

Mediapipe와 LSTM을 활용한 Skeleton Keypoints 기반 현장 사고 감지 시스템

Accident Detection System Based on Skeleton Keypoints using Mediapipe and LSTM

요 약

사고 상황에서의 빠른 조치는 인적, 물적 손실을 최소화하는 가장 효과적인 방법이다. 그러나 관리 감독이 부족한 작업환경에서는 빠르게 사고를 감지하고 조치하기 어렵다. 이러한 문제를 해결하기 위해 인공지능을 활용한 낙상 감지 시스템이 개발되었다. 그러나 작업 현장의 특성상 낙상 감지만으로는 모든 사고를 감지하기에 부족하다. 따라서, 본 논문에서는 낙상 뿐만 아니라 기절 등 분류 기준을 추가한 사고 감지 시스템을 제안한다. 본 연구에서는 Mediapipe 라이브러리와 장단기 메모리(Long Short Term Memory: LSTM) 기반 순환 신경망(Recurrent Neural Network: RNN)을 사용하여 사고 발생 여부를 영상을 통해 판단할 수 있는 인공지능 모델을 구현하였다. 자체 실험 결과, 구현 모델이 낙상상태 뿐만이 아닌 기절상태 또한 사고로 감지하면서도 우수한 정확도를 보임을 확인하였다.

keyword: LSTM, RNN, 인공지능, 사고 감지

Abstract

Quick reaction in accidents situation is the most effective way to minimize human and material losses. But, it is difficult to detect and react accident rapidly at the workplaces with insufficient management supervision. To alleviate this problem, fall detection systems using AI(Artificial Intelligence) has been researched. However, on characteristic of workplaces, it is insufficient to detect all accidents by using fall detection only. Thus, in this paper, we suggest accident detection system to detect not only fall state but also stun state as accidents. In this work, we implemented an AI model detecting accidents with video using Mediapipe and LSTM. Our experiment confirmed that our proposed model shows high accuracy although our implemented model detect fall state as well as stun state as accidents.

keyword: LSTM, RNN, AI, accident detection

1. 서 론

현대 사회에서는 경제, 문화, 공업 등의 다양한 분야에
서 산업 발달이 이루어지고 있다. 산업현장에서의 산업
재해는 인명피해를 초래하는 심각한 문제이다.

2020년 고용노동부의 조사에 따르면 사업체 내 근로자
수를 기준으로 5인 미만에서 전체 사고의 35.4%(312명),
5~49인에서 45.6%(402명)의 사고사망자가 발생한 반면, 5
0~299인은 14.9%(131명) 300인 이상은 4.2%(37명)가 발생
하였다 [1]. 즉 소규모 사업장에서의 사고 사망 비율이
대규모 사업장 대비 높음을 알 수 있다. 이는 상대적으로
로 인원이 적은 사업장일수록 각 작업환경을 적절하게

관리 감독하기 어렵고, 사고 발생 시 빠른 조치를 기대
하기 어렵기 때문이다. 이를 해결하기 위해 우리나라의
건설 산업 현장의 경우 국토교통부 주관으로 안전수준
평가를 시행하고 있지만, 그 주기가 길거나 점검인력이
부족하여 실효성이 떨어진다는 [2].

이처럼 안전 관리가 부족한 현장에서 발생한 사고를
신속하게 탐지하기 위해 본 연구에서는 인공신경망(Arti-
ficial Neural Network: ANN) 기반의 사고 감지 시스템을
구현하였다. 인공신경망이란 인간의 뇌 속 뉴런을 모방
한 학습 모델로, 인공지능은 학습한 내용을 바탕으로 현
재 상황의 위험도를 예측하는 기능을 수행해 왔다 [3].

사고 상태를 감지하기 위한 기존 기술로 낙상 감지(fall

detection) 기술이 있다. 낙상 감지 시스템 중 뛰어난 감지 성능을 보이는 기술로는 카메라를 통해 취득한 영상 데이터를 인공신경망 구조의 일종인 장단기 메모리(Long Short Term Memory: LSTM) 기반 순환 신경망(Recurrent Neural Network: RNN)을 사용한 인공지능을 통해 낙상을 감지하는 기술이 있다 [4]. 그러나, 작업 현장에서 사고를 당한 사람의 자세를 넘어짐(fall) 하나만으로 감지하기엔 부족함이 있어 개선이 필요하다.

