https://doi.org/10.24191/jeesr.v20i1.005

Deep Learning Optimisation Algorithms for Snatch Theft Detection

Nurul Farhana Mohamad Zamri, Nooritawati Md Tahir*, Megat Syahirul Amin Megat Ali and Nur Dalila Khirul Ashar

***Abstract*— Learning algorithms related to deep learning use bells and whistles, called hyperparameters. Hence, this study conducted numerical analysis, specifically backpropagation gradients and gradient-based optimization for snatch-theft detection. Here, snatch theft images and augmented images were used to perform the experimental study to determine the optimum hyperparameter values. Next, the value of epoch and learning rate was obtained after careful analysis based on each training option. Results achieved showed that epoch value of 20 and learning rate corresponding to 0.0001 was the optimum values. Findings from this study can be used as a practical guide in determining the importance of the most optimum hyperparameters.**

***Index Terms*— Deep learning neural network, image classification, optimisation, snatch theft detection.**

I. Introduction

CRIMES worldwide include murder, shooting, drug traffic, concealment, fraud, and black marketing. The general trends of each different type of crimes in any country are vital to be analysed to understand better the most prevalent crimes and the reason that contributes to higher crime rates in certain areas. Crime is one of the research areas that can be explored related to crime prediction and detection in urban cities. Additionally, crime is not random. Crime is planned, or the criminals see an opportunity. It usually occurs whenever the victim's and criminal's activity space are intercepts [1]. A person's activity space consists of a place's lifestyle, for example, work, home, shopping, entertainment area or school. For instance, a person is walking in the vicinities. They feel unperturbed since the neighbourhood or environs are familiar to them and might not be aware of changes in these familiar surroundings or the presence of criminals. This is one of the situations that lead to the occurrence of crimes.

On the other hand, street crime is a criminal act usually in urban cities, especially in public spaces. Pick pocketing, mugging, and snatch theft are some examples of street crime [1]. Pick pocketing is a crime in which the offender steals from the pocket of the victim. Criminals may also work in teams, where one person distracts the targeted victim while others steal or bump into the victim to distract. They typically work in crowded areas as it is easy to blend in with other people. The criminal act that threatens the victim in order to steal any valuable pieces is called mugging or robbery, whilst snatch theft is a criminal act of forcibly stealing valuable pedestrian pieces such as necklaces, mobile phones, and handbags through felony and misdemeanour. This can also be a two-person job, one rides a motorcycle, and the other person will steal from their target victims [2]. Snatch theft is a worrying problem for pedestrians since it can lead to accidents and further cause fear, anxiety, and trauma. For instance, in today's modern world, everyone owns a mobile phone and often glance at the phone while walking. Therefore, this action can cause them to lose their sense of surrounding and easily be the victims of snatch theft. Hence in this study, snatch theft is the main focus to be further investigated [3].

Conversely, deep learning neural network (DLNN) is used in image recognition [4], target detection [5], speech recognition [6], machine translation [7], image title generation [8] and the latest technology of driverless cars [9]. However, training and optimisation of DLNN are indeed challenging. Therefore, the simulation and numerical analysis of determining the optimum parameter values of the DLNN model will be investigated, namely the epoch, learning rate, training accuracy and the time consumption for training the model.

II. Related Work

A. Snatch Theft

As we know, some of the major crimes faced across the world include street and theft crimes. Petty theft, such as snatch theft, is increasing drastically, especially in the global economic downturn. One of the previous works is Umair Muneer et al. [10] that reported the use of VGG19 convolution neural network (CNN) to recognise large-scale images or videos of snatch theft crime. They achieved 81% of accuracy in detecting snatch theft using 21 videos as the database. The author occurs some limitations while using MatLab Image Labeller app where the author used it to select features or object that usually occurs during snatching. As they need to consider the number of

This manuscript is submitted on the 23th August 2021 and accepted on 16th December 2021. This work was supported by the Ministry of Higher Education (MOHE) Malaysia, Grant No: 600-IRMI/FRGS 5/3 (394/2019) Sponsorship File No: FRGS/1/2019/ TK04/UITM/01/3.

