MediaPipe-based LSTM-Autoencoder Sarcopenia Anomaly Detection and Requirements for Improving Detection Accuracy

HyeRin Yoon[1], Eunah Jo[2], Seungjae Ryu[2], Jun-Il Yoo[3] and Jin Hyun Kim[1,*]

[1]*Gyeongsang National University, Jinju-si , 53828 , South Korea*
[2]*Gyeongsang National University, Jinju-si , 53828 , South Korea*
[3]*Gyeongsang National University Hospital, 52828 , South Korea*

Abstract

MediaPipe is a leaning-based human pose detection technology that detects the position and movement of a person's body, face, fingers, etc., from videos. Nowadays, many orthopedic studies put efforts into finding a biomarker of orthopedic diseases from the correlation between gait and orthopedic, using MediaPipe. This paper presents the results of applying the LSTM(Long Short Term Memory)-Autoencoder-based anomaly detection technique for orthopedic diseases, e.g., sarcopenia disease and the capability of distinguishing the normal and abnormal gait. We compare the sensitivity of the anomaly detection based on 5 human body points in predicting sarcopenia so as to find the primary gait features of human body. In addition, we present four environmental factors affecting MediaPipe Recognition that can improve the accuracy of anomaly detection using MediaPipe. Our anomaly detection approach detects 92% (35) of sarcopenia patients from 38 patients.

Keywords

AI, MediaPipe, LSTM-Autoencoder, Anomaly detection, Sarcopenia, Gait analysis, YOLO

1. Introduction

In recent years, AI(Artificial Intelligence) that can help doctors' decisions based on accumulated data is positioned as a research flow in the medical area[1][2]. In addition, AI is an effective diagnostic way because it can provide immediate diagnosis and severity analysis for patients at a low cost. Many orthopedic studies use gait images or video to find the status of musculoskeletal-related diseases[3][4]. Chen et al. presented a hybrid prediction model that combines an LSTM model and an SVM classifier to provide a functional description of the knee joint to patients with osteoarthritis, one of the musculoskeletal diseases[5]. Mirelman et al. proposed the RUSBoost(Random Under-Sampling Boosting) classification algorithm to examine the correlation between gait changes, one of the symptoms of Parkinson's disease, and the severity of Parkinson's disease[6]. The problems mentioned in such orthopedic papers are that the lack of quantitative gait analysis systems leads to doctors' subjective decisions or that the correlation between gait and disease is unclear.

Image tracking systems, such as marker-based VICON, are mainly used as data acquisition methods for developing quantitative gait analysis systems[7][8]. However, applying marker-based systems can only be used in limited environments and is time-consuming and expensive. Also, for unskilled experts revising and collecting the data is difficult. Various orthopedics research is trying to solve these problems using MediaPipe, a markerless-based gait tracking model[9]. MediaPipe can gait analysis using only video taken by smartphones that anyone can easily access and use without high-spec cameras or sensors. In addition, since MediaPipe extracts more joint points than OpenPose, mainly used for gait analysis, it is possible to develop a more accurate gait analysis system by using data to which MediaPipe is applied.

This paper figures out joint angles of normal and abnormal gaits with sarcopenia from the DMS (Data Management System) of DEEVO Co. Ltd with MediaPipe. Based on the joint angle data of pedestrians, we train the LSTM-Autoencoder model with normal gait data and propose anomaly detection method to detect abnormal gait. In this paper, abnormal gait data are from patients with sarcopenia, a representative musculoskeletal disease. Sarcopenia is a disease in which muscle mass and muscle strength are overly reduced with increasing age, resulting in abnormalities in physical function. In addition, the cause of sarcopenia and the body part where the symptoms appear is different among sarcopenia patients.

Therefore, we use the right and left knee, right and left hip, and nose-shoulder angle data for anomaly detection. This paper provides a sarcopenia detection technique using the LSTM-Autoencoder model and anomaly detection method and its results. In addition, this paper presents four environmental factors affecting MediaPipe recog-

1st International Workshop on Intelligent Software Engineering, December 06, 2022, Busan, South korea

dbsfls00@gnu.ac.kr (H. Yoon); chosd4603@gnu.ac.kr (E. Jo); ruru0213@gnu.ac.kr (S. Ryu); furim@daum.net (J. Yoo); jin.kim@gnu.ac.kr (J. H. Kim)

nition and their solutions: frame imbalance, headless, clothing, and background that affect recognition error when applying MediaPipe.