본 논문에서는 감시카메라 영상에서 얻어진 데이터에서 Mediapipe [5]를 사용하여 키포인트와 키포인트의 좌표를 추출하고, 추출된 키포인트의 좌표값을 통해 선 자세(stand), 앉은 자세(sit), 숙인 자세(fold), 기절한 자세(s-tun), 넘어진 자세(fall)의 5가지 자세로 분류(classification)하여 사고를 감지하고 알림을 전송하는 LSTM 기반 인공지능 모델을 제안한다.

2. 배경 지식

2.1. 인공신경망(Artificial Neural Network)

본 절에서는 인공신경망과 연구에서 사용한 인공신경망 모델들에 대하여 설명한다. 인공신경망은 인간의 뇌속 뉴런의 동작 원리를 모방하여 구축한 학습 알고리즘으로 입력층, 은닉층, 출력층의 3계층으로 구성된다 [6]. 입력층은 초기에 데이터가 입력되는 계층, 은닉층은 데이터가 드러나지 않는 계층, 출력층은 학습된 데이터가 담긴 계층을 의미한다. 각각의 계층들은 여러 개의 퍼셉트론으로 구성되어 있으며, 각 퍼셉트론들은 효과적인 출력값을 만들기 위해 활성화 함수(activation function)를 사용한다 [6].

활성화 함수는 입력된 데이터의 가중 합을 출력 신호로 변환하는 함수로 시그모이드(sigmoid)나 하이퍼볼릭 탄젠트(hyperbolic tangent: tanh), 소프트맥스(softmax) 등의 함수를 사용하며 본 논문에서는 활성화 함수로 하이퍼볼릭 탄젠트와 소프트맥스 함수를 사용하였다.

2.2. 장단기 메모리 (Long Short Term Memory: LSTM)

본 논문에서는 인공신경망 모델의 일종인 LSTM 기반 RNN에 대하여 설명한다.

그림 1은 RNN의 구조도이다. RNN이란 은닉층의 퍼셉트론에서 나온 결과를 출력층 방향으로 보낼과 동시에 은닉층의 다음 입력으로 보내 순환하며 학습하는 신경망이다 [7]. RNN은 영상과 같이 시간의 흐름에 따라 변화하는 데이터를 효과적으로 학습할 수 있는 인공신경망 모델이다. 그러나, 장시간 학습을 진행하는 경우 학습 초

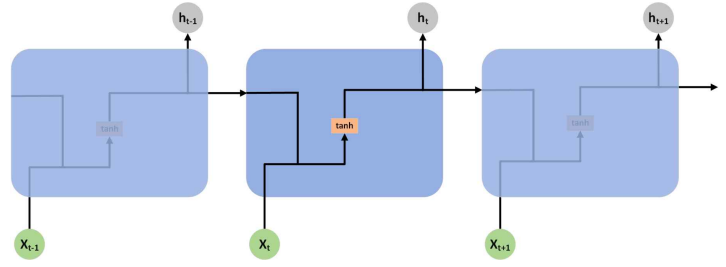


그림 1 RNN 구조도

Fig. 1 RNN structure diagram

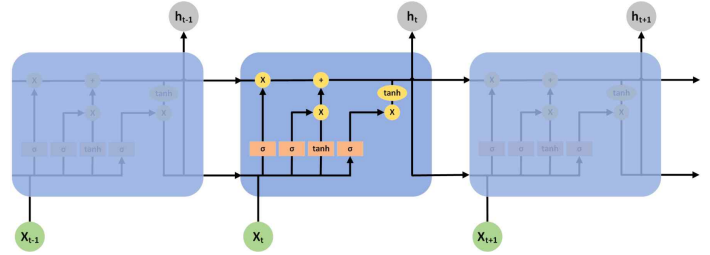


그림 2 LSTM 구조도

Fig. 2 LSTM structure diagram

기의 결과가 미치는 영향이 미미해져 기울기 소실 문제가 발생하여 학습 능력이 크게 저하되는 문제가 있다 [8]. 이러한 문제를 해결하고자 개발된 퍼셉트론 구성 방법이 LSTM이다.

그림 2는 LSTM 기반 RNN의 구조도이다. LSTM은 장기 기억 메모리 역할을 수행하는 셀(Cell) 상태와 연결의 강도를 조절하는 망각, 입력, 출력 게이트로 구성되어 있다. LSTM은 망각 게이트를 통해 이전 신경망 정보가 현재 신경망에 끼치는 영향을 결정하고, 이를 통해 RNN이 가지고 있던 기울기 소실 문제를 개선할 수 있다 [7].