Nurul Farhana Mohamad Zamri, Nooritawati Md Tahir and Megat Syahirul Amin Megat Ali are with the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Selangor, Malaysia.

*Nooritawati Md Tahir is also with the [2]Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, Selangor, Malaysia (email: nooritawati@ieee.org)

Megat Syahirul Amin Megat Ali is also with the Microwave Research Institute (MRI), Universiti Teknologi MARA, Selangor, Malaysia.

Nur Dalila Khirul Ashar is with the School of Electrical Engineering, Universiti Teknologi MARA, Johor, Malaysia.

*Corresponding author
Email address: nooritawati@ieee.org

moving targets that imitates the act, pace and amount of motion in snatch video. Thus, it can be overcome by carefully selecting/labelling each of images frame in the video.

Besides that, Koichiro Goya et al. [11] implemented public system security and a video surveillance system consisting of four sequential blocks: data collection, object detection, feature extraction, and scene classification. In their research, tracking was based on human motion, using optical flow and background subtraction, and then calculating the object's speed in the target scene during the occurrence of snatch. The security system is limited to longitudinal viewpoint of object whereas if the snatch occurs far from targeted location, the system consumes more time to detect situation. Therefore, it is best to expect that the snatch theft incident will not occur in the same location.

As mentioned earlier, the criminal act of snatching pedestrians' necklaces, cell phones, handbags and many other valuable belongings is known as snatch theft. Run and rob tactics were common methods used by criminals while committing the crimes. Figure 1 shows example scenarios of snatch theft on how the predator usually snatches an item. It was based on two person's that pact about it, one handles the motorcycle, and the other will snatch from the targeted victim or run toward the victim whenever there are chances [2] whilst other figure below show the preparatory cooperates on committing the crime. These are some of the common ploys used by perpetrators during committing snath theft crimes.

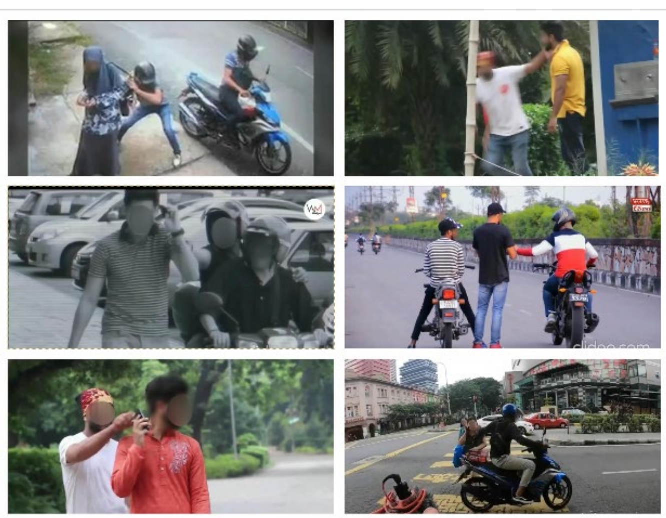


Fig. 1.Example of snatch theft scenarios

B. Optimisation of Neural Network Parameters

Choosing an appropriate hyperparameter for DLNN is a critical task or step for building an effective and accurate model. The process of setting hyperparameters requires extensive numerical analysis and several trial-and-error approaches. There is no definite rule of thumb to select these hyperparameters' optimum values, especially the epoch values, learning rates, batch size, augmenting data and many more. Adjusting model hyperparameters is indeed vital because hyperparameters control the behaviour of the DLNN models.

In addition, the selections of hyperparameters also impact the performance of the training model, which further to be used to predict the outcome of the testing or unseen data. Most of the DL models' performance depends on these hyperparameter values.

1) Epoch

In real-time, large datasets cannot be directly used simultaneously as inputs to the DL model. The entire dataset was segmented prior as inputs to DLNN. Firstly, an epoch was passed forward and backwards only once through the neural network [12]. One epoch in the network can lead to the curve under fitting, which might be overcome by increasing the number of epochs. Note that the epoch number is a hyperparameter that defines the number of times the learning algorithm iterates or functions during the entire training stage of the dataset [13]. As the number of epochs increases, the weights of the network change as well. The curve goes from under fitting to best fitting to over fitting [14]. The number of epochs has no specific value; it relates to the diversity of the dataset. The batch size is the total number of training samples provided and differs from the batch quantity. Conversely, iteration is the number of batches required to complete an epoch. The number of batches is equal to the number of iterations in a period.

