Abstract: Cardiovascular diseases are the leading cause of death worldwide. Therefore, getting help in time makes the difference between life and death. In many cases, help is not obtained in time when a person is alone and suffers a heart attack. This is mainly due to the fact that pain prevents him/her from asking for help. This article presents a novel proposal to identify people with an apparent heart attack in colour images by detecting characteristic postures of heart attack. The method of identifying infarcts makes use of convolutional neural networks. These have been trained with a specially prepared set of images that contain people simulating a heart attack. The promising results in the classification of infarcts show 91.75% accuracy and 92.85% sensitivity.

Keywords: image analysis; heart attack detection; convolutional neural networks

1. Introduction

Cardiovascular diseases are the leading cause of death throughout the world, according to the World Health Organisation (WHO) [1]. Age-related illnesses such as heart problems used to appear as a person gets older. Moreover, the average age is increasing worldwide according to the World Bank [2]. In some countries like Italy, Greece, and Japan, the population older than 65 years exceeds 20% of the total, and the percentage tends to increase. Moreover, many older people worldwide prefer to live independently at home rather than move to a nursing centre. However, the decision to live alone increases the likelihood of not receiving timely assistance during an emergency, and even more so when a person lives in remote locations.

For this reason, many researchers have been developing methods and mechanisms to automatically detect abnormal events over the last decades [3,4]. A heart attack is an example of a situation in which timely care makes the difference between life and death. When a person experiences a heart attack, one main symptom is a strong pain in the chest [5]. A person living alone will find it very difficult to ask for help because of the pain caused by the infarct. Therefore, it is necessary to provide mechanisms that automatically detect events that affect a person’s health as in the case of heart attacks.

The strong pain in the chest due to a heart attack, present in most cases [6], leads to a position in which the person brings their hands to the chest and the upper body moves forward. This posture could be useful to detect a heart attack using computer vision techniques. However, as far as we know, there is no project based on human postures and gestures capable of detecting a possible infarction by processing a single image.
This article proposes a method for detecting heart attack events using a non-invasive method. That is, it does not require a person to wear a series of devices, a specialist to directly monitor the patient, or the person to operate a system in some way. This work included the creation of a data set of images with people simulating infarction conditions for training the convolutional neural networks (CNNs) constructed. Our proposal has achieved 91.75% accuracy and 92.85% sensitivity in detecting people with a potential heart attack. It consists of a visual monitoring system that enables the identification of people with apparent postures of infarcts, using CNNs to analyse colour images. The proposed system is intended to be added to larger systems such as intelligent health environments [7,8] or accompanying robots [9,10], enriching their capabilities.

The remainder of the paper is organised as follows. Section 2 presents several related works and their strategies in the recognition of human activity. Section 3 shows the methodology used to develop, train, and test the proposal. This section also presents the results obtained and a comparison with related works. Finally, some relevant conclusions and possible future works are described in Section 5.

2. Related Work

Many researchers are concentrating their efforts on solving different challenges in health care [11,12], specifically through human activity recognition [13,14]. This has opened the door to the detection of abnormal events in an automatic manner. Fall detection is by far the most commonly faced challenge and the top topic in health environments [15,16], but there exist other challenges like monitoring Parkinson’s disease [17] or even recognising emotional states [18,19].

Regardless of the event to be detected, it is possible to classify all publications into two principal approaches according to the acquisition sensors of the input data, which are non-visual and visual. For non-visual sensors, the use of accelerometers and gyroscopes [20–22] should be highlighted. Although these devices provide high precision, their main drawback is that they require that people wear them all the time, which is uncomfortable and not always possible. The second class uses cameras and is always based on computer vision to analyse the captured images or videos.

Regarding the image-based approach, the use of the Microsoft Kinect device is a solution often found in the literature [12,23–25]. This sensor simplifies some tasks like background subtraction and/or skeleton generation but limits or reduces the accuracy after a few meters of distance [26]. Another option consists in analysing the provided colour information directly by using artificial intelligence techniques [26,27]. In fact, the use of neural networks, and especially CNNs, has obtained excellent classification results in identifying particular events in recent years [24,27,28].