3. 관련 연구

영상 기반 행동 인식 시스템은 영상에서 특징을 검출하여 행동을 인식한다. 검출 방식은 크게 두 가지로, 영상에서 얻을 수 있는 사람의 외형이나 전체적인 윤곽 정보를 이용하는 방식과 사람의 관절과 같은 키포인트들을 이용하는 방식이 있다 [4]. 키포인트들을 이용하는 방식의 경우 인체 관절의 위치 정보를 포함하여 행동 인식에 도움이 되는 주요 특징을 얻기 쉬운 장점이 있다.

[9]에 따르면, 사람의 관절과 같은 키포인트를 추출하는 기법으로는 OpenPose와 BlazePose가 있다. OpenPose는 사람의 관절마다 히트 맵을 생성해서 각 관절 좌표에 대해 조정하는 방식으로, 사람의 관절 키포인트를 추출하여 이용한다. 여러 사람의 경우 최소한의 비용으로 키포인트를 추출하는 것이 가능하나, 한 사람의 경우 관절 키포인트를 추출하여 이용하기엔 모델이 너무나 과도하여 실시간 처리에는 부적합하다 [9]. 반면, Mediapipe 라

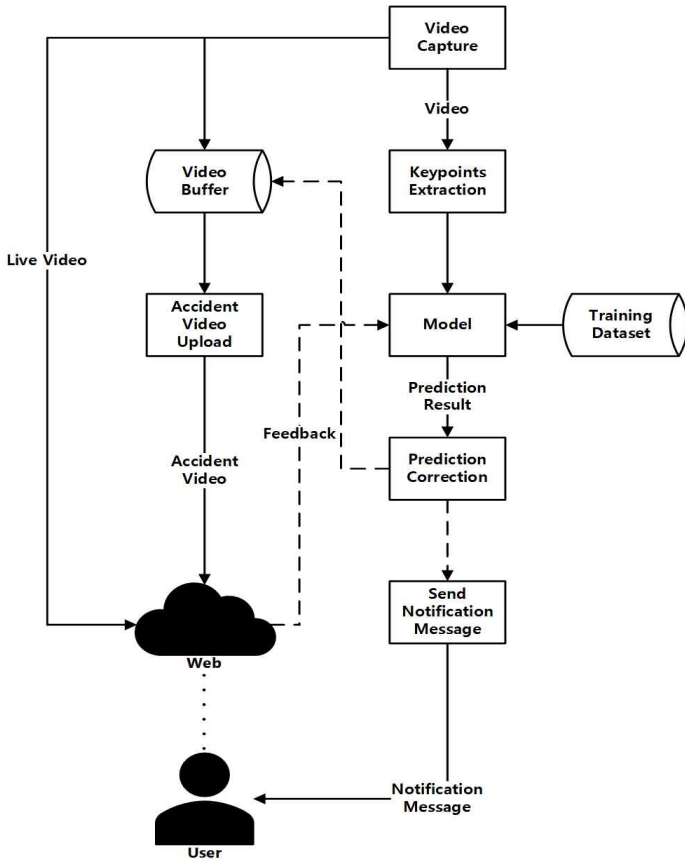


그림 3 전체 시스템 구조도
Fig. 3 Overall system structure diagram

이브러리에서 사용하는 BlazePose의 경우 한 사람일 경우에도 최적화 과정을 거쳐 가벼운 모델인 합성곱 신경망(Convolutional Neural Network: CNN)을 통해 추출하므로 실시간 처리에 적합하다 [10]. 따라서, 본 연구에서는 BlazePose 기반의 Mediapipe를 활용하여 키포인트를 추출하였다.

[11]에서는 영상에서 OpenPose를 이용하여 사람의 키포인트를 추출하고 CNN과 SVM(Support Vector Machine)을 사용하여 낙상을 검출하는 시스템을 제안하였다.

[12]에서는 일반적인 RNN에서 발생하는 기울기 소실 문제를 해결하기 위한 LSTM, 게이트 순환 구조(Gate Recurrent Unit; GRU)를 적용한 낙상 인식 시스템을 제안하였다. [12]에서 2D 카메라를 사용하여 키포인트를 추출하고 기존 RNN 모델과 LSTM, GRU 기반의 모델 사용시의 낙상 인식 정확도를 비교한 결과, LSTM을 사용하였을 때의 정확도가 가장 우수함을 보였다.