One of the critical issues in training the NN using the sample data is that it might lead to over fitting. When the number of epochs that trains the NN model is too large, the training model learns the patterns specific to the sample data to a large extent. This situation prevents the model to perform well in the new or unseen dataset. The model has high accuracy on the training set or sample data, but it could not achieve good accuracy on the testing dataset. In other words, the model loses its generalisation ability due to over fitting based on the training data. The model is trained to produce the best period [15] to reduce over fitting and improve the generalisation ability of NNs. In addition, the training data is dedicated to verifying the model to check the model's performance after each training phase. This study monitors the training set's loss, accuracy, and validity to determine the number of epochs after the model begins too over fit.

2) Learning Rate

The next hyperparameter is the learning rate. Every time the model changes its weight, it responds to the estimated error that is further used to control the changes in the model [16]. Training is more reliable if the learning rate is low; however, optimisation consumes longer time to minimise the loss function. If the learning rate is high, the training may not converge or diverge [17]. The weight change can be significant as well that the optimiser exceeds the minimum range and increases the loss. The learning rate is a configurable hyperparameter required for training the neural network, with a small positive value, usually between 0.0 and 1.0. There are several ways to choose a good starting point for the learning rate. For example, it starts with a more significant value, for instance, 0.1, and then reduces exponentially to 0.01, 0.001, and so on.

Moreover, the learning rate controls the speed that the model can adapt to the training database. The smaller the learning rate, the smaller the weight change for each update and the more training time require, whilst a higher learning rate leads to rapid changes and lesser training time. A higher learning rate may cause the model to converge too fast to a sub-optimal solution, while a lower learning rate may cause difficulty. The challenges of training DLNNs include careful selection of the learning rate since it is one of the essential hyperparameters. The challenges of training DLNNs require a careful learning rate selection since it is one of the critical hyperparameters. For instance, Leslie N. Smith et al. [18] described a technique for selecting a series of learning rates for NNs. The method trains the network from a low learning rate, increases the learning rate in batches, and then records each batch's learning rate and training loss. It was found that in the case of a low learning rate, the loss reduces, the training accelerates until the learning rate becomes too large and the loss increases. The learning rate continuously varies between these bounds until the optimum results are achieved.

3) *Stochastic Gradient Descent*

Stochastic Gradient Descent (SGD) is an iterative learning algorithm that uses a training dataset to update the model. An optimisation algorithm can train machine learning algorithms in DL. The algorithm's task is to determine a set of internal model parameters that perform well on specific performance indicators such as logarithmic loss or mean square error.

Note that the gradient refers to the calculation of the error gradient or error slope, and the descent refers to the decline along the slope to a certain minimum error level. Once the search process occurs in multiple discrete steps, the algorithm iterates, each step slightly improves the model parameters. SGD is an optimisation algorithm that uses the training dataset to estimate the error gradient of the model's current state. Then, it uses the backpropagation error algorithm to update the weights of the model.

C. *Data Augmentation*

The machine gains more learning experience by providing more data. For deep learning models, more data is required to building high-performance models. Additionally, the fed data needs to be more diverse. Otherwise, the model might face the problem of over fitting [19]. One of the ways to solve the problem of limited data is to perform different transformations on the available data and synthesise it as a new dataset. The method of synthesising new data from available data is called data augmentation. Data augmentation is the process of increasing the size of database using the original available data.

For example, in this snatch theft detection, the motorcycle in an image may not be properly positioned in all pictures. The image can be located at the left or right view of the motorcycle. The DL model should accurately detect the snatch theft situation without being affected by factors like daylight, night, or any other environmental condition. Data enhancement techniques can be applied to the data generation process after pre-processing and before training. For the test dataset, it can be directly used to test without any conversion. A small dataset can generate image conversion and training models for all data at the same time. At the same time, large datasets can generate unique transformed images for each batch in each period. Simple transformations for data augmentation purposes include flip, rotate, translate, crop, zoom, and adjusting the brightness of the images. [20].

