ISSN 2278-3091
Volume 9, No.1.3, 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse6791.32020.pdf
https://doi.org/10.30534/ijatcse/2020/6791.32020

Human Falling Recognition using Shallow Convolutional Neural Network

Amirah Syahirah Ali[1], Zaidah Ibrahim[2], Zarina Zainol[3]
[1]Faculty of Computer and Mathematical Sciences, UiTM Shah Alam, Selangor, Malaysia, amirahsyahirah0797@gmail.com.
[2]Faculty of Computer and Mathematical Sciences, UiTM Shah Alam, Selangor, Malaysia, zaidah@tmsk.uitm.edu.my.
[3]Faculty of Computer and Mathematical Sciences, UiTM Shah Alam, Selangor, Malaysia, zarina@tmsk.uitm.edu.my

ABSTRACT

Human falling happens whenever there is a slight change in the position of the human body from sitting, standing or lying to a lower slanting position. This situation mostly happens among old people due to their physical weaknesses associated with aging. It is critical in health security to detect human falling because it can lead to serious consequences such as bone fractures or even death. The problem with the current technology such as wearable sensors is that the elderly often forgets and sometimes reluctant to wear them. Therefore, vision-based approach is proposed to recognize human falling activity using shallow Convolutional Neural Network (CNN). Experiments have been conducted to train and test the CNN model to determine good parameters that achieve the highest accuracy using UR Fall dataset. The experimental results show that human falling activity recognition using shallow CNN is still able to achieve 100% accuracy.

Key words : Convolutional Neural Network, human falling recognition, UR Fall dataset.

1. INTRODUCTION

Human activity recognition (HAR) is one of the popular studied computer vision applications including video surveillance [1][2] and health care [3]-[5] which includes elderly falling activity recognition [6]-[11]. Fall happens when there is a sudden change of body position from lying, standing or sitting position to a lower position [7]. It is a common activity that occurs among senior citizens due to their physical weaknesses associated with aging. Falling can cause negative consequences for their health. In fact, it is one of the major causes of their deaths [9]. However, we must bear in mind that falling cannot be completely As the population of the elderly is rising, this led to the increasing demand of surveillance system that can support falling detection in the healthcare industry [10].

As a result, nowadays human falling activity recognition is becoming a popular computer vision problem [11]. Due to an increase in demand for the technology related to falling detection, it has become an important research area in assistive technology [12]. It is a process where it recognizes the activity or actions performed by a human where it can improve and facilitate the rehabilitation process of patients [13].

Human falling activity recognition can be conducted either using sensor-based [12] or vision-based approaches [14]. Sensor-based is an approach where patients or users wear wearable sensors like accelerometers or help buttons [15]. It can be attached under the armpit or around the waist [10]. But the problem is that the elderly often forgets and sometimes reluctant to wear them [15]. Besides that, the use of a help button has also been introduced, but it can be ineffective if the person is unconscious after the falling incident [15]. A vision-based approach uses camera to detect and recognize falling activity [14]. The highest accuracy for vision-based approach for falling detection and recognition is 0.97 using UR Fall Detection dataset but this achievement may not be sufficient since it involves life and death situation [14]. Thus, accuracy that is better than 0.97 is still necessary to recognize falling activity.

Nowadays, world advances in medical technology thus resulting in a growing population of senior citizens [16]. Falling is a main health problem among the elderly and can cause negative consequences to their health [17] especially if not being entertained fast [18]. It can lead to fracture and serious consequences [16]. It can also turn to severe injuries, loss of independence, and even death [14]. There is also a chance that they may get into a coma, brain trauma or paralysis. Sometimes if unconscious for a long period of time with late rescue may worsen the situation [19][20]. In addition, the rate of falling incidents among the elderly above the age of 65 is increasing [17].

Currently, deep learning is a popular technique in computer vision. It has achieved excellent performance in various image recognition problems such as handwriting word recognition [21], iris recognition [22], and traffic sign recognition [23]. Thus, the aim of this study is to apply Convolutional Neural Network (CNN), a popular technique under deep learning, to recognize human falling activity.

This paper is organized as follows. Section 2 explains the architecture of CNN followed by the experimental results and analysis. The last section summarized this paper.