기존 연구들의 낙상 감지 시스템은 사람의 자세를 단순히 정상 상태와 낙상 상태의 두 종류로만 나누어 판단하는 문제가 있다. 따라서 본 논문에서는 실제 산업현장에서 감시카메라로 주로 이용하는 2D 카메라와 Mediapipe를 이용하여 사람의 키포인트를 추출하였고, 관련

연구들의 실험 결과 우수한 성능을 보인 LSTM 기반 RNN 모델을 사용하여 사람의 자세를 서 있는 자세, 앉은 자세, 숙인 자세, 기절한 자세, 넘어진 자세의 5가지로 분류하여 사고를 감지하는 시스템을 구현하였다.

4. LSTM 기반 사고 감지 시스템의 구현

행동 인식 시스템을 활용하여 사고를 감지하고 조기에 대응할 수 있도록 인공지능 모델을 사용한 사고 감지 시스템을 구현하였다. 전체 시스템 구조는 그림 3과 같은 구조로 되어 있으며 데이터 수집 단계, 사고 감지 단계, 사고 알림 단계의 3단계로 되어 있다.

4.1 데이터 수집

4.1.1 영상 획득

영상 입력 장치를 통해 원본 영상을 획득하는 단계이다. 사고 감지 정확도를 높이기 위해선 학습에 사용한 영상과 비슷한 각도로 신체 전부가 장애물 없이 영상 내에 존재하도록 하는 것이 유리하다. 키포인트를 추출하기 위한 영상 입력 장치로는 3D 깊이 카메라 혹은 2D 카메라를 사용한다. 그러나 3D 깊이 카메라의 경우 2D 카메라보다 가시거리가 짧고 화각이 좁으며 실제 산업현장에서 감시카메라로 사용하는 카메라 또한 주로 2D 카메라를 사용하므로 구현 모델에서는 2D 카메라를 통해 영상 데이터를 획득하고 낙상 감지에 이용하였다 [4].

4.1.2 키포인트 추출

Mediapipe에서 추출하는 키포인트는 총 33개이다. 이중 11개의 키포인트가 얼굴에 존재하며 이는 사람의 자세를 인식하는 본 모델에는 적합하지 않은 키포인트 분포라고 판단하였다. 부적합한 키포인트 제거를 통한 정확도 향상을 위해 키포인트 사용 여부에 따라 실험군을 나누어 실험하였다. 사용 데이터셋은 오버샘플링을 적용하기 전의 자체 제작 데이터셋+외부 데이터셋을 사용한 모델에 RP정규화를 수행하고 3회 진행하여 평균 정확도를 계산하였다. 그 결과는 표 1과 같이 모든 키포인트를

표 1 키포인트 사용 별 정확도
Table 1 Accuracy by keypoint usage

Keypoint usage type	Average Acc
All keypoints	82.67%
Remove some keypoints on the face	83.81%
Remove keypoints on the arm	85.59%
Remove some keypoints on the face and arm	87.53%
Remove some keypoints on the face, arm and foot	85.63%

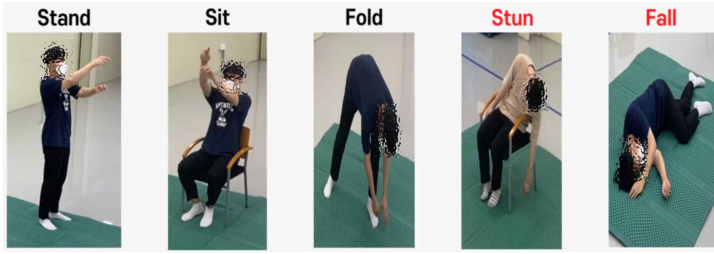


그림 4 자세 클래스
Fig. 4 Posture class

사용하였을 때 82.67%, 얼굴 일부 제거 83.81%, 팔 일부 제거 85.59%의 평균 정확도를 보여 모든 키포인트를 사용한 경우에 비해 성능 향상을 확인할 수 있었다.

성능 향상을 보였던 두 경우를 조합하여 사용하면 정확도가 더욱 올라갈 것으로 보고 얼굴 일부와 팔의 키포인트를 제거한 경우와 얼굴 일부와 팔에 더하여 발 일부 키포인트까지 제거한 경우의 실험을 수행하였다. 그 결과 얼굴 일부와 팔의 키포인트를 제거하였을 때 87.53%의 평균 정확도를 보여 확실한 성능 증가세를 보였지만, 발 일부의 키포인트를 추가로 제거하면 85.63%로 얼굴 일부와 팔의 키포인트를 제거하였을 때 비해 오히려 낮아지는 결과를 보였다. 실험 결과 가장 높은 정확도를 보였던 경우를 반영하여 얼굴 일부와 팔의 키포인트를 제거하고 16개의 키포인트를 사용하도록 구현 모델에 적용하였다.