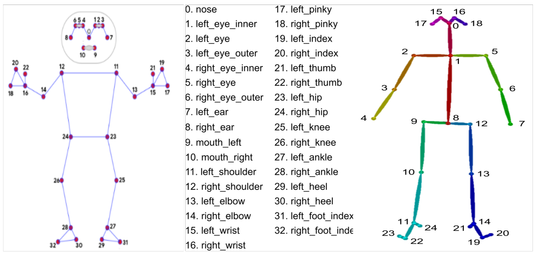
In this paper is organized as follows: Section 2 explains the data pre-processing process and research method. In Section 3, we propose the experimental results. In Section 4, we describe environmental proposals for using MediaPipe. Finally, we conclude this paper and present future work in section 5.

2. METHODS

This section describes the data extraction, preprocessing process, and the model used in the study.

2.1. MediaPipe Description and Data Extraction

MediaPipe is an open-source gait analysis system similar to OpenPose developed by Google in June 2019[10]. Like OpenPose, MediaPipe can extract joint points of a person's face, hand, and body in real-time from a photo or video. However, as shown in Table. 1, MediaPipe extracts more joint points than OpenPose and can express body joints and faces more delicately.



(a) (b)

Figure 1: MediaPipe and OpenPose Joint Points

Table 1
Number of MediaPipe and OpenPose joint points

	OpenPose	MediaPipe
Hand	21	21
Face	70	468
Body	25	33

This study extracts 33 body points from video data using the pose model of MediaPipe, as shown in Fig. 1(a). And we get the angle data of the right and left knee, right and left hip, and nose-shoulder using the DMS(Data Management System) of DEEVO Co. Ltd.

2.2. Data Description and Pre-processing Process

This paper is offered lateral gait video data of 78 sarcopenia patients and 29 normal persons from the Gyeongsang National University Hospital Orthopedic. After pre-processing, this paper uses lateral gait image data from 38 sarcopenia patients and 21 normal persons.

In the data pre-processing process, we remove parts of videos that can cause recognition errors when applying MediaPipe, such as parts that are not on the lateral of pedestrians or that the camera frame covers the joints. And we apply MediaPipe to extract joint points and calculate the angles of both knees, hips, and nose-shoulder using the DMS of DEEVO Co. Ltd. And we exclude the video data with spike values of the angle data caused by the MideaPipe's recognition error that can affect the model training. MediaPipe's recognition errors refer to these cases: recognizing the background as a pedestrian's joint because of clothing in color similar to the background, or the wrong joint when the pedestrian's body part is out of the camera frame, or an angled object as a pedestrian's joint. This paper constructs lateral gait video data of 38 sarcopenia patients and 21 normal patients. The LSTM-Autoencoder model trains using this angle data and predicts sarcopenia with anomaly detection.

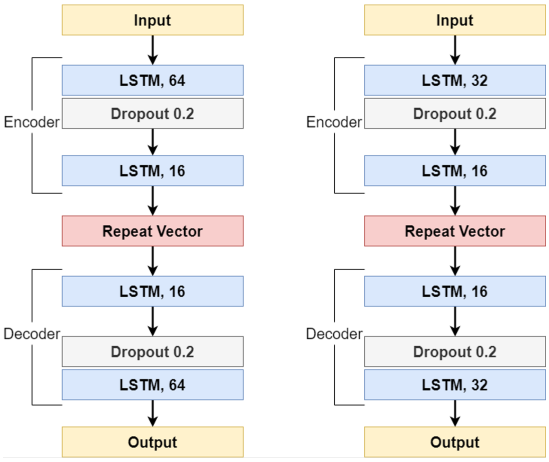
2.3. Sarcopenia Gait Detection using LSTM-Autoencoder

LSTM is a deep learning neural network algorithm and is a model created to overcome the vanishing problem, which is a drawback of RNN. An Autoencoder is a neural network that compresses input data and restores the compressed data to its original form. LSTM-Autoencoder is a model that combines these two models. The LSTM-Autoencoder has a low loss value for the time series data as learned and a relatively high loss value for the untrained data[11]. Therefore, the LSTM-Autoencoder in this paper trains the normal gait data and determines the threshold for anomaly detection.