2.CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

Deep learning has the ability to automatically extract features from the images which removes the necessity of handcrafted feature detectors and descriptors for classification or recognition as in the machine learning approach [24]. Krizhevsky et al. [25] first trained the deep ConvNets also known as Convolutional Neural Network (CNN) in extremely huge image datasets consisting of over 15 million labeled images. The remarkable results lead to its popularity in diverse pattern recognition domains.

CNN has various layers where the output of a previous layer is passed as input to the next layer [25]. Figure 1 illustrates the different layers of CNN architecture.

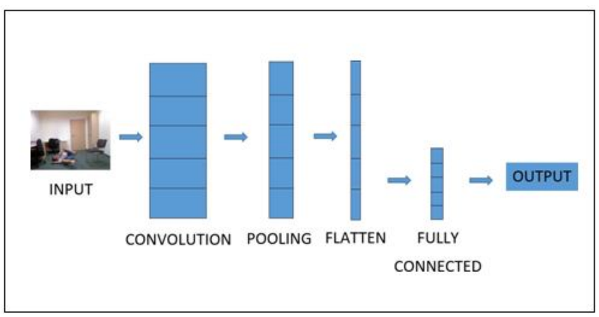

Figure 1: CNN architecture.

At the very beginning of this process, an input image of size 64x64 is broken down into pixels. CNN receives an input image of a three dimensional matrix where the sizes of first two dimensions equal to the width and height of the image while the third dimension corresponds to the number of color channels. In this research, the color channel is 3 for RGB images. What this means is that a single color image actually provides three images as input to the CNN. A grayscale or binary image has the channel of 1 for the third dimension.

The main purpose of the convolution layer is to extract features from the input image. The convolutional layer is always the first layer in CNN. A convolution is a linear operation that involves the multiplication of an array of input data and a two-dimensional array of weights called a filter of kernel and the result of this is the convolved feature map. During a convolution, the filters slide over the input feature map's grid horizontally and vertically, one pixel at a time extracting each pixel. Figure 2 shows an example of how the computation is being performed in the convolution layer. By looking at Figure 2, we can see that the first filter slides over the input image for channel 1, one pixel at a time starting from the top left. The filter multiplies its own values with the overlapping values of the image and adds all of them up to output a single value for each overlap until the entire image is traversed. CNN leans from multiple filters which correspond to multiple features.

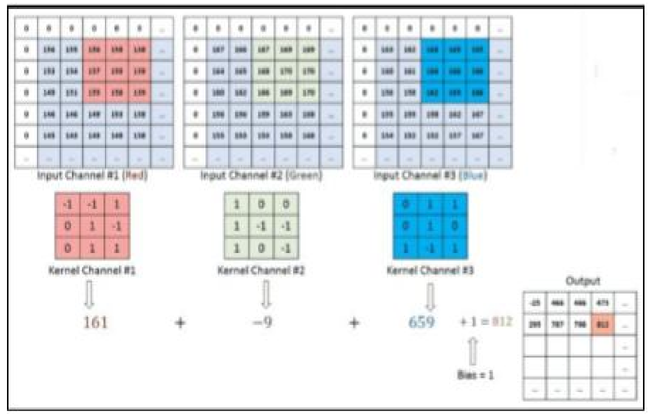

Figure 2: Sample computation in convolution layer [26].

Once a feature map has been created, it is passed to a nonlinearity or activation function. The purpose of applying the activation function is to increase the non-linearity in the images because images are naturally non-linear. One of the popular activation functions is Rectified Linear Unit (ReLU). ReLU applies the function f(x) = max (0, x) to all of the values in the feature map where it changes all the negative values to 0 but maintain the original nonzero values. Figure 3 shows an example of how the result looks like after applying ReLU activation function.

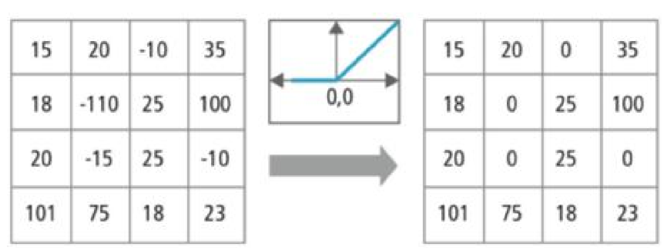

Figure 3: Sample computation in ReLU layer [26].

Another building block of a CNN is pooling layer. It progressively reduces the size of the input representation, thus reducing the amount of parameters and computation. The result is a pooled feature map. It also helps to control overfitting**,** a situation where the model is being trained too well. As a result, it provides very good accuracy on trained data but not on test data.