The most traditional approach to identify specific human events by analysing images is the identification of the human’s posture [29–31]. This can be achieved by analysing the silhouette of the person [16,25], obtaining the skeleton [32–34], processing the complete information of the person’s image, which is the approach used in this work, or even a combination of techniques [35].

3. Methodology

The objective of this work is the development of an algorithm to automatically detect when a person is having a possible heart attack by analysing images using CNNs. This work followed the steps described next and depicted in Figure 1: (1) generation of a set of images to be used for training, validation, and testing; (2) design of CNNs in charge of the identification of heart attacks; (3) training and validation of the CNNs; and, (4) performing of tests to determine the accuracy of the system. It is important to clarify that steps (3) and (4) were repeated until satisfactory results were achieved, as shown in the figure.
3.1. Creation of the Image Data Set

A data set of images with people in non-infarction situations and others with a possible infarction was generated. The data set is made up of images created by the authors plus others downloaded from the Internet. All these pictures contain only one person. In regards to images tagged as an infarct, all people show a posture in which they have one or both hands on the chest. As regards to no infarct situations, images where people are performing daily activities were used. Some of these latter images include similar postures to a possible infarct, but they are not labelled as such. The initial image data set contained a total of 1520 images, 760 images of class “Infarct” and 760 images of class “No Infarct”. To increase the number of images in the data set, the following actions were performed:

- Each image was scaled to a maximum size of $256 \times 256$ pixels, maintaining the original proportion, both for “Infarct” and “No Infarct” images. This was done in order to reduce the amount of data to be processed during the augmented data technique.
- After that, the images were classified into two categories, “Infarct” and “No Infarct”. Furthermore, each category was split into the three subcategories of training, validation, and testing, as shown in Table 1. This process was done manually in order to ensure that the images from each subcategory would not be repeated. This means that an image that is being used for training, for instance, will not be used again for validation or testing purposes, thus avoiding an alteration of the final results.
- As the CNNs only have to infer a possible heart attack, people were extracted from the background of the image by reducing the noise caused by the variation of the background in order to improve the training set (see Figure 2).
  - People were automatically located in each image. For this purpose, the Mask R-CNN software for object detection and instance segmentation was used [36].
  - From the previous result, the background of the image was removed, replacing the value of each pixel with purple colour.
- Finally, the augmented data technique was applied [37]. Data augmentation is a process for generating new samples by transforming training data with the target of improving the accuracy.
and robustness of the classifiers. In this case, each original picture generated 20 more different images. For this, six transformations were combined and applied to each image (rotation, increase/decrease in width or height, zoom, horizontal flip, and brightness change).

Table 1. Data in selected distribution: Images per class and set.

| Class     | Training 70% | Validation 15% | Test 15% | Total |
|-----------|--------------|----------------|----------|-------|
| Infarct   | 532          | 114            | 114      | 760   |
| No Infarct| 532          | 114            | 114      | 760   |
| Total     | 1064         | 228            | 228      | 1520  |

**Figure 2.** Location and extraction of a person. (1) Original image. (2) Location of the person. (3) Background removal, cropping, and re-sizing.

As a result, a total of 31,920 images were obtained by adding the augmented images to the original pictures, as shown in Table 2. It is important to mention that each image generated by the transformations during the data augmentation process was kept in the same set to which the original image belonged, ensuring that both the original image and its transformations belonged to no more than one set.