D. *Performance Measure*

The performance of the classifier is evaluated to confirm the effectiveness of the classification stage [21], [22]. The performance of the classifier are measured using the values of accuracy, positive and negative predictive value, error rate, sensitivity/call rate, specificity, precision, and confusion matrix [21]–[26]. The error rate is used to measure classification performance because the learning method can optimise the error rate and will not produce sub-optimal results [26]. However, the error rate is not suitable for a wide range of unbalanced data, if the size of positive and negative data is not equal.

The confusion matrix represents two parameters, specifically matrix rows and matrix columns. This is a performance measure for machine learning classification, where the output can be two or more classes [27]. The matrix rows are the actual categories or benchmarks being tested, and the matrix columns are the target categories or categories of the specific benchmarks. The matrix evaluates the behaviour of multiple classifiers in vision systems [21], [28], [29]. The advantage of the matrix lies in its prior information known as the confusion matrix [21].

Recall that this study is to evaluate snatch theft detection. Each test can classify the occurrence of theft. The test result can be positive that categorises the snatching event, or negative to categorise the occurrence of non-snatching. The test result of each subject may not match the actual condition of the subject. True Positive (TP) implies accurate recognition of 'snatch theft'; False Positive (FP) implies inaccurate recognition of 'snatch theft'. Meanwhile, True Negative (TN) are 'normal' conditions and can be correctly identified as 'normal', and False Negative (FN) represents 'snatch theft' is incorrectly identified as 'normal'.

On the other hand, accuracy (Acc) calculates whether the test correctly identifies each class. The equation of accuracy is as in (1). In addition, sensitivity (Sens) is to determine the actual positive probability as expressed in (2) [30] while specificity (Spec) is a measure to determine the likelihood of actual negative as described in (3) [30].

$$accuracy = \left[\frac{TP + TN}{TP + FP + TN + FN}\right] \times 100 \tag{1}$$

$$sensitivity = \left[\frac{TP}{TP + FN}\right] \times 100 \tag{2}$$

$$specificity = \left[\frac{TN}{TN + FP}\right] \times 100 \tag{3}$$

III. Methodology

This section elaborates on the selection of optimum training parameters for the snatch theft detection model. A step by step on the parameter's selection will be elaborated. Here, the snatch theft datasets are videos attained from the YouTube or Google. Two types of datasets were used for this study, known as 'Normal' or 'Anomaly', to represent the occurrence of snatch theft. Two default hyperparameters used are epoch and learning rates, as in Table 1.

TABLE 1
Default Value For Training Options

Training Option	Value
Epoch	30
Learning Rates	0.001

A total of 2,000 images equally represents the two categories are used as the dataset. Here, 70% of the dataset is used as training and the remainder as testing [30] , [31]. Thus, 700 images from each category are used as training and the remaining 300 images as testing or unseen data.

A. Choosing Optimum Learning Rates

Firstly, the model is trained with a value range from 0.0 to 1.0 to select the most suitable learning rate. The learning rate value starts at 1.0 and decreases as follows: 0.1, 0.01, and 0.0001, respectively. The initial learning rate used for training is specified as a positive scalar. Note that low learning rate is requires longer time while higher learning rate might lead to sub-optimal results.

B. Choosing Optimum Number of Epochs

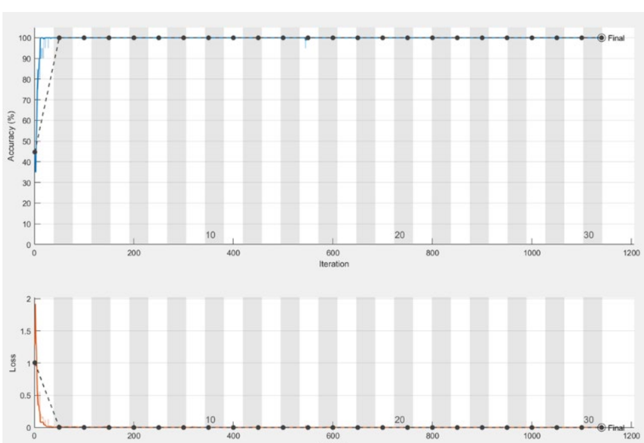


Fig. 2.Training progress epoch = 30 with learning rates of 0.0001

Figure 2 shows the training progress by setting the number of epochs as the default value, 30 and the learning rate at 0.0001. According to [15], Figure 2 depicts over fitting based on the training data used. Therefore, it is suggested to overcome the over fitting by training the data with different epoch values.