There are two popular types of pooling that are max-pooling and average pooling. Max-pooling takes only the maximum values from each sub-image while average pooling computes the average from each sub-image to produce the pooled feature map. Figure 4 shows some sample result of max-pooling and average pooling.

The final fully connected layer contains softmax as in Equation (1) which outputs a probability value from 0 to 1 for each type classification.

$$a = softmax(n) = \exp(n)/\mathrm{sum}(\exp(n))$$

Equation (1)

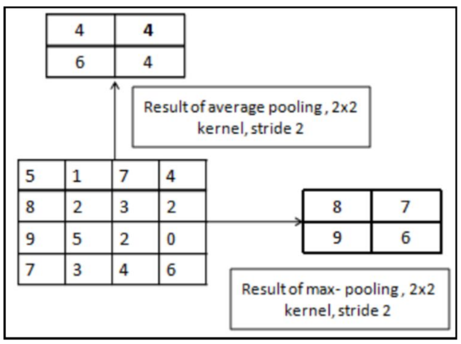

Figure 4: Sample computation in pooling layer.

3. EXPERIMENTAL RESULTS AND ANALYSIS

The dataset used in this research is UR Fall Dataset, retrieved from [16]. This dataset contains 30 different folders that represent 30 videos which have been converted to 100 frames of human falling and 30 different videos of other human activities such as sitting, walking and sleeping. 70% of the images were used for training while 30% were used for testing. Figure 4 and Figure 5 show examples of some of the images used from this dataset.

The performance of the CNN can be improved by adding more convolve and pooling layers and fine-tuning the parameters within the layers. Table 1 shows the recognition performance accuracy produced by investigating the parameters such as applying different types of activation function and pooling function, different number of convolutional layers and pooling layers, with or without dropout, different percentage of dropout, and various classifiers. Dropout is a regularization technique where it randomly select neurons and ignored it during training. Due to time constraints, in this experiment, parameters that are fixed are the number of kernel filters, number of pooling layers which is 1 layer and the size of the kernel which is 3x3.

. By referring to Table 1, we can see that the CNN model with the listed parameter values can achieve accuracy of 1.0 when fed with UR Fall Dataset. This experimental results show that using hyperbolic tangent (tanh) as activation function with average pooling did not achieve accuracy 1.0 on any sigmoid or softmax classifier compared to other parameters. Tanh takes the range from -1 to 1 where it suffers from losing the gradient. On the other hand, when comparing ReLU and ELU activation functions, ReLU takes the range from 0 to infinity which avoids and resolves the problem of eliminating the gradient. All the negative values will become 0 while ELU is very similar to ReLU except it can produce negative outputs. Figure 6 illustrates the accuracy and loss for ReLU with max-pooling and softmax classifier of UR Fall Dataset without dropout while Figure 7 shows the result with dropout.

By referring to Table 2, we can see that the model with the above combination of parameters can achieve accuracy of 1.0 when fed with UR Fall Dataset although the dropout increases from 0.1 to 0.2.

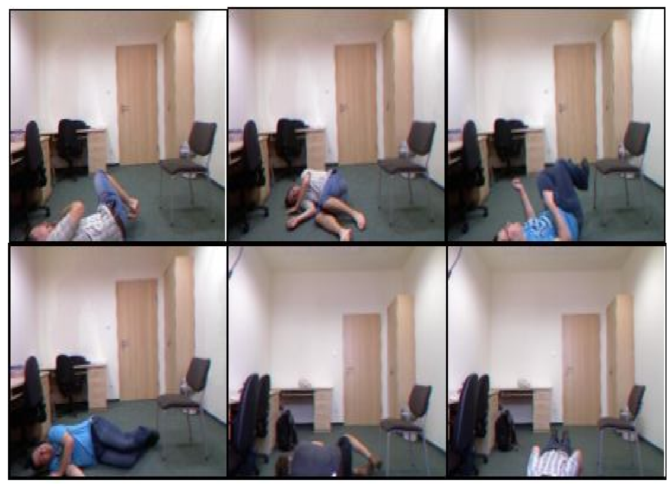

Figure 4 : Examples of different falling activities.

Figure 5 : Examples of non-falling activities.