Table 2. Amount of original images and total number after data augmentation.

| Class     | Initial Training | Training Augmented | Initial Validation | Validation Augmented | Initial Testing | Testing Augmented | Final   |
|-----------|------------------|--------------------|--------------------|---------------------|----------------|------------------|--------|
| Infarct   | 532              | 10,640             | 114                | 2280                | 114           | 2280             | 15,960 |
| No Infarct| 532              | 10,640             | 114                | 2280                | 114           | 2280             | 15,960 |
| Total     | 1064             | 21,280             | 228                | 4,560               | 228           | 4560             | 31,920 |

**Figure 3** introduces an example of the resulting images, both of people with a heart attack and others in different positions. The data set includes images of the same people in different postures, as this allows the neural networks to identify the key features in the posture rather than to recognise the person shown in the image.
3.2. Design of Convolutional Neural Networks

At this stage, convolutional neural networks were built to identify the person’s posture, considering the defined size of the input images. Different tests were carried out with several layer combinations and configurations in order to reach the final model (presented in Figure 4).

![Figure 4. Convolutional neural network architecture.](image)

At the beginning of the network, the proposed design has five convolutional blocks, where each block is firstly composed of a convolution layer to highlight the general features in the image. Then, a max pooling layer is provided to keep the number of variables of the network low, in this way maintaining a size easy to compute.

In the middle of the network, just after the convolution blocks, there is a dropout layer that prevents the generated model from presenting an envelope training, mainly due to the limited amount of data. After this, a flatten layer allows changing the 2D design of the convolutional layers to a vectorial one so that the values generated in the previous layers are passed to the traditional neuron layers.

At the end of the network, ten layers composed of traditional neurons are arranged, each with 128 neurons, which deliver the result of forward propagation to a softmax function with two outputs. These will classify whether there is or is not a person with a heart attack in the image.

3.3. Training, Validation and Test

The robustness of a model is not only given by the accuracy rate obtained during training but also by its precision when tested with unknown data. For this reason, it is necessary to add robustness at
the training stage. Therefore, a part of the available data set must be split for training, another for validation, and another for testing. To ensure the best distribution of data for training, validation, and testing, the design of the networks were tested with several distributions. Table 3 shows the four data distributions that were assessed.

Table 3. Percentage of data set images used in each distribution for training, validation, and testing.

| Distribution | Training | Validation | Test |
|--------------|----------|------------|------|
| A            | 80%      | 15%        | 5%   |
| B            | 70%      | 15%        | 15%  |
| C            | 70%      | 0%         | 30%  |
| D            | 80%      | 0%         | 20%  |

After carrying out the proves with the different data distributions, it became evident that the design maintained its robustness regardless of the selected distribution, as they present similar accuracy rates (see Figure 5). The figure shows the accuracy (left) and loss (right) at each step for the four distributions during training. From the shown graphs, it stands out that the learning speed in all distributions is similar. All four distributions require less than 100,000 steps to reach their best result.

![Figure 5. Comparison of training accuracy (left) and loss (right) for the four distributions.](image)

For that reason, only one distribution was selected to validate the results. We chose the B distribution (70%–15%–15% for training, validation, and testing, respectively), which is a typical configuration in many other applications based on neural networks [38–40]. Please consider Table 1 again for a description of the number of images used for this distribution.

Once the training images were selected, the model was built using the Tensorflow [41] framework. A precision of 99% was achieved during training, which allowed 91.75% accuracy and 92.85% sensitivity in the test set, defining a learning rate of 0.003 and using the gradient descent optimiser. Both the source code and the image data set can be downloaded from https://github.com/Turing-IAC/IHC/Heart-Attack-Detection-In-Images.

4. Results

For the assessment of the developed system, 2394 images of the “Infarct” class and 2394 of the “No Infarct” class were taken. The sensitivity, specificity, and accuracy (in percentages) of the proposed CNN-based method for heart attack detection were calculated as:

\[
sensitivity = \frac{TP}{TP + FN} \times 100, \tag{1}
\]

\[
specificity = \frac{TN}{TN + FP} \times 100, \tag{2}
\]
\[
\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100, \tag{3}
\]

where \(TP\) (true positives) is the number of images correctly identified as an infarct, \(FN\) (false negatives) is the number of images incorrectly identified as no infarct, \(FP\) (false positives) is the number of images incorrectly identified as an infarct, and \(TN\) (true negatives) is the number of images correctly identified as no infarct. In this case, \(FP = 224, FN = 171, TP = 2223,\) and \(TN = 2170\). These numbers obtained a sensitivity of 92.85\%, a specificity of 90.64\%, and an accuracy of 91.75\%.