Table 2
Hyper-parameters: RH(Right hip), LH(Left hip), RK(Right knee), LK(Left knee), NS(Nose shoulder)

	RH	LH	RK	LK	NS
Optimizer	Adam	Adam	Adam	Adam	Adam
Epoch	200	150	150	200	150
Batch size	16	32	32	64	64
Learning Rate	0.0001	0.0001	0.0001	0.0001	0.0001

Fig. 2 shows the model structure of this study. The model that trained the right hip data has the structure shown in Fig. 2(a). And the models that trained the body parts except for the right hip have the structure as shown



(a) (b)

Figure 2: LSTM-Autoencoder Model Architecture

in Fig. 2(b). In the model structure of Fig. 2(a), (b), the number of features in the encoder's first layer and the decoder's last layer is different. Our model trains each body part's 18 normal gait data based on the hyper-parameters in Table. 2. And we determine the threshold by averaging each data's most significant loss values.

3. RESULTS

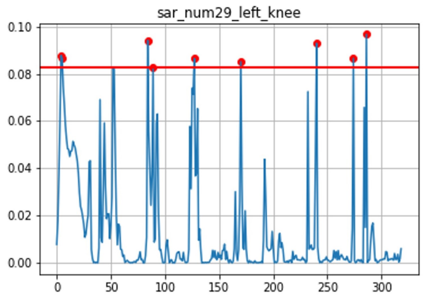


Figure 3: Anomaly Detection Result Graph

Fig. 3 shows a graph detecting outliers in the left knee of sarcopenia patient who have video ID 29. The red line is the threshold obtained through training, and the red dots are outliers.

Fig. 4 shows the number of sarcopenia patients detected by each body part. The number of sarcopenia anomaly detected is 35 out of 38, showing an accuracy of about 92%. The number detected in both knees is more than 30, and the nose-shoulder is 10. On the other hand, the right and left hips show relatively low results with 7 and 5 sarcopenia patients, respectively.

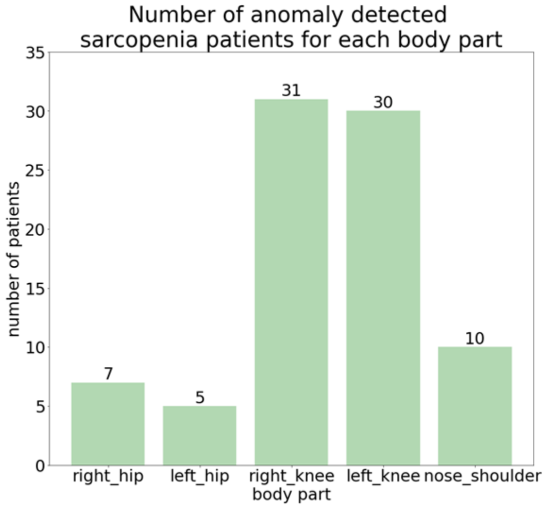
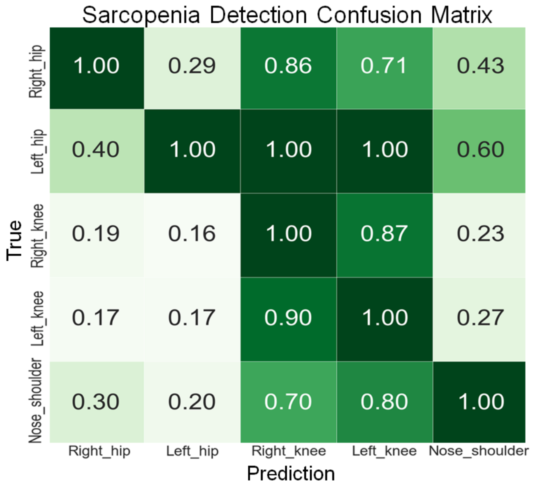


Figure 4: Number of anomaly detected sarcopenia patients for each body part



	Right_hip	Left_hip	Right_knee	Left_knee	Nose_shoulder
Right_hip	1.00	0.29	0.86	0.71	0.43
Left_hip	0.40	1.00	1.00	1.00	0.60
Right_knee	0.19	0.16	1.00	0.87	0.23
Left_knee	0.17	0.17	0.90	1.00	0.27
Nose_shoulder	0.30	0.20	0.70	0.80	1.00

