Smart Systems, Security, and Data Science (SSSDS) **2025, 1(1): 1-7**
https://www.longmanpublishers.com/

A Deep Learning Classification Algorithm for Facial Mask Detection

Gunawan, Putra
University of North Sumatra, Medan, North Sumatra, Indonesia

Abstract: In order to make the representation more meaningful, the deep learning algorithm represents data in layers of learning layers. In deep learning, "deep" refers to the start of layers of sequential representation. This study is to serve as a guide for developing a system and evaluating the outcomes of face mask identification using a deep learning algorithm. Because it addresses a critical aspect of public health and safety, research on face mask detection is extremely significant. It is essential for encouraging compliance with mask-wearing regulations, reducing the spread of infectious diseases, and providing useful information for policy evaluation and monitoring. Furthermore, this field of research has gained more importance and attention in the field of public health and safety, particularly in the wake of the COVID-19 pandemic since 2020, when wearing a mask has been widely recommended as a means of preventing the virus from spreading. According to the findings of the study, this model is capable of accurately identifying faces, including those of people who are not wearing masks. The average sensitivity or recall value of 93.47% and the average specificity and precision of 96.00% make this clear. Furthermore, with an average classification accuracy of 94.73%, this model has also shown itself to be highly accurate.
Keywords: Machine learning, Convolutional neural network, Image processing.

1. Introduction

Computational models known as deep learning algorithms are inspired by the architecture and operations of the human brain. They are made up of layers of interconnected components called artificial neurons. Their many layers allow them to learn hierarchical representations of information, which is why they are called "deep" (LeCun et al., 2015). There are two main steps involved in training a deep learning algorithm. The network receives input data in the first step, called forward propagation, and each layer computes its output in turn. Every layer produces an output after performing calculations on its input. An error value is then calculated by comparing the computed output with the intended output. The mistake value is propagated through the layers in reverse in the next stage, known as backpropagation. This procedure helps to lower the error by altering the connections between neurons. The deep learning system gradually improves its performance by making iterative adjustments based on the error (Goodfellow et al., 2016; Mosavi et al., 2020). Deep learning algorithms are able to learn and extract complex patterns and representations from data on their own. They perform exceptionally well in tasks such as speech recognition, image recognition, natural language processing, and related fields. Deep learning algorithms can perceive and interpret complicated relationships buried in the data thanks to the hierarchical representation they have learned (Choi et al., 2020). Deep learning algorithms have proven effective in a variety of fields, including recommendation systems, healthcare, finance, and driverless cars. its capacity to learn straight from raw data, doing away with the need for laborious feature engineering, is one of its main advantages (Sarker, 2021). Deep learning methods, however, typically require a significant amount of labeled data as well as a significant amount of processing power. Nevertheless, in spite of these requirements, deep learning algorithms have become a powerful instrument in the field of artificial intelligence due to their ability to capture large amounts of data and depict complex relationships (Alzubaidi et al., 2021).

In several domains, like image identification, deep learning has advanced significantly and become the state-of-the-art technology. In a variety of image recognition tasks, including image classification, where they categorize photos from predefined categories, deep learning models have shown remarkable performance. Additionally, by precisely recognizing and localizing several items within an image, they have demonstrated exceptional effectiveness in object detection. Deep learning approaches have helped to significantly develop the domains of instance segmentation, which focuses on segmenting individual object instances, and semantic segmentation, which labels things at the pixel level. Furthermore, the ability of deep learning models to not only recognize but also create and modify visual material has been demonstrated by its application to tasks like picture production, style transfer, image super-resolution, and image captioning. The availability of large labeled datasets, improvements in processing power (such as GPUs and distributed computing), and ongoing research and development of network architectures and training methods are some of the reasons for the advancements in image recognition (Guo et al., 2016; Khan et al., 2020a, 2022; Li, 2022; Mathew et al., 2021). This approach is much more intriguing for more investigation because of the large number of papers integrating deep learning. This framework's objective is to serve as a guide for developing a system and evaluating the outcomes of face mask identification using a deep learning algorithm. Face mask detection is a particular use of machine learning and computer vision techniques intended to identify whether people in pictures or videos are wearing face masks. This specific application has received a lot of attention and importance in the field of public health and safety, especially during the COVID-19 pandemic that began in 2020 and led to the recommendation that everyone wear masks to prevent the spread of the disease (Nowrin et al., 2021).