Although the detection of heart attacks using the strategy proposed here is unprecedented, there are multiple works focused on identifying specific human activities and abnormal behaviours. This is the reason why the present work compares the obtained accuracy with reference works that identify other types of events (see Table 4). As shown, our method overtakes previous works with similar approaches in terms of accuracy. This demonstrates that our proposal can be used as an effective way to classify events in human activities, even when the initial sample is limited.

A strategy similar to the one presented in this work, using RGB-D (red, green, blue plus depth) images, colour subtraction, and a CNN, was presented for the identification of fall events [24]. The difference lies in the source of the images. In this case, a Kinect sensor was used, also obtaining depth information that was used to eliminate the background. The accuracy is poorer, by far (74%). Regarding the use of specialised sensors such as the Kinect, the main problems of their implementation lies (a) in the increase in costs and (b) in the fact that there are already effective methods to extract a person from the background without their use.

Table 4. Comparison between similar approaches.

| Work | Model | Activity Identified | Accuracy | Data Source | Data Set |
|------|-------|---------------------|----------|-------------|---------|
| [24] | CNN   | Fall                | 74\%     | RGB-D input, background subtraction | Own, captured with Kinect |
| [34] | Hierarchical maximum entropy Markov model | talking on the phone, drinking water, talking, relaxing, writing | 84.7\%  | RGB-D input, Skeleton | Own, captured with Kinect |
| [27] | CNN and RNN/LSTM | Fall  | 88.9\%  | RGB input  | Not reported |
| [25] | CNN and RNN/LSTM | Fall  | 60.76\% | RGB-D input, background subtraction | Own, captured with Kinect |
| [26] | CNN   | Fall                | 65.0\%  | RGB input  | YouTube Fall data (YTFD) set |
| Ours | CNN   | Infarct             | 91.7\%  | RGB input, background subtraction | Own, Internet images |

As in the previous case, the Kinect sensor is used in another approach [34]. Instead of subtracting the background, this paper makes use of the skeleton delivered by the proper device. It is worth noting that the proposal not only focuses on a single activity but tries to identify up to 12 activities in 5 different environments. It makes use of the hierarchical maximum-entropy Markov model, obtaining a global accuracy of 84%.

Unlike in the previous approaches, the data source is traditional RGB (red, green, blue) images in another paper [27]. In addition, this paper extracts the person from the background using a CNN. The difference with our approach is that the authors train an RNN/LSTM (recurrent neural
network/long short-term memory) network using skeletons generated with 14 joints to detect a fall. They communicate an 88.9% accuracy, close to our proposal.

Another proposal [25] makes use of several of the aforementioned strategies. On the one hand, it uses the Kinect device to generate binary images, leaving the background in black and the silhouette of the person in white. Subsequently, the generated images are analysed by a network that starts with convolution layers followed by LSTM layers. This is the most striking aspect of the proposal, its accuracy being the lowest of all the compared papers.

Finally, a work that starts with the creation of a fall data set constructed with YouTube videos has been presented [26]. The novel element of this proposal with respect to other works is the creation of dynamic images. These types of images are an amalgam of several frames of a video in a single image, including the changes occurring in a window of time. This generates images with shadows or strokes coming from the subtraction of pixels. After generating the images, these are passed through the pretrained Oxford’s Visual Geometry Group VGG-16 model [42], which helps to reduce the processing needs due to the learning transfer strategy. However, the results are not very good, which may be due to insufficient data in the training data set or because the type of images used for training does not present any advantage in comparison to traditional images.