Figure 5: Sarcopenia Detection Confusion Matrix

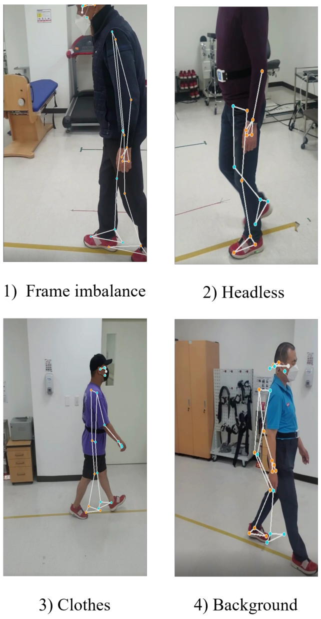
Fig. 5 is a Confusion Matrix showing rates of predicting sarcopenia in other body parts when predicted in a particular body part. For example, 0.29 in the first column and second row means the ratio of the number of predicted sarcopenia through the Right hip and Left hip to the number of predicted sarcopenia through the Right hip. When both knees are predicted to be positive, the rate of other body parts also predicted to be positive is about 85% and about 28% for both hips and nose-shoulder. In other words, both knees have a higher anomaly detection ratio than both hips and nose-shoulder. And among both knees, the anomaly detection rate of the right knee is higher.

4. SUGGESTING ENVIRONMENTS TO USE MEDIAPIPE

The most significant advantage of MediaPipe is that gait analysis is possible with a smartphone without a high-spec camera. It is easy to analyze for an inexperienced expert. Also, as shown in Table. 1, MediaPipe recognizes more joint points than OpenPose, enabling accurate gait analysis. However, there are environmental proposals when taking a video to increase the recognition accuracy of MediaPipe.

Therefore, this paper presents the four environmental factors affecting MediaPipe recognition and their solutions: frame imbalance, headless, clothing, and background.

4.1. 4 Environmental Factors Affecting MediaPipe Recognition


1) Frame imbalance 2) Headless

3) Clothes 4) Background

Figure 6: 4 Environmental Factors Affecting MediaPipe Recognition

1) Frame imbalance: It is a case where some body part of a pedestrian is out of frame. In Fig. 6(1), there is a recognition error because the pedestrian's left foot is out of the frame.

2) Headless: It is a case in which only the body below the neck is shown in the video. In Fig. 6(2), a recognition error occurred because a video frame does not include the pedestrian's head and shoulder.

3) Clothing: It is a case where pedestrians wear clothing similar in color to the background, clothing too large or too loose for their bodies. In Fig. 6(3), MediaPipe recognized the knee as the angled hem of the large black shorts.

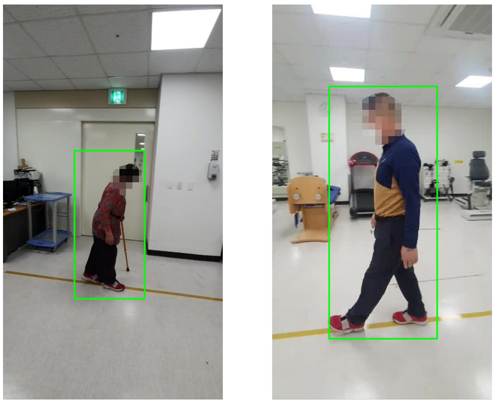
4) Background: It is a case where the video included angular objects such as drawers and doors. In Fig. 6(4), a recognition error is caused by a drawer behind the pedestrian.

4.2. Proposals to increase the recognition accuracy of MediaPipe

In the pre-processing process of subsection 2.2, we analyze the angle data of the video is applying MediaPipe. More than 50% of the total video data have a MediaPipe recognition error, resulting in spike values that can affect the model training. The recognition error is caused by the four environmental factors MediaPipe recognition described in subsection 4.1: frame imbalance, headless, clothing, and background. Only video editing cannot block all recognition errors for video data, including any of the four environmental factors MediaPipe recognition. Therefore, we present four environmental proposals when taking a video.