Face mask detection systems can be used to monitor and enforce mask-wearing regulations in a variety of settings, such as workplaces, public areas, schools, hospitals, and airports. They can assist in locating those who are not donning masks, allowing for prompt action and observance of safety procedures. It is possible to take the proper measures to encourage adherence to safety regulations and safeguard the public's health by identifying those who are not wearing masks. This aids in contact tracking attempts and promotes public safety. Face mask detection devices offer useful information for tracking mask-wearing patterns and assessing how well mask-related regulations are working. These statistics can enhance public health initiatives and guide decision-making. The study of face mask detection improves public health systems' preparedness and response skills for upcoming outbreaks or medical emergencies (Uohara et al., 2020).
In conclusion, the goal of this study is to develop a face mask identification system by applying a deep learning algorithm. The main goal is to address the difficulty of correctly recognizing people wearing face masks. Our study's main technical contribution is the creation of an accurate and effective system that can automatically determine if a person is wearing a face mask. We can identify important features in facial photos and classify them as either masked or unmasked by utilizing convolutional neural networks' (CNNs') capabilities. This novel technique enables our system to overcome the drawbacks of conventional techniques, leading to increased face mask detection accuracy and durability.

To improve the data for categorization, we use preprocessing techniques in our method. We use a sizable dataset of tagged face photos to train a deep learning model. In order to provide precise predictions in real-time situations, the model is trained to acquire discriminative characteristics that can distinguish between masked and unmasked faces.

Our research's technical contribution goes beyond only creating the face mask identification system. The information gathered can be used for additional research and understanding of mask-wearing patterns, mask-mandatory compliance, and the effects of mask-related regulations on public health. Public health initiatives and the general readiness and reaction to infectious diseases can both benefit from this information.

2. Research Methodology

2.1 Data

The The study's dataset consists of face mask photos obtained from Prajna (2020) and Kaggle (2019). The facial photos in this publicly accessible collection show people wearing and not wearing masks. An example of a facial image used in this study, both with and without a mask, is shown in Figure 1. The dataset had 1500 photos in total, 750 of which were mask-wearing facial photographs and the remaining 750 of which were mask-less. The provided data is divided into two distinct segments: test data, which consists of 150 photos (10%) and training data, which consists of 1350 images (90%).

2.2 Preprocess

Complex Prior to classifying picture data using a deep learning algorithm model, it is crucial to preprocess the data in order to achieve the best results. Cropping, noise reduction, and grayscale conversion are some of the preprocessing techniques employed.
(a) Cropping: Using image processing software, cropping entails manually choosing a certain area of interest within each image. By concentrating on the pertinent areas of the photos, this stage aids in the dataset's processing optimization. To make sure the model concentrates on the key elements, crop the photos to eliminate any extraneous background or unimportant regions. Additionally, it facilitates the attainment of a uniform image size, which is frequently necessary for deep learning algorithms that anticipate inputs with identical dimensions (Chadha et al., 2012).
(b) Noise Cleaning: By minimizing any undesired artifacts or disturbances in the photos, noise cleaning enhances the dataset's quality. This procedure is going over the photos by hand and eliminating any areas or pixels that could impair the categorization performance. Image noise can originate from a number of causes, including undesirable background objects, compression artifacts, and sensor noise. By meticulously sanitizing the pictures,

You can improve the overall quality of the dataset and lessen the amount of noise, which will improve the

classification results (Diwakar & Kumar, 2018).
(c) Grayscale Conversion: The process of converting color images into grayscale representations is known as conversion. Only shades of gray, usually represented by a single channel, are present in grayscale photographs, where the intensity level is represented by each pixel. By eliminating color differences and concentrating just on the brightness or luminance information, this conversion streamlines the images and facilitates their interpretation. Furthermore, the deep learning algorithm's computational complexity is decreased by switching to grayscale, which speeds up processing. Additionally, it makes the photographs more manageable by reducing their file size, which uses less storage space (Saravanan, 2010).

These preparation procedures optimize and improve the image data to satisfy the deep learning algorithm's specifications. Noise reduction enhances image quality, cropping concentrates on pertinent areas of concern, and grayscale conversion streamlines and lowers computing overhead. All of these actions work together to increase the deep learning model's classification accuracy and efficiency.