5. Conclusions

This paper has introduced a method to identify a possible infarct in RGB images. Our experiments obtained 91.75% accuracy and 92.85% sensitivity in the detection of people in postures associated with infarcts. Our CNN-based algorithm could be implemented in smart and health environments or accompanying robots.

As far as we know, there is no similar proposal to identify a possible infarct using computer-vision-based non-invasive methods. Therefore, the development of the proposal presented in this article should be considered of great importance since it could prevent the death of people.

The paper has shown that convolutional neuronal networks (CNNs), along with adequate data sets, allow quickly and accurately finding and detecting different image patterns that are useful in the health care and medical fields. Likewise, CNNs demonstrate stability in the different forms of distribution of the available data set images for training, validation, and testing. However, it is not always possible to count on a sufficiently large set of data to perform the training of a CNN. In this sense, data augmentation has demonstrated itself to be fundamental to improving the training.

As future work, it is intended to expand the data set and to build other neural network architectures. Moreover, it is our aim to widen the proposal to consider other types of troubles that affect the well-being and health conditions of people’s lives.

Author Contributions: All authors contributed equally to this work.

Acknowledgments: This work was partially supported by the Spanish Ministerio de Economía, Industria y Competitividad, Agencia Estatal de Investigación (AEI)/European Regional Development Fund (FEDER, UE) under grant DPI2016-80894-R.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations
The following abbreviations are used in this manuscript:

| Abbreviation | Definition                      |
|--------------|--------------------------------|
| CNN          | convolutional neural network    |
| FN           | false negative                  |
| FP           | false positive                  |
| TN           | true negative                   |
| TP           | true positive                   |
| WHO          | World Health Organisation       |
References

1. World Health Organization. *The Top 10 Causes of Death;* WHO: Geneva, Switzerland, 2018.

2. The World Bank. *Population Ages 65 and above (% of Total);* The World Bank: Washington, DC, USA, 2017.

3. Yahaya, S.W.; Lotfi, A.; Mahmud, M. A Consensus Novelty Detection Ensemble Approach for Anomaly Detection in Activities of Daily Living. *Appl. Soft Comput.* 2019, 83, 105613. [CrossRef]

4. Dhiman, C.; Vishwakarma, D.K. A review of state-of-the-art techniques for abnormal human activity recognition. *Eng. Appl. Artif. Intell.* 2019, 77, 21–45. [CrossRef]

5. Patel, A.; Fang, J.; Gillespie, C.; Odom, E.; Luncheon, C.; Ayala, C. Awareness of heart attack signs and symptoms and calling 9-1-1 among U.S. adults. *J. Am. Coll. Cardiol.* 2018, 71, 808–809. [CrossRef]

6. Goff, D.C.; Sellers, D.E.; McGovern, P.G.; Meischke, H.; Goldberg, R.J.; Bittner, V.; Hedges, J.R.; Allender, P.S.; Nichaman, M.Z.; for the REACT Study Group. Knowledge of Heart Attack Symptoms in a Population Survey in the United States: The REACT Trial. *JAMA Intern. Med.* 1998, 158, 2329–2338. [CrossRef]

7. Mshali, H.; Lemlouma, T.; Moloney, M.; Magoni, D. A survey on health monitoring systems for health smart homes. *Int. J. Ind. Ergon.* 2018, 66, 26–56. [CrossRef]

8. Fernández-Caballero, A.; Martínez-Rodrigo, A.; Pastor, J.M.; Castillo, J.C.; Lozano-Monasor, E.; López, M.T.; Zangróñiz, R.; Latorre, J.M.; Fernández-Sotos, A. Smart environment architecture for emotion detection and regulation. *J. Biomed. Inform.* 2016, 64, 55–73. [CrossRef] [PubMed]

9. Tang, D.; Yusuf, B.; Botzheim, J.; Kubota, N.; Chan, C.S. A novel multimodal communication framework using robot partner for aging population. *Expert Syst. Appl.* 2015, 42, 4540–4555. [CrossRef]