4. CONCLUSION

This research has presented the accuracy performance of CNN for falling activity recognition. Experimental results show that recognition accuracy of 1.0 can be achieved using shallow CNN with just one convolutional layer with ReLU, one max-pooling layer, softmax at the fully-connected layer. By fine-tuning the parameters within the layers, the performance accuracy can be improved without having to add more layers. Future work involves recognizing other types of human activities for real-time video surveillance system.

Table 1: Experimental results of CNN with different activation and pooling functions.

No	Activation Function at First Convolutional	First Pooling Layer	Activation Function at Fully Connected Layer	Classifier	Accuracy
1	ReLU	MaxPooling	ReLU	Sigmoid	0.9999
2	ReLU	MaxPooling	ReLU	Softmax	1.0000
3	ReLU	AvgPooling	ReLU	Sigmoid	1.0000
4	ReLU	AvgPooling	ReLU	Softmax	0.9999
5	eLU	MaxPooling	eLU	Sigmoid	1.0000
6	eLU	MaxPooling	eLU	Softmax	1.0000
7	eLU	AvgPooling	eLU	Sigmoid	0.9999
8	eLU	AvgPooling	eLU	Softmax	1.0000
9	tanh	MaxPooling	tanh	Sigmoid	0.9999
10	tanh	MaxPooling	tanh	Softmax	1.0000
11	tanh	AvgPooling	tanh	Sigmoid	0.9999
12	tanh	AvgPooling	tanh	Softmax	0.9999

Table 2: Experimental results of CNN with different values of dropout.

No	Activation Function at First Convolutional Layer	First Pooling Layer	Activation Function at Fully Connected Layer	Classifier	Dropout(%)	Accuracy
1	ReLU	MaxPooling	ReLU	Sigmoid	0.1	1.0000
2	ReLU	MaxPooling	ReLU	Softmax	0.1	1.0000
3	ReLU	AvgPooling	ReLU	Sigmoid	0.1	1.0000
4	ReLU	AvgPooling	ReLU	Softmax	0.1	0.9999
5	eLU	MaxPooling	eLU	Sigmoid	0.1	1.0000
6	eLU	MaxPooling	eLU	Softmax	0.1	1.0000
7	eLU	AvgPooling	eLU	Sigmoid	0.1	0.9999
8	eLU	AvgPooling	eLU	Softmax	0.1	0.9999
9	tanh	MaxPooling	tanh	Sigmoid	0.1	0.9999
10	tanh	MaxPooling	tanh	Softmax	0.1	0.9999
11	tanh	AvgPooling	tanh	Sigmoid	0.1	0.9999
12	tanh	AvgPooling	tanh	Softmax	0.1	1.0000
13	ReLU	MaxPooling	ReLU	Sigmoid	0.2	0.9999
14	ReLU	MaxPooling	ReLU	Softmax	0.2	0.9999
15	ReLU	AvgPooling	ReLU	Sigmoid	0.2	0.9998
16	ReLU	AvgPooling	ReLU	Softmax	0.2	0.9998
17	eLU	MaxPooling	eLU	Sigmoid	0.2	0.9999
18	eLU	MaxPooling	eLU	Softmax	0.2	1.0000
19	eLU	AvgPooling	eLU	Sigmoid	0.2	0.9998
20	eLU	AvgPooling	eLU	Softmax	0.2	1.0000
21	tanh	MaxPooling	tanh	Sigmoid	0.2	0.9998
22	tanh	MaxPooling	tanh	Softmax	0.2	0.9999
23	tanh	AvgPooling	tanh	Sigmoid	0.2	0.9999
24	tanh	AvgPooling	tanh	Softmax	0.2	0.9999

