사물인터넷



최 상 일

<https://orcid.org/0000-0002-9272-7367>
e-mail : schoi@gwnu.ac.kr
2000년 강릉대학교 컴퓨터공학과(학사)
2008년 아이오와주립대학교 컴퓨터과학과
(석사)
2015년 네브래스카주립대학교 컴퓨터공학과
(박사)

2015년 ~ 2016년 스와질랜드기독대학교 컴퓨터공학과 교수
2016년 ~ 2019년 아주대학교 소프트웨어학과 강의전담교수
2019년 ~ 현 재 강릉원주대학교 컴퓨터공학과 교수
관심분야: 사물인터넷, 머신러닝 및 딥러닝, 보행분석



이 소 연

<https://orcid.org/0000-0002-9146-688X>
e-mail : freeyeon96@gmail.com
2021년 강릉원주대학교 컴퓨터공학과
(학사)
2021년~현 재 피지알디 개발팀 매니저
관심분야: 모바일 및 웹 프로그래밍