C. Data Augmented or Not Augmented

In this study, the accuracy of the proposed model is evaluated and validated using both augmentation and non-augmentation images as database. Figure 3 shows example of original or non-augmented images while Figure 4 depicts some examples of augmented images. The accuracy results of each training are compared.

Fig. 3.Example of normal image


Fig. 4.Example of images with augmentation process

IV. Result & Discussion

This section discusses the experimental analysis conducted, and the results attained. Figure 5 shows the learning rate at 1.0, 0.1, 0.01, 0.001 and 0.0001. Result shows that the training model reaches 50% accuracy with the learning rate set at 1.0, 0.1 and 0.01, while the learning rate at 0.001 and 0.0001 shows that 100% accuracy is achieved. Hence, it proves that 0.001 and 0.0001 are more suitable values as learning rates. The next step is to decide whether 0.001 or 0.0001 is better.

Further, the data is trained with another two different values of the epoch. Figure 6(a) shows that the accuracy reaches 99.0% with 15 as the epoch value, while Figure 6(b) shows a more promising accuracy at 99.5% with an epoch value of 20. Hence in this study, an epoch value of 20 is chosen. Upon completing this numerical analysis, the optimum epoch and learning rate values are 20 and 0.0001, respectively.

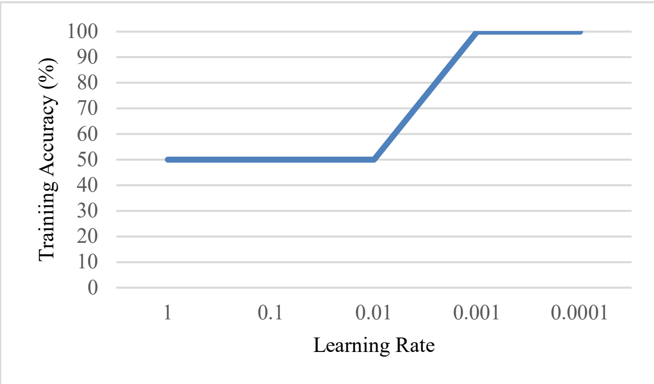
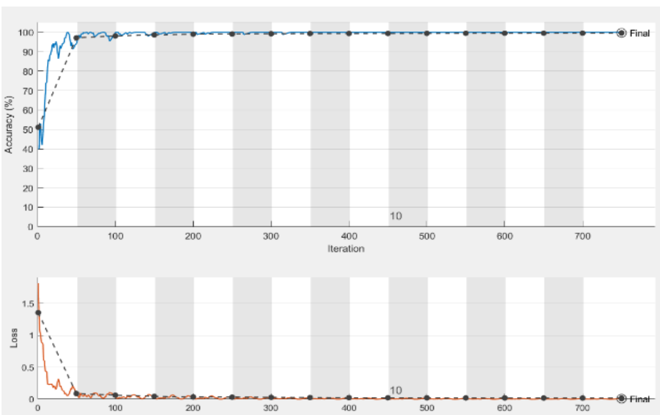
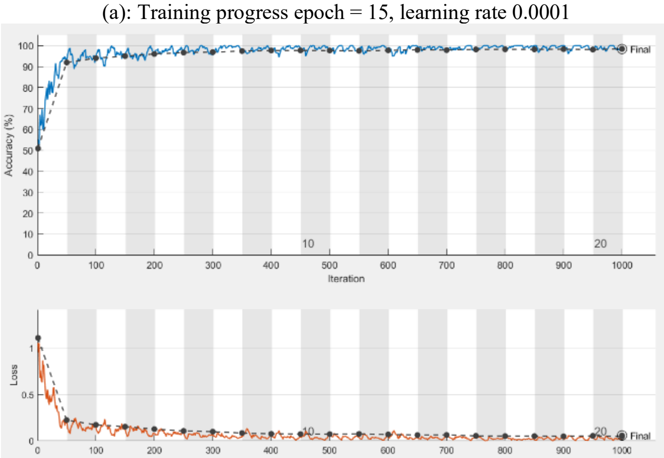