데이터가 가진 키포인트들에 대하여 각 키포인트의 위치 좌표값의 크기가 크게 차이 나는 경우 각 키포인트 별로 모델의 가중치 설정에 반영되는 정도가 달라진다. 이 경우 결과에 영향을 미쳐 영상 내 인물의 위치에 따라 같은 자세임에도 다른 좌표값을 가지게 되는 문제가 생긴다. 이와 같은 문제를 해결하기 위해 각 좌표값의 크기를 조정하여 결과에 미치는 영향을 균일하게 만드는 기법인 정규화를 수행하였다.

정규화 기법은 인물을 기준으로 키포인트들의 좌표를 식 1의 공식을 적용해 0 ~ 1의 값을 가지도록 설정해주는 기법인 최소-최대 정규화를 사용하였다 [13].

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad \dots\dots\dots(1)$$

4.2 사고 감지

4.2.1 인공지능 분류

정규화를 마친 데이터는 사고 감지를 위해 학습된 인공지능을 사용하여 그림 4와 같이 서 있는 자세(Stand), 앉은 자세(Sit), 숙인 자세(Fold), 기절한 자세(Stun), 넘어진 자세(Fall)의 5가지 클래스로 분류한다. 기절한 자세와 넘어진 자세로 분류되면 사고로 판단하고, 감지한 사고를 알리기 위한 알림 단계로 진행한다. 이때, 잘못된 키포인트 추출 등의 오류로 인한 판단 착오를

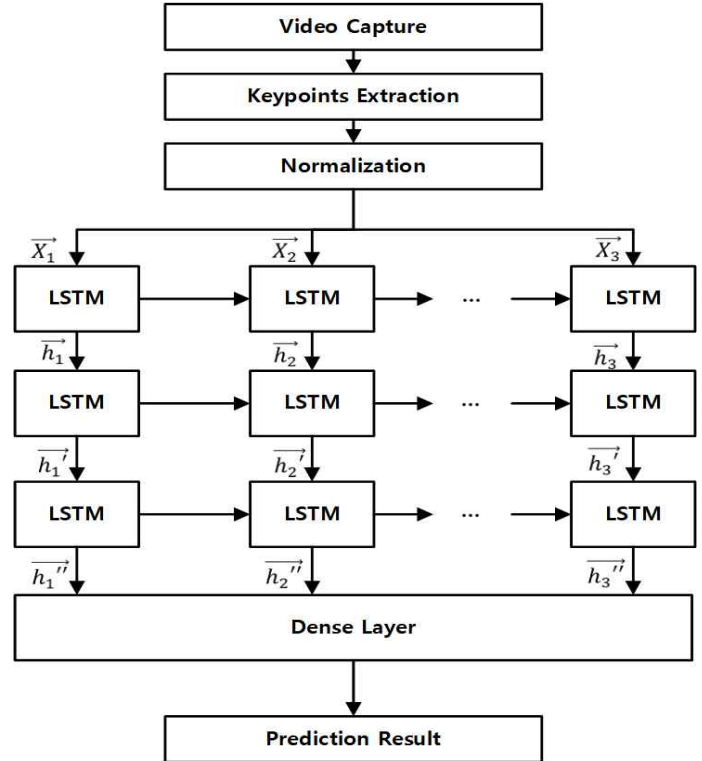


그림 5 학습 모델 구조도
Fig. 5 Training model structure diagram

줄이기 위하여 약 1.5초 이상 연속으로 사고 상황으로 판단하는 경우만 사고 알림 단계로 진행하도록 구현하였다.