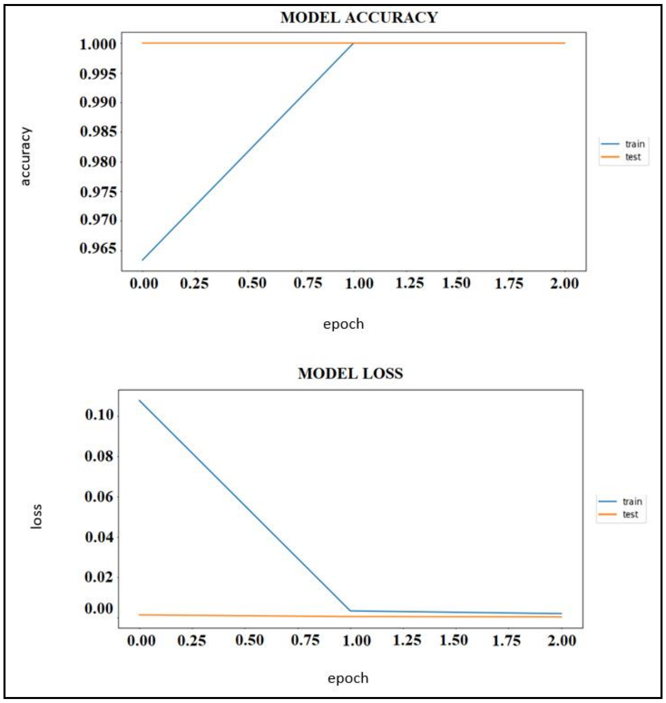

Figure 6: Graph model accuracy and loss for relu with max pooling and softmax classifier of UR Fall Dataset without dropout

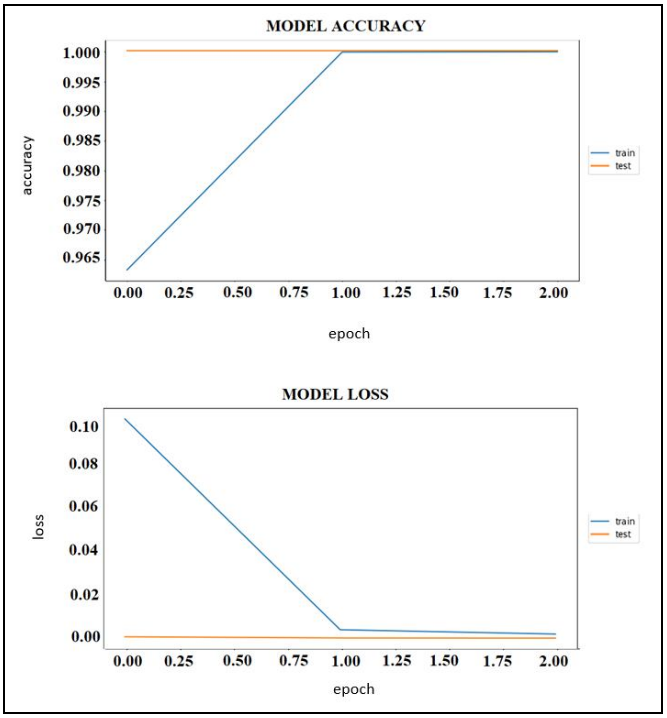

Figure 7: Graph model accuracy and loss for relu with max pooling and softmax classifier of UR Fall Dataset with dropout.

ACKNOWLEDGEMENT

The authors thank the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia, for sponsoring this research.