Fig. 5.Learning Rates vs Training Accuracy



(a): Training progress epoch = 15, learning rate 0.0001



(b): Training progress epoch = 20, learning rate 0.0001

Fig. 6.Results of Training based on different epoch and learning rate

Moreover, Figure 7 shows the training results obtained using the augmentation and non-augmentation datasets based on the proposed model. Each accuracy offers an almost similar pattern, precisely 99.87% for non-augmentation and 99.54% for augmented database. Note that some of the captured images are towards the most left or extreme-right view of the motorcycle. Hence, all these factors affected the model during training.

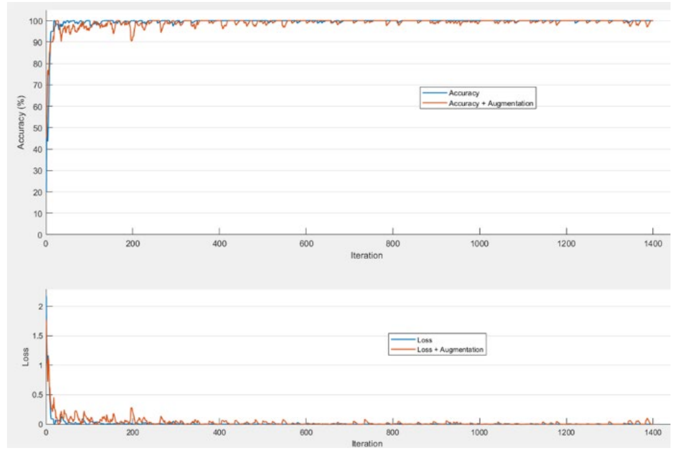


Fig. 7. Training progress of accuracy with and without augmentation

Next, in this study, the hyperparameter values obtained from the training stage are further utilised to classify the testing or unseen dataset to confirm that the proposed model worked well based on each performance measure values. As mentioned earlier, accuracy, specificity and sensitivity are used to evaluate and validate the effectiveness of the proposed model. Table 2 shows the testing results using 20 epochs and the learning rate set at 0.0001. The accuracy obtained during testing is 98.3%, with sensitivity and specificity of 96.7% and 100%, respectively. Here, sensitivity resembles that snatch theft is correctly classified as 'Anomaly' while specificity represents that non-snatch theft is correctly classified as 'Normal'.

TABLE 2
PERFORMANCE MEASURE OF THE SNATCH THEFT DETECTION MODEL

CATEGORY	ACTUAL		
	ANOMALY (Snatch theft)	**NORMAL (Non snatch)**	
ANOMALY (Snatch theft)	48.3% (290)	0.0% (0)	96.7% (Sens)
NORMAL (Non snatch)	1.7% (10)	50.0% (300)	100% (Spec)
	96.7% (Sens)	100% (Spec)	98.3% (Acc)

V. CONCLUSION

In conclusion, this study investigates, evaluates, validates, and establish an approach for selecting the best optimum hyperparameters for the snatch theft detection model. Results show that the learning rates of epoch set as 20 and 0.0001 are the optimum value for the snatch theft DL model using both actual and augmented images as database. The results obtained are 98.3% accuracy, 96.7% sensitivity and 100% specificity. Future work includes utilising more complex and intricate crime scene images as database.

ACKNOWLEDGMENT

This research was funded by the Ministry of Higher Education (MOHE) Malaysia, Grant No: 600-IRMI/FRGS 5/3

(394/2019), Sponsorship File No: FRGS/1/2019/ TK04/UITM/01/3. The authors would like to thank the College of Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia for the facilities provided in this research.

REFERENCES

[1] Y. V. Truntsevsky, I. I. Lukiny, A. V. Sumachev, and A. V. Kopytova, "A smart city is a safe city: The current status of street crime and its victim prevention using a digital application," *MATEC Web Conf.*, vol. 170, 2018, doi: 10.1051/matecconf/201817001067.