1) Frame: MediaPipe is a real-time gait analysis system that recognizes pedestrians for each video frame. However, the frame imbalance described in subsection 4.1 can cause MediaPipe's recognition errors in cases: where the distance between the camera and the pedestrian is not kept constant, that a body part is out of video frame due to the difference between the camera's moving speed and the human's gait speed. As a result of anomaly detection using angle data of normal videos with recognition errors and sarcopenia patients' videos, they are all detected as sarcopenia patients. We suppose this result to be the effect of the spike values of the angle data caused by the recognition error. Therefore, if we use a video with frame imbalance, a normal person may be predicted as a sarcopenia patient. Also, it may be difficult to determine the severity of sarcopenia.

In this paper, we propose an ideal frame and maximum frame range that does not cause frame imbalance when taking a video, using the coordinates of the bounding box of the pedestrian center of the video through the YOLO algorithm.



(a) Long distance (b) Short distance

Figure 7: Images of sarcopenia patients taken from a long and short distance

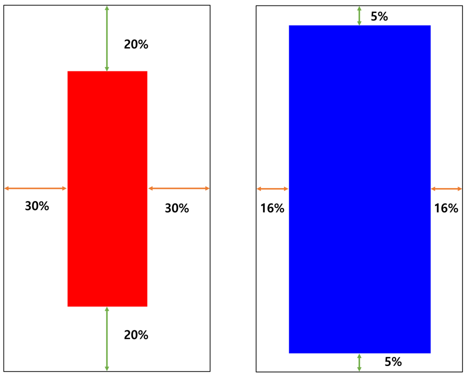
This paper uses the data from 38 sarcopenia patients in subsection 2.3 and the YOLO algorithm to determine the ideal and maximum frame range. YOLO(You Only Look Once) is an algorithm that detects people or things through object detection in an image or video. For each video frame with a pedestrian in the center of each video data, YOLO v3[12] draws a bounding box centered on the pedestrian, as shown in Fig. 7. Also, YOLO v3 offers the coordinates of the bounding box's upper left and lower right. Based on the average of each coordinate value, we get the top, bottom, left, and right margins of the remainder of the video frame except for the bounding box. And this paper obtains the top, bottom, left, and right average margins of the 38 sarcopenia patients' video data.

This paper classified each video data as long and short distance, using the bottom, left, and right margins. If it is larger than the average of each margin, it is classified as long distance, and if it is smaller, it is classified as short distance. The reason for excluding the top margin is that it could be affected by the height of the pedestrian. Then, each video data is divided into long and short distances by the frequency of the previously classified distance. Finally, we divide the video data into 13 long-distance and 25 short-distance samples.

Table 3
Minimum and Total Average

Margin	Top	Bottom	Left	Right
Minimum(%)	2.2	8.2	16.4	15
Total Average(%)	17.1	18.5	33.5	28.1

TABLE. 3 shows the minimum of the top, bottom, left, and right margins in 38 sarcopenia video data. Also, it shows the average of the top, bottom, left, and right margins in total sarcopenia video data. Based on this, this paper proposes an ideal and maximum frame range that can recognize pedestrians.



(a) Ideal frame range (b) Maximum frame range

Figure 8: Frame Range

Fig. 8 shows the ideal and the maximum frame range for recognizing pedestrians. We calculated the frame ranges using the average of the top, bottom and left, right margins for the minimum and total averages in Table. 3. Fig. 8(a) shows the ideal frame range. The top and bottom margins are 20%, and the left and right margins are 30%. Fig. 8(b) shows the maximum frame range. The top and bottom margins are 5%, and the left and right margins are 16%. Therefore, we recommended taking a video within the ideal frame range of Fig. 8(a) to recognize pedestrians when applying MediaPipe. However, if not, it is good to maintain the maximum frame range in Fig. 8(b). When taking a video to apply MediaPipe, we can improve the recognition accuracy by following these proposals. Also, it can reduce the recognition error caused by 2) headless.

3) Clothing: It is better to wear clothing that shows body shape than clothing that covers joint points, such as skirts, shorts, and loose clothing. Also, clothing that contrasts colors with the background can reduce recognition errors.

4) Background: MediaPipe tends to recognize angled objects as joints, so it is good to avoid angled objects in the background if possible. We can remove or cover angled objects in a limited environment before taking a video. However, there is a problem that it cannot resolve recognition errors caused by angle objects in environments where objects cannot be restricted, such as a street or a house.