10. Wilson, G.; Pereyda, C.; Raghunath, N.; de la Cruz, G.; Goel, S.; Nesaei, S.; Minor, B.; Schmitter-Edgecombe, M.; Taylor, M.E.; Cook, D.J. Robot-enabled support of daily activities in smart home environments. *Cogn. Syst. Res.* 2019, 54, 258–272. [CrossRef] [PubMed]

11. Haider, D.; Yang, X.; Abbasi, Q.H. Post-surgical fall detection by exploiting the 5 G C-Band technology for eHealth paradigm. *Appl. Soft Comput.* 2019, 81, 105337. [CrossRef]

12. Pilco, H.; Sanchez-Gordon, S.; Calle-Jimenez, T.; Pérez-Medina, J.L.; Rybarczyk, Y.; Jadán-Guerrero, J.; Maldonado, C.G.; Nunes, I.L. An Agile Approach to Improve the Usability of a Physical Telerehabilitation Platform. *Appl. Sci.* 2019, 9, 480. [CrossRef]

13. Sahoo, S.P.; Ari, S. On an algorithm for human action recognition. *Expert Syst. Appl.* 2019, 115, 524–534. [CrossRef]

14. Khemchandani, R.; Sharma, S. Robust least squares twin support vector machine for human activity recognition. *Appl. Soft Comput.* 2016, 47, 33–46. [CrossRef]

15. Alazrai, R.; Momani, M.; Daoud, M.I. Fall Detection for Elderly from Partially Observed Depth-Map Video Sequences Based on View-Invariant Human Activity Representation. *Appl. Sci.* 2017, 7, 316. [CrossRef]

16. Sokolova, M.V.; Serrano-Cuerda, J.; Castillo, J.C.; Fernández-Caballero, A. A fuzzy model for human fall detection in infrared video. *J. Intell. Fuzzy Syst.* 2013, 24, 215–228. [CrossRef]

17. Cho, C.W.; Chao, W.H.; Lin, S.H.; Chen, Y.Y. A vision-based analysis system for gait recognition in patients with Parkinson’s disease. *Expert Syst. Appl.* 2009, 36, 7033–7039. [CrossRef]

18. Lin, C.J.; Lin, C.H.; Wang, S.H.; Wu, C.H. Multiple Convolutional Neural Networks Fusion Using Improved Fuzzy Integral for Facial Emotion Recognition. *Appl. Sci.* 2019, 9, 2593. [CrossRef]

19. Meza-Kubo, V.; Morán, A.L.; Carrillo, I.; Galindo, G.; García-Canseco, E. Assessing the user experience of older adults using a neural network trained to recognize emotions from brain signals. *J. Biomed. Inform.* 2016, 62, 202–209. [CrossRef]

20. Micucci, D.; Mobilio, M.; Napoletano, P. UniMiB SHAR: A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones. *Appl. Sci.* 2017, 7, 1101. [CrossRef]

21. Guan, Y.; Ploetz, T. Ensembles of deep LSTM learners for activity recognition using wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2017, 1, 11:1–11:28. [CrossRef]

22. Hur, T.; Bang, J.; Huynh-The, T.; Lee, J.; Kim, J.I.; Lee, S. Iss2Image: A novel signal-encoding technique for CNN-based human activity recognition. *Sensors* 2018, 18, 3910. [CrossRef]

23. Stone, E.E.; Skubic, M. Fall detection in homes of older adults using the microsoft kinect. *IEEE J. Biomed. Health Inform.* 2015, 19, 290–301. [CrossRef] [PubMed]
24. Adhikari, K.; Bouchachia, H.; Nait-Charif, H. Activity recognition for indoor fall detection using convolutional neural network. In Proceedings of the 15th IAPR International Conference on Machine Vision Applications, Nagoya, Japan, 8–12 May 2017; pp. 81–84. [CrossRef]