[2] S. R. B. Md Sakip and M. N. Moihd Salleh, "Linear Street Pattern in Urban Cities in Malaysia Influence Snatch Theft Crime Activities," *Environ. Proc. J.*, vol. 3, no. 8, p. 189, 2018, doi: 10.21834/e-bpj.v3i8.1386.

[3] R. Latimaha, Z. Bahari, and N. A. Ismail, "Examining the linkages between street crime and selected state economic variables in Malaysia: A panel data analysis," *J. Ekon. Malaysia*, vol. 53, no. 1, pp. 59–72, 2019, doi: 10.17576/JEM-2019-5301-6.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Handb. Approx. Algorithms Metaheuristics*, pp. 1–1432, 2012, doi: 10.1201/9781420010749.

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[6] P. Tzirakis, J. Zhang, and W. Schuller, "END-TO-END SPEECH EMOTION RECOGNITION USING DEEP NEURAL NETWORKS Department of Computing , Imperial College London , London , UK Chair of Embedded Intelligence for Health Care and Wellbeing , University of Augsburg , Germany," *Icassp 2018*, pp. 5089–5093, 2018.

[7] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 4, no. January, pp. 3104–3112, 2014.

[8] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 3156–3164, 2015, doi: 10.1109/CVPR.2015.7298935.

[9] B. Huval *et al.*, "An Empirical Evaluation of Deep Learning on Highway Driving," pp. 1–7, 2015, [Online]. Available: http://arxiv.org/abs/1504.01716.

[10] U. M. Butt, S. Letchmunan, F. H. Hassan, S. Zia, and A. Baqir, "Detecting video surveillance using VGG19 convolutional neural networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, pp. 674–682, 2020, doi: 10.14569/ijacsa.2020.0110285.

[11] K. Goya, X. Zhang, K. Kitayama, and I. Nagayama, "A method for automatic detection of crimes for public security by using motion analysis," *IIH-MSP 2009 - 2009 5th Int. Conf. Intell. Inf. Hiding Multimed. Signal Process.*, pp. 736–741, 2009, doi: 10.1109/IIH-MSP.2009.264.

[12] Sagar Sharma, "Epoch vs Batch Size vs Iterations | by SAGAR SHARMA | Towards Data Science." [Online]. Available: https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9.

[13] Jason Brownlee, "Difference Between a Batch and an Epoch in a Neural Network," *Machine Learning Mastery*, 2018. https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/.

[14] U. Vijay, "What is epoch and How to choose the correct number of epoch | by Upendra Vijay | Medium," *medium*, 2019. https://medium.com/@upendravijay2/what-is-epoch-and-how-to-choose-the-correct-number-of-epoch-d170656adaaf.

[15] "Choose optimal number of epochs to train a neural network in Keras - GeeksforGeeks," *Geeks for Geeks*, 2020. .

[16] J. Brownlee, "Understand the Impact of Learning Rate on Neural Network Performance," *Machine Learning Mastery*. pp. 1–27, 2019, [Online]. Available: https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/.

[17] P. Surmenok, "Estimating an Optimal Learning Rate For a Deep Neural Network," *Toward Data Science*. pp. 1–6, 2017, [Online]. Available: https://towardsdatascience.com/estimating-optimal-learning-rate-for-a-deep-neural-network-ce32f2556ce0.

[18] L. N. Smith, "Cyclical learning rates for training neural networks," *Proc. - 2017 IEEE Winter Conf. Appl. Comput. Vision, WACV 2017*, no. April, pp. 464–472, 2017, doi: 10.1109/WACV.2017.58.

[19] N. Balla, "Five Popular Data Augmentation Techniques In Deep Learning." .

[20] A. K, "What is Data Augmentation & how it works?," *Great Learning Blog*. 2018, [Online]. Available: https://www.mygreatlearning.com/blog/understanding-data-augmentation/.

[21] J. R. Parker, "Rank and response combination from confusion matrix data," *Inf. Fusion*, vol. 2, no. 2, pp. 113–120, 2001, doi: 10.1016/S1566-2535(01)00030-6.