REFERENCES

1. Muhammad, K., Ahmad, J. and Sung, W. B. **Early Fire Detection Using Convolutional Neural Networks during Surveillance for Effective Disaster Management**, *Neurocomputing*, Vol. 288, 2018, pp. 30-42. https://doi.org/10.1016/j.neucom.2017.04.083
2. Ping, W., Li, L., Yinli, J. and Guiping, W., **Detection of Unwanted Traffic Congestion based on Existing Surveillance Sytem using in Freeway via a CNN-architecture Trafficnet**, *13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2018.
3. Kiranyaz, S., Ince, T. and Gabbouj, M. **Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks**, *IEEE Transactions on Biomedical Engineering*, Vol. 63, Issue 3, 2015. https://doi.org/10.1109/TBME.2015.2468589
4. Chen, M., Shi, X., Zhang, Y., Wu, D., and Guizani, M. **Deep Feature Learning for Medical Image Analysis with Convolutional Autoencoder Neural Network**, *IEEE Transactions on Big Data*, 2017. https://doi.org/10.1109/TBDATA.2017.2717439
5. Alhathi, A., "**Convolutional Neural Network based on Transfer Learning for Medical Forms Classification 34**", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, No. 6, pp.3405-3411, December 2019. https://doi.org/10.30534/ijatcse/2019/115862019
6. Akour, M., Al Qasem, O., Alsghaier, H. and Al-Radaideh, K., "**The Effectiveness of Using Deep Learning Algorithms in Predicting Daily Activities**", International Journal of Advanced Trends in Computer Science and Engineering, Vol 8, No. 5, pp. 2231-2235, October 2019. https://doi.org/10.30534/ijatcse/2019/57852019
7. Noury, N., Fleury, A., Rumeau, P., Bourke, A. K., Laighin, G. O., Rialle, V. and Lundy, J. E. **Fall detection-principles and methods**, *IEEE 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007.
8. Ambrose, A. F., Paul, G. and Hausdorff, J. M. **Risk Factors for Falls among Older Adults: A Review of the Literature**, *Maturitas*, Volume 75, Issue 1, May 2013, pp. 51-61.
9. Yu,Xinguo**, Approaches and Principles of Fall Detection for Elderly and Patient**, *IEEE 10th International Conference on e-Health Networking, Applications and Services*, 2008.
10. Jiunna, W. O. V., Sabrib, N., & Ibrahim, Z. **Image-Based Human Fall Recognition Using Gaussian Mixture Model And Support Vector Machine**, *International Journal of Control Theory and Applications*, Volume 10, Number 20, 2017, pp. 339-344.
11. Nizam, Y., Haji Mohd, M. N., & Abdul Jamil, M. M., **A Study On Human Fall Detection Systems: Daily Activity Classification And Sensing Techniques**, *International Journal of Integrated Engineering*, Vol 8, No 1, 2016.
12. Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., & Li, Z. **A Review on Human Activity Recognition Using Vision-Based Method**, *Journal Of Healthcare Engineering*, 2017.
13. Dobhal, T., Shitole, V., Thomas, G., & Navada, G. **Human Activity Recognition using Binary Motion Image and Deep Learning**, *Procedia Computer Science*, Volume 28, 2015, pp. 178-185.
14. Nunez-Marcos, A., Azkune, G., & Arganda-Carreras, I. **Vision-Based Fall Detection With Convolutional Neural Networks**, *Journal of Wireless Communications And Mobile Computing*, 2017.
15. Rougier, C., Meunier, J., St-Arnaud, A., & Rousseau, J. **Fall Detection from Human Shape and Motion History Using Video Surveillance**, *21st International Conference On Advanced Information Networking And Applications Workshops*, 2007.
16. Lu, K.-L., & Chu, E. **An Image-Based Fall Detection System For The Elderly**, Journal *of Applied Science*, Vol. 8, Issue 10, 2018. https://doi.org/10.3390/app8101995
17. Adhikari, K., Bouchachia, H., & Nait-Charif, H. A**ctivity Recognition For Indoor Fall Detection Using Convolutional Neural Network**, 15th *IAPR International Conference On Machine Vision Applications*, 2017.
18. Shanny, T. Redmond, S. J., Marschollwk, M. and Lovell, N. H. **Assessing Fall Risk Using Wearable Sensors: A Practical Discussion**, Zeitschrift *fur Gerontologie und Geriatrie*, Volume 45, Issue 8, 2012, pp. 694-706.
19. Wu, Falin, Zhao, Hengyang, Zhao, Y. and Zhong, H., **Development of a Wearable-Sensor-Based Fall Detection System**, *International Journal of Telemedicine and Applications*, Volume 2015, 2015.
20. Kwolek, B., & Kepski, M. (2014). **Human Fall Detection on embedded platform using depth maps and wireless accelerometer**. *Computer Methods and Programs in Biomedicine*, 117(3).
21. Poznanskim. A. and Wolf, L., **CNN-N-Gram for Handwriting Word Recognition**, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. https://doi.org/10.1109/CVPR.2016.253
22. Nguyen, K., Fookes, C., Ross, A. and Sridharan, S., **Iris Recognition with off-the-shelf CNN Features:A**

Deep Learning Perspective, *IEEE Access Visual Surveillance and Biometrics: Practices, Challenges, and Possibilities*, Vol. 6, 2017.

23. Shustanov, A. and Yakimov, P. **CNN Design for Real-Time Traffic Sign Recognition**, *Procedia Engineering 201*, 2017, pp. 718-725.
24. Y. LeCun, Y. Bengio, and G. Hinton, **Deep learning**, *Nature*, vol. 521, pp. 436–444, 2015. https://doi.org/10.1038/nature14539
25. A. Krizhevsky, I. Sutskever, and G. E. Hinton, **Imagenet classification with deep convolutional neural networks**, *in Advances in Neural Information Processing Systems*, pp. 1097–1105, Lake Tahoe, Nevada, 2012.
26. Saha, S., **A Comprehensive Guide to Convolutional Neural Networks – the ELI5 way**, 2018, Retrieved on 10 March 2020, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.