4.2.2 인공지능 모델 학습

인공지능 모델은 입력층, 은닉층, 출력층으로 이루어져 있으며 그림 5와 같은 형태로 학습이 이루어지게 된다. 우선 정규화를 마친 데이터셋에서 1프레임당 16개의 키포인트를 추출하고, 추출된 키포인트의 X 좌표와 Y 좌표를 32차원 벡터 형태의 데이터로 만든다. 사람의 동작 흐름을 학습에 반영할 수 있도록 30프레임에 해당하는 데이터를 학습 단위로 사용하여 30개의 32차원 벡터 \vec{X}_n ($n=1\sim30$)를 하나의 학습 데이터로 입력층에 입력시켜 학습을 수행하게 된다. 입력층은 32차원 벡터 \vec{X}_n 을 입력받는 30개의 LSTM 셀로 이루어져 있다. 은닉층은 활성화 함수로 하이퍼볼릭 탄젠트(tanh)를 사용하며 출력 \vec{h}_m ($m=1\sim30$)를 발생시키는 LSTM 셀이 30개씩 3층으로 완전 연결(fully connected) 되어진 완전 연결 계층과 Dense 층으로 이루어져 있다. 출력층은 활성화 함수로 소프트맥스 함수를 사용한 dense 층으로 이루어져 있으며 분류하고자 하는 자세의 수인 5개의 퍼셉트론을 가지고 있다. 학습에 사용한 최적화 함수로는 Adam을 사용하였으며 배치 크기는 96, 학습 횟수(epoch)는 90이다.

4.3 사고 알림

인공지능 분류를 통해 사고를 감지하면 메신저 앱의 notification API를 활용하여 관리자에게 알림과 웹앱 링크를 전송하고, 관리자는 웹을 통해 현장의 실시간 영상과 사고 발생 당시의 3초간 영상을 확인하여 조치할 수 있도록 한다. 사고로 판단한 영상 중 실제 사고가 아닌 영상에 대해서는 관리자가 피드백을 제공하고, 피드백은 모델의 학습에 반영되어 매일 자정(00:00)에 새로운 환경의 데이터로 학습을 수행하도록 구현하였다.

5. 실험 결과 및 평가

5.1 실험 환경

본 논문에서 구현한 모델의 실험 환경은 표 2와 같다. 운영체제는 Windows 10, CPU는 Intel(R) Core(TM) i7-6700, GPU는 Nvidia Geforce GTX 1050, 램은 DDR3 24GB 램을 사용하였으며 저장 장치로는 하드 디스크를 사용하였다. 사용 언어는 Python 3.8 버전을 사용하였고 라이브러리는 TensorFlow 2.4.1 버전과 Keras 2.4.3 버전을 사용하였다. 사용 인공지능망으로는 LSTM 기반 RNN모델을 사용하였으며 배치(batch) 사이즈는 96, 학습 횟수(epoch)는 90이다. 활성화 함수는 은닉층에 하이퍼볼릭 탄젠트, 출력층에 소프트맥스 함수를 사용하였으며 최적화 함수로는 Adam을 사용하였다.

5.2 사용 데이터셋

학습 및 검증을 위한 실험에 사용한 데이터셋은 표 3과 같다. 자체 제작 데이터 193,110프레임과 JEI(Journal of Electronic Imaging)에서 발췌한 외부 데이터 59,760프레임[14]을 합하여 총 252,870프레임의 데이터를 사용하였고, 이때 자세별 프레임과

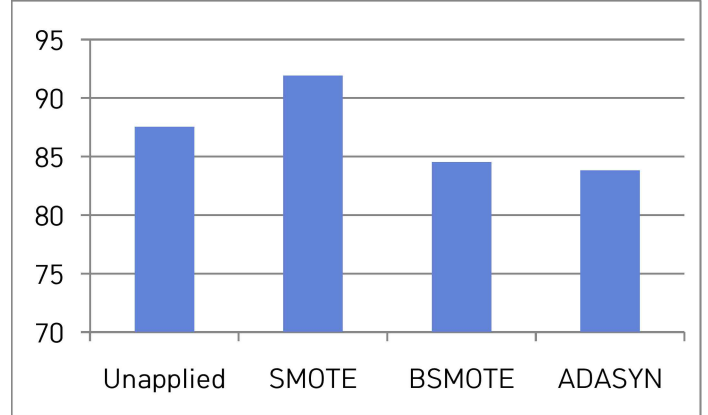


그림 6 오버샘플링 기법별 정확도

Fig. 6 Accuracy by oversampling method

비율은 stand 90,720프레임(35.9%), sit 50,040프레임(19.8%), fold 21,600프레임(8.5%), stun 25,830프레임(10.2%), fall 64,680프레임(25.6%)이다.

정확도에 영향을 미치는 데이터 불균형 문제 [15] 해결을 위해 다양한 오버샘플링 방법을 적용하여 실험하였다. 실험 결과는 5.3절에서 후술하며, 소수 클래스와 다수 클래스 사이에 소수 클래스에 해당하는 임의의 새로운 데이터들을 생성하는 SMOTE(Synthetic Minority Over-Sampling Technique) [15] 오버샘플링을 적용하였다. 오버샘플링 결과 클래스당 90,720프레임을 가진 5개 클래스, 총 450,450프레임의 데이터셋이 되었다.