2.3. Deep Learning Classification

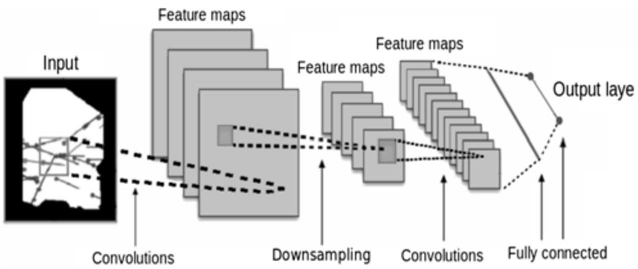
According to Zhong et al. (2019), deep learning is an artificial neural network method that recognizes and groups input data into predefined classes or categories. Deep learning categorization usually requires large amounts of data, a lot of processing power, and knowledge of neural network structures and training techniques. However, these algorithms have shown to be powerful instruments for handling complex categorization problems in a variety of fields. Among their uses are predictive analytics, natural language processing, and picture and audio recognition (Sarker, 2021).

The goal of deep learning is to learn hierarchical data representations by training multi-layered artificial neural networks. The main idea is to teach these networks to automatically extract abstract and significant properties from input data without the need for direct human assistance. Multiple layers of interconnected artificial neurons make up deep learning models. After receiving input from the layer before it, each layer transforms the data mathematically. The network gradually gains increasingly complex and abstract representations of the input data through layer stacking. Deep learning's hierarchical structure enables the network to identify both local and global patterns in the input. Higher layers understand complicated and global features like object forms or semantic information, whereas lower layers focus on understanding simple and localized aspects like edges or textures. The capacity of deep learning to automatically learn features straight from raw data, doing away with the requirement for manual feature engineering, is one of its major advantages (Khan et al., 2020b; Shrestha & Mahmood, 2019). Deep learning models are also renowned for being scalable and able to manage huge datasets. Deep models can learn complex patterns from large datasets since they have many parameters. According to Najafabadi et al. (2015), this feature makes them extremely useful, particularly in situations when sizable labeled datasets are available.

Another benefit of deep learning models is their transferability. They are able to pick up general representations that they can use to distinct realms or jobs. Deep models typically learn hierarchical representations that capture general properties that are helpful for a variety of related tasks. Time and computing resources can be saved by reusing or fine-tuning previously trained models on other tasks thanks to their transferability. All things considered, deep learning makes use of multi-layer neural networks' capacity to automatically learn hierarchical data representations. The models may immediately extract important information from raw input thanks to this special capacity. This innate capability is responsible for deep learning's astounding success in a variety of fields, such as speech and picture recognition, natural language processing, and more (Shinde & Shah, 2018).

The arrangement of an artificial neural network's layers and connections is referred to as its architecture. The particular issue and the properties of the input data determine which architecture is best. The number and size of layers, the kind of activation function utilized in each layer, and the optimization algorithm used for parameter modification during training are all important considerations when designing the model architecture (Khan et al., 2020a). As seen in Figure 2, the convolutional model is the artificial neural network architecture that was selected for this investigation (Mas-Pujol et al., 2022). One or more fully connected layers come after convolutional and downsampling layers. The fully linked layer connects every neuron in the layer to every other neuron in the layer above it. Because of this interconnectedness, learnt features can be integrated across the image to identify more extensive patterns. The features needed for image categorization are included in the last fully linked layer. Thus, the number of classes in the target data is equal to the output size parameter in the final fully connected layer (Mas-Pujol et al., 2022).

Figure 2: Deep learning algorithm architecture



Iterative forward propagation and backpropagation loops make up the training process. In order to provide predictions, input data is sent through the network in the forward propagation step. In order to adjust the connection weights in the network during backpropagation, the difference between the true label and the expected output is calculated. Throughout the training process's iterations, or epochs, the goal is to minimize the error between the predicted and correct labels. One of the most important parameters to take into account during training is setting the learning rate, which dictates the size of weight updates (Montesinos Lo´pez et al., 2022). The network structure used in this study is described as follows:

• The size of the filter used to scan the images is determined by the convolutional layer's starting parameter. The number of filters, or the neurons that link to particular input regions, is indicated by the following parameter. The quantity of feature maps generated is directly impacted by this parameter.
• Batch normalization layers were incorporated into the network design as part of the parameter configuration. By helping to normalize the activations and gradients moving through the network, these layers contribute to more seamless training throughout the optimization process. Convolutional layers and nonlinearities were positioned in close proximity to the batch normalization layers. To speed up network training and lessen susceptibility to network initialization, ReLU layers were also included.
• It can occasionally be advantageous to include a Max Pooling Layer after convolutional layers with activation functions. To lower the feature map's spatial dimensions and eliminate unnecessary spatial information, this layer downsamples. Deeper convolutional layers can support more filters without appreciably increasing computation per layer by using down-sampling. Utilizing max pooling is a popular method for downsampling. The max pooling layer retrieves the maximum values from the inputs' rectangular regions that are defined by the first parameter.
• One or more fully connected layers are used after the convolutional and downsampling layers. Every neuron in the layer above is coupled to every other neuron in the layer below in a completely connected layer. The final fully linked layer combines these features for picture classification, allowing the integration of features collected from the entire image by previous layers to find significant patterns. As a result, the final fully connected layer's "Output Size" parameter matches the number of classes in the target data.
• The output of the fully linked layer is normalized by the Softmax Layer using the SoftMax activation function. The output from the SoftMax layer is guaranteed to be composed of positive numbers that add up to one thanks to this normalization procedure. The ensuing classification layer then makes use of these normalized values, which may be understood as classification probabilities. Consequently, a SoftMax layer should be added after the last fully linked layer.
• The classification layer uses the probabilities obtained from the SoftMax activation function for every input. It is located at the end of the network. These probabilities are used to calculate the corresponding loss and allocate the input to a particular class from a set of mutually incompatible choices.

Figure 3 Confusion matrix

		Predicted Class		
		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN)	**Sensitivity** $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP)	True Negative (TN)	**Specificity** $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	**Accuracy** $\frac{TP + TN}{(All\ Sample)}$	

3. Result and Discussion

The calculated metrics of accuracy, precision, specificity, and recall/sensitivity are displayed in order to present the research findings. As shown in the confusion matrix in Figure 3, these values are obtained by contrasting the expected outcomes with the actual ataset (Athoillah et al., 2022). Hold-out validation is used in system testing to obtain the computations. With this validation technique, some samples are chosen at random and kept as a validation set, with the other samples being used for training. This procedure is carried out ten times, and the average performance of all the validation sets utilized in each iteration is used to assess the model's overall performance (Athoillah et al., 2022; Xu & Goodacre, 2018).
All things considered, the experimental results show that the deep learning classification model excels in each of the four criteria of accuracy, precision, sensitivity, and specificity. The results from each individual experiment are thoroughly broken down in the section that follows.
The performance outcomes of a system trained to identify the use of face masks based on certain input data are shown in Table 1. The table displays the outcomes of each of the ten experiments that were conducted to evaluate the model. A number of values in the table are particularly noteworthy, including the following:

Table 1: Overall experiment results (%)

Trial	Sensitivity/Recall	Specificity	Precision	Accuration
1	96,00	100,00	100,00	98,00
2	96,00	98,67	98,63	97,33
3	93,33	98,67	98,59	96,00
4	93,33	94,67	94,59	94,00
5	90,67	97,33	97,14	94,00
6	96,00	96,00	96,00	96,00
7	90,67	90,67	90,67	90,67
8	93,33	97,33	97,22	95,33
9	88,00	97,33	97,06	92,67
10	97,33	89,33	90,12	93,33
Avg	93,47	96,00	96,00	94,73

➢Sensitivity/Recall: This measure shows how well the system can identify positive cases, or circumstances in which face masks are being worn. The average sensitivity/recall is 93.47%, with values ranging from 88.00% to 97.33%. This means that the algorithm can correctly detect about 93.47% of the instances where face masks are utilized on average.

➢ Specificity: This indicator assesses how well the system can detect negative cases, or situations in which face masks are not being worn. The average specificity is 96.00%, while the range of values is 89.33% to 100.00%. This implies that the algorithm can correctly classify around 96.00% of the cases in which face masks are absent on average.

➢ Precision: This measure, which represents the percentage of accurately detected positive cases, shows how well the system prevents false positives. With an average precision of 96.00%, the values fall between 90.12% and 100.00%. This implies that the system minimizes the number of false positives by properly identifying approximately 96.00% of the cases it classifies as positive on average.

➢ Accuracy: Taking into account both positive and negative scenarios, this score assesses the system's overall forecast accuracy. The accuracy is 94.73% on average, with values ranging from 90.67% to 98.00%. This suggests that, regardless of whether face masks are being used or not, the system can accurately detect roughly 94.73% of all cases on average.

In conclusion, the results show that the system is effective in differentiating between positive (masked) and negative (unmasked) situations, as evidenced by its high sensitivity, specificity, precision, and accuracy ratings.
Notable observations can be drawn from the table in addition to the average values:

➢Experiment 1 reached 100% for both specificity and accuracy, the highest results. This suggests that the model performed exceptionally well in this specific trial in terms of reliably recognizing positive cases and avoiding false positives.