5. CONCLUSION AND FUTURE WORK

This paper proposes a sarcopenia detection technique using the LSTM-Autoencoder-based anomaly detection method. And it presents four environmental factors MediaPipe recognition and their solutions: frame imbalance, headless, clothing, and background that affect recognition error when applying MediaPipe.

As a result of the technique, outliers are detected in 35 out of 38 sarcopenia patients with 92% accuracy. From Fig. 5, both knees are the most sensitively detecting body parts. Among both knees, the right knee has a higher anomaly detection rate. We suppose that the low anomaly detection rates in parts except for knees are due to characteristics that distinguish between sarcopenia and normal being uncertain. Therefore, this paper expects a better performance of a model that can respond sensitively to data in the future.

This paper presents four environmental factors MediaPipe recognition and their solutions: frame imbalance, headless, clothing, and background that affect recognition when applying MediaPipe. Frame imbalance is a case where some body part of a pedestrian is out of frame. It can reduce recognition error by maintaining the previously proposed ideal frame range for pedestrian recognition when taking a video or by retaining the maximum frame range when impossible. This frame range can also solve recognition errors due to the headless that appears only on the body below the neck in the video. And clothing that shows the pedestrian's body shape well and in color contrast with the background can make fewer recognition errors. However, in the background case, there is a problem in that the recognition error cannot be entirely solved for the video taken in an environment where angular objects cannot be restricted, such as a street or a house.

For future work, we intend to research how to improve the accuracy of gait analysis by solving the background problem among the four environmental factors affecting MediaPipe recognition.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1A2C1014855).

References

[1] T. Perepelkina, A. B. Fulton, Artificial intelligence (ai) applications for age-related macular degeneration (amd) and other retinal dystrophies, in: Seminars in ophthalmology, volume 36, Taylor & Francis, 2021, pp. 304–309.

[2] R. Vaishya, M. Javaid, I. H. Khan, A. Haleem, Artificial intelligence (ai) applications for covid-19 pandemic, Diabetes & Metabolic Syndrome: Clinical Research & Reviews 14 (2020) 337–339.

[3] K. R. Kaufman, C. Hughes, B. F. Morrey, M. Morrey, K.-N. An, Gait characteristics of patients with knee osteoarthritis, Journal of biomechanics 34 (2001) 907–915.

[4] S. P. Messier, R. F. Loeser, J. L. Hoover, E. L. Semble, C. M. Wise, Osteoarthritis of the knee: effects on gait, strength, and flexibility, Archives of physical medicine and rehabilitation 73 (1992) 29–36.

[5] F. Chen, X. Cui, Z. Zhao, D. Zhang, C. Ma, X. Zhang, H. Liao, Gait acquisition and analysis system for osteoarthritis based on hybrid prediction model, Computerized Medical Imaging and Graphics 85 (2020) 101782.

[6] A. Mirelman, M. Ben Or Frank, M. Melamed, L. Granovsky, A. Nieuwboer, L. Rochester, S. Del Din, L. Avanzino, E. Pelosin, B. R. Bloem, et al., Detecting sensitive mobility features for parkinson's disease stages via machine learning, Movement Disorders 36 (2021) 2144–2155.

[7] N. Goldfarb, A. Lewis, A. Tacescu, G. S. Fischer, Open source vicon toolkit for motion capture and gait analysis, Computer Methods and Programs in Biomedicine 212 (2021) 106414.

[8] A. Pfister, A. M. West, S. Bronner, J. A. Noah, Comparative abilities of microsoft kinect and vicon 3d motion capture for gait analysis, Journal of medical engineering & technology 38 (2014) 274–280.

[9] A. Patil, D. Rao, K. Utturwar, T. Shelke, E. Sarda, Body posture detection and motion tracking using ai for medical exercises and recommendation system, in: ITM Web of Conferences, volume 44, EDP Sciences, 2022, p. 03043.

[10] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, et al., Mediapipe: A framework for building perception pipelines, arXiv preprint arXiv:1906.08172 (2019).

[11] P. Liu, X. Sun, Y. Han, Z. He, W. Zhang, C. Wu, Arrhythmia classification of lstm autoencoder based on time series anomaly detection, Biomedical Signal Processing and Control 71 (2022) 103228.

[12] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, arXiv preprint arXiv:1804.02767 (2018).