5.3 실험 및 평가

그림 6은 오버샘플링 미적용, SMOTE, BSMOTE(Borderline SMOTE), ADASYN(ADaptive SYNthetic sampling) 네 가지 오버샘플링을 적용하여 실험하고 그 정확도를 비교한 차트이다. 실험 결과 오버샘플링 미적용(87.53%), SMOTE(91.90%), BSMOTE(84.52%), ADASYN(83.80%)의 평균 정확도를 보여주었다.

표 2 실험 환경

Table 2 Experiment enviroment

OS	Windows 10
CPU	Intel(R) Core(TM) i7-6700
GPU	Nvidia Geforce GTX 1050
RAM	DDR3 24GB
Storage	Hard Disk Drive
Batch size	96
Epoch	90
Neural Network	RNN based on LSTM
Active Function	hyperbolic tangent, softmax
Optimize Function	Adam
Programming language	Python 3.8
Library	TensorFlow 2.4.1 / Keras 2.4.3

표 3 사용 데이터셋

Table 3 Used dataset

	Data	Number of frame (ratio)
by type	internal data	193,110 frame (76.4%)
	external data(JEI)	59,760 frame (23.6%)
by posture	stand	90,720 frame (35.9%)
	sit	50,040 frame (19.8%)
	fold	21,600 frame (8.5%)
	stun	25,830 frame (10.2%)
	fall	64,680 frame (25.6%)

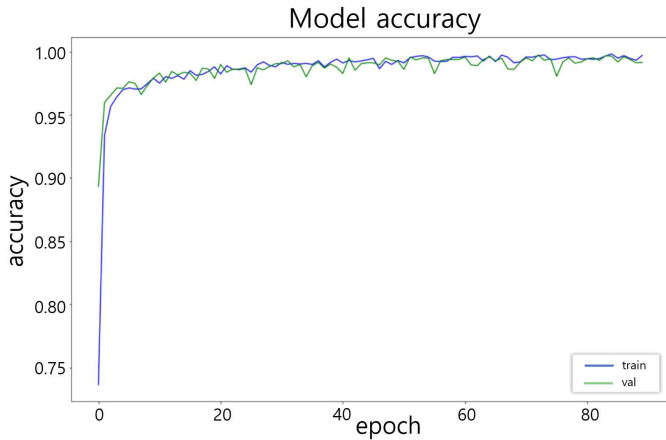


그림 7 학습 횟수별 정확도
Fig. 7 Accuracy by epoch

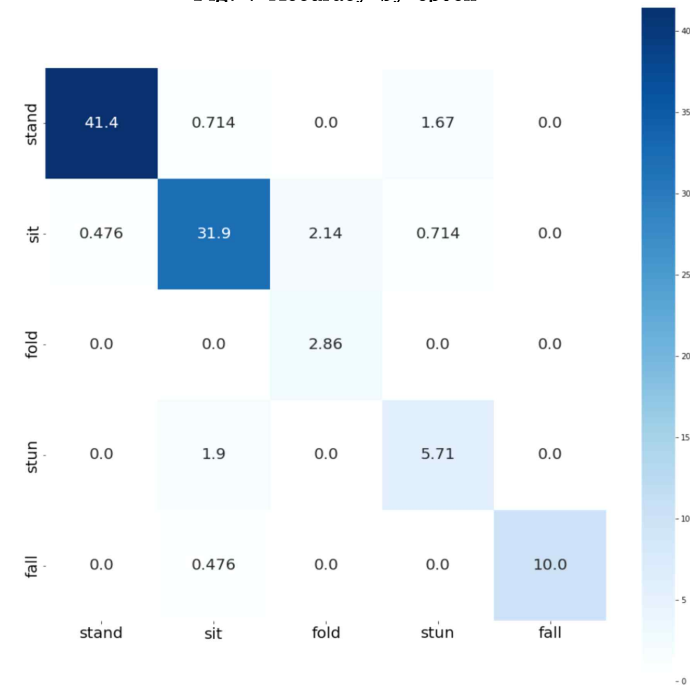


그림 8 랩 테스트 결과 혼동 행렬
Fig. 8 Confusion matrix from lab test result

그림 7는 제안 모델의 학습 횟수(epoch)에 따른 정확도의 변화를 나타낸 그래프이다. 0.75의 정확도로 시작하여 가파른 상승 곡선을 그리는 파란색 그래프는 학습 정확도를 나타내며 0.90에서 시작하는 초록색 그래프는 검증 정확도를 나타낸다. 두 그래프 모두 학습 횟수의 증가에 따라 안정적인 상승 개형을 보여 문제없이 학습이 이루어지고 있음을 보여주고 있다.