[22] P. Sharma, Y. P. S. Berwal, and W. Ghai, "Performance analysis of deep learning CNN models for disease detection in plants using image segmentation," *Inf. Process. Agric.*, vol. 7, no. 4, pp. 566–574, 2019, doi: 10.1016/j.inpa.2019.11.001.

[23] J. T. Wei, Z. Zhang, S. D. Barnhill, K. R. Madyastha, H. Zhang, and J. E. Oesterling, "Understanding artificial neural networks and exploring their potential applications for the practicing urologist," *Urology*, vol. 52, no. 2, pp. 161–172, 1998, doi: 10.1016/S0090-4295(98)00181-2.

[24] C. Ferri, J. Hernández-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recognit. Lett.*, vol. 30, no. 1, pp. 27–38, 2009, doi: 10.1016/j.patrec.2008.08.010.

[25] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.

[26] T. Joachims, "A Support Vector Method for multivariate performance measures," *ICML 2005 - Proc. 22nd Int. Conf. Mach. Learn.*, pp. 377–384, 2005, doi: 10.1145/1102351.1102399.

[27] S. Narkhede, "Understanding Confusion Matrix," *Nursing times*, vol. 85, no. 39. pp. 35–37, 2018, [Online]. Available: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62.

[28] J. J. Remus, C. S. Throckmorton, and L. M. Collins, "Expediting the identification of impaired channels in cochlear implants via analysis of speech-based confusion matrices," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2193–2204, 2007, doi: 10.1109/TBME.2007.908336.

[29] N. D. Marom, L. Rokach, and A. Shmilovici, "Using the confusion matrix for improving ensemble classifiers," *2010 IEEE 26th Conv. Electr. Electron. Eng. Isr. IEEEI 2010*, pp. 555–559, 2010, doi: 10.1109/EEEI.2010.5662159.

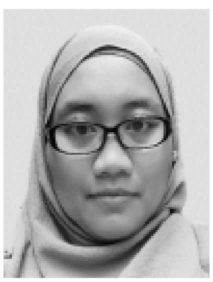
[30] S. Lhessani, "What is the difference between training and test dataset?," *Medium*. 2019.

[31] J. Brownlee, "Train-Test Split for Evaluating Machine Learning Algorithms," *Machine Learning Mastery*. pp. 1–17, 2020.

Nurul Farhana Mohamad Zamri graduated with B.Eng (Honors) in Electronics Engineering from Universiti Teknologi Mara (UiTM), Malaysia. Currently, she is a Research Assistant and Masters student in Electrical Engineering at the School of Electrical Engineering, College of Engineering, UiTM . Her research concentrates in detection of Snatch Theft.

Nooritawati Md Tahir is a lecturer at the College of Engineering, Universiti Teknologi MARA Selangor Malaysia. Her research interest is in the field of Image Processing, Pattern Recognition and Computational Intelligence. She has authored and co-authored more than 250 indexed publications with SCOPUS h-index of 18. In addition, she is a Certified Trainer for Postgraduate Training in Innovation and Entrepreneurship awarded by TCD-UCD Dublin, Ireland, Stanford Faculty Fellow organised by Stanford Technology Ventures Program, Stanford University, US, Chartered Engineer (CEng) from the Institution of Engineering and Technology and Engineering Council UK as well as Senior Member (SMIEEE) of the Institution of Electrical, Electronics Engineer (IEEE).

Megat Syahirul Amin Megat Ali received his B.Eng. (Biomedical) from University of Malaya, Malaysia, M.Sc. in Biomedical Engineering from University of Surrey, United Kingdom, and Ph.D. in Electrical Engineering from Universiti Teknologi MARA, Malaysia. He is a senior lecturer at the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA. His research interests include biomedical signal processing and artificial intelligence. Dr. Megat is currently a research fellow at the Microwave Research Institute, Universiti Teknologi MARA.



Nur Dalila Khirul Ashar received the MSc degree from the Department of Electrical Engineering (Computer Control and Automation), National University of Singapore (NUS), Singapore in 2013. Currently she serves as a senior lecturer since 2003 with the Universiti Teknologi MARA Malaysia in Johor. Her research interest areas are control system, instrumentation space weather and machine learning applications.