25. Lin, H.Y.; Hsueh, Y.L.; Lie, W.N. Abnormal event detection using Microsoft kinect in a smart home. In Proceedings of the 2016 International Computer Symposium, Chiayi, Taiwan, 15–17 December 2016; pp. 285–289. [CrossRef]

26. Fan, Y.; Levine, M.D.; Wen, G.; Qiu, S. A deep neural network for real-time detection of falling humans in naturally occurring scenes. *Neurocomputing* 2017, 260, 43–58. [CrossRef]

27. Lie, W.N.; Le, A.T.; Lin, G.H. Human fall-down event detection based on 2D skeletons and deep learning approach. In Proceedings of the International Workshop on Advanced Image Technology, Chiang Mai, Thailand, 7–9 January 2018; pp. 1–4. [CrossRef]

28. Yang, H.; Zhang, J.; Li, S.; Lei, J.; Chen, S. Attend It Again: Recurrent Attention Convolutional Neural Network for Action Recognition. *Appl. Sci.* 2018, 8, 383. [CrossRef]

29. Newell, A.; Yang, K.; Deng, J. Stacked hourglass networks for human pose estimation. In *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer International Publishing: Cham, Switzerland, 2016; Volume 9912, pp. 483–499. [CrossRef]

30. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2D human pose estimation: New benchmark and state of the art analysis. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3686–3693. [CrossRef]

31. Fernández-Caballero, A.; Sokolova, M.V.; Serrano-Cuerda, J.; Castillo, J.C.; Moreno, V.; Castiñeira, R.; Redondo, L. HOLDS: Efficient Fall Detection through Accelerometers and Computer Vision. In Proceedings of the 2012 Eighth International Conference on Intelligent Environments, Guanajuato, Mexico, 26–29 June 2012; pp. 367–370. [CrossRef]

32. Zhao, C.; Chen, M.; Zhao, J.; Wang, Q.; Shen, Y. 3D Behavior Recognition Based on Multi-Modal Deep Space-Time Learning. *Appl. Sci.* 2019, 9, 716. [CrossRef]

33. Rojas-Albarracín, G.; Carbajal, C.A.; Fernández-Caballero, A.; López, M.T. Skeleton simplification by key points identification. In *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6256, pp. 30–39. [CrossRef]

34. Sung, J.; Ponce, C.; Selman, B.; Saxena, A. Unstructured human activity detection from RGBD images. In Proceedings of the IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 842–849. [CrossRef]

35. Castillo, J.C.; Carneiro, D.; Serrano-Cuerda, J.; Novais, P.; Fernández-Caballero, A.; Neves, J. A multi-modal approach for activity classification and fall detection. *Int. J. Syst. Sci.* 2014, 45, 810–824. [CrossRef]

36. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988. [CrossRef]

37. Fawzi, A.; Samulowitz, H.; Turaga, D.; Frossard, P. Adaptive data augmentation for image classification. In Proceedings of the International Conference on Image Processing, Phoenix, AR, USA, 25–28 September 2016; pp. 3688–3692. [CrossRef]

38. Valipour, M. Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms. *Metervol. Appl.* 2016, 23, 91–100. [CrossRef]

39. Esfe, M.H.; Saedodin, S.; Sina, N.; Afrand, M.; Rostami, S. Designing an artificial neural network to predict thermal conductivity and dynamic viscosity of ferromagnetic nanofluid. *Int. Commun. Heat Mass Transf.* 2015, 68, 50–57. [CrossRef]

40. Turabieh, H.; Mafarja, M.; Li, X. Iterated feature selection algorithms with layered recurrent neural network for software fault prediction. *Expert Syst. Appl.* 2019, 122, 27–42. [CrossRef]

41. Google Inc. *TensorFlow*; Google Inc.: Mountain View, CA, USA, 2019.

42. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014, arXiv:1409.1556.