그림 8은 제안 모델의 랩 테스트 결과를 혼동 행렬로 나타내고 시각화한 것으로 세로축은 정답, 가로축은 예측 결과이다. 구현 모델의 예측과 정답 라벨이 정확히 일치하는 경우의 정확도는 약 91.90%이고 안전 상태(stand, sit, fold)와 위험 상태(stun, fall)로 분류하는 경우의 정확도는 95.20%로 우수한 정확도를 보여주었다.

6. 결론

본 논문에서는 LSTM 기반 인공지능 모델을 통해 사고를 감지하고 알람을 전송하는 시스템을 구현하였다. 감시카메라를 통해 입력된 작업 현장의 영상을 사전 학습된 인공지능 모델을 사용하여 자세를 분류하고, 사고 상황으로 판단되면 관리자에게 사고 상황이라는 알람을 보낸다. 실험을 통해, 구현된 시스템은 기존 낙상 감지 시스템보다 더욱 다양한 사고를 감지하면서도 약 95.20%의 우수한 정확도를 보임을 확인할 수 있었다. 관리 감독이 부족한 작업 현장에 적용하면 의식을 잃어 스스로 조치할 수 없는 사고 발생 시 빠른 초동 대처를 가능하게 하여 인명사고를 방지하는 것에 큰 도움이 될 것으로 생각한다.

참고 문헌

- [1] Ministry of Employment and Labor Industrial Accident Prevention Policy Division [Online]. Available: <https://ieic.kdi.re.kr/policy/materialView.do?num=212812&topic> (in Korean)
- [2] Occupational Safety and Health Research Institute [Online]. Available: <https://www.kosha.or.kr/oshri/publication/research/ReportSearch.do?mode=download&articleNo=419749&attachNo=237052> (in Korean)
- [3] Y. C. Kim, W. S. Yoo and Y. Shin, "Application of Artificial Neural Networks to Prediction of Construction Safety Accidents," Journal of The Korean Society of Hazard Mitigation, Vol. 17, No.1, pp. 7-14, 2017. (in Korean)
- [4] S. Jeong, S. Kang and I. Chun, "Human-skeleton based fall-detection method using LSTM for manufacturing industries," Proc. of 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), pp. 1-4, 2019.
- [5] Mediapipe, [Online]. Available: <https://mediapipe.dev>
- [6] A. K. Jain, J. Mao and K. M. Mohiuddin, "Artificial neural networks: A tutorial," Computer, Vol. 29, No. 3, pp. 31-44, 1996.
- [7] S. Alex, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," Physica D: Nonlinear Phenomena, Vol. 404, pp. 132306, 2020.
- [8] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 6, No. 3, pp. 107-116, 1998.
- [9] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang and M. Grundmann, "Blazepose: On-device real-time

- body pose tracking,” arXiv:2006.10204, 2020.
- [10] K. O’Shea and R. Nash, “An introduction to convolutional neural networks,” arXiv preprint arXiv:1511.08458, 2015.
 - [11] Z. Huang, Y. Liu, Y. Fanh and BKP. Horn, “Video-based fall detection for seniors with human pose estimation,” Proc of 2018 4th International Conference on Universal Village (UV), pp 1-4, 2018.
 - [12] C. B. Lin, Z. Dong, W. K. Kuan and Y. F. Huang, “A framework for fall detection based on openpose skeleton and lstm/gru models,” Applied Sciences 11.1, pp. 329, 2020.
 - [13] S. Patro and K. K. Sahu. “Normalization: A preprocessing stage,” arXiv preprint arXiv:1503.06462, 2015.
 - [14] I. Charfi, J. Mitian, J. Dubois, M. Atri and R. Tourki. “Optimised spatio-temporal descriptors for real-time fall detection: comparison of SVM and Adaboost based classification,” Journal of Electronic Imaging (JED), Vol. 22. No. 4, pp. 17, 2013.
 - [15] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kelmeier, “SMOTE: synthetic minority over-sampling technique,” Journal of artificial intelligence research, Vol. 16, pp. 321-357, 2002.