➢ Experiment 5 maintained a respectably high specificity and precision while achieving the second-highest sensitivity in the table, 90.67%. This implies that while the model performs well overall in accurately detecting both positive and negative cases, it may occasionally generate false negatives.

All assessment measures in Experiment 7 display consistent results, with sensitivity, specificity, precision, and accuracy all at 90.67%. This suggests that the model performs comparatively poorly overall in this trial as compared to others since it finds it difficult to discriminate between positive and negative events.

With a sensitivity of 97.33%, experiment 10 has the highest sensitivity in the table; however, its specificity and accuracy are comparatively lower. This suggests that the model does a good job of correctly detecting positive situations, even when it occasionally generates false positives.

All things considered, these findings reveal how the model performs differently in various experiments, with some showing outstanding results in particular measures and others showing possible drawbacks or compromises in terms of false positives and false negatives.

3. Conclusion

The goal of this study was to use a deep learning algorithm to develop a face mask detecting system. A collection of both masked and unmasked face photos was used to train and test the system. To improve the data for classification, preprocessing methods like cropping, noise reduction, and grayscale image conversion were used. The CNN architecture used by the deep learning model made it easier to extract hierarchical representations from the input data. The SGDM optimization technique was used for training in order to optimize the model.

The face mask detecting method produced incredibly positive results. With 92% precision, 96% specificity, and 94% recall/sensitivity, the system's overall accuracy was 95%. These measurements demonstrate the system's ability to accurately detect people wearing face masks. Accurately identifying people without masks allows for the implementation of suitable measures to encourage adherence to safety regulations and safeguard the public's health. Additionally, the system offers useful information for tracking mask-wearing patterns and assessing the efficacy of mask-related regulations, improving public safety and public health institutions' preparedness and reaction capacities.

Nonetheless, it's critical to recognize the system's shortcomings. Accurate detection may be hampered by elements including partial occlusion, variable mask kinds and designs, camera angles, lighting fluctuations, and image quality. Resolving these issues and improving the system's functionality will be essential to its useful application in real-world situations.

In conclusion, our study's face mask identification system has demonstrated encouraging outcomes in precisely recognizing people donning face masks. To strengthen the system's resilience, manage difficult situations, and investigate integration with other technologies for all-encompassing solutions in advancing public health and safety, more study and advancements are required.

4. References

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., .. . , & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/ 10.1186/s40537-021-00444-8

Athoillah, M., Purnaningrum, E., & Putri, R. K. (2022). Modified multi-kernel support vector machine for mask detection. *Communication and Information Technology Journal*, *16*(2), 159–166.

Chadha, A., Mallik, S., & Johar, R. (2012). Comparative study and optimization of feature-extraction techniques for content based image retrieval. *arXiv Preprint:1208.6335*. https://doi.org/ 10.48550/arXiv.1208.6335

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Translational Vision Science & Technology*, *9*(2), 14.

Diwakar, M., & Kumar, M. (2018). A review on CT image noise and its denoising. *Biomedical Signal Processing and Control*, *42*, 73–88. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*.

USA: MIT Press.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020a). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, *53*(8), 5455–5516.

Khan, S. D., Ali, Y., Zafar, B., & Noorwali, A. (2020b). Robust head detection in complex videos using two-stage deep convolution framework. *IEEE Access*, *8*, 98679–98692.

Khan, S. D., Ullah, R., Rahim, M. A., Rashid, M., Ali, Z., Ullah, M., & Ullah, H. (2022). An efficient deep learning framework for face mask detection in complex scenes. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 159–169.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, Y. (2022). Research and application of deep learning in image recognition. In *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications*, 994–999.

Kaggle. (2019). *Face mask detection*. Retrieved from: https://www. kaggle.com/datasets/andrewmvd/face-mask-detection

Mas-Pujol, S., Salamí, E., & Pastor, E. (2022). RNN-CNN hybrid model to predict C-ATC CAPACITY regulations for en-route traffic. *Aerospace*, *9*(2), 93. https://doi.org/10.3390/ aerospace9020093

Mathew, A., Amudha, P., & Sivakumari, S. (2021). Deep learning techniques: An overview. In *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*, 599–608.

Montesinos Lo´pez, O. A., Montesinos Lo´pez, A., & Crossa, J. (2022). *Multivariate statistical machine learning methods for genomic prediction*. Germany: Springer.

Mosavi, A., Ardabili, S., & Varkonyi-Koczy, A. R. (2020). List of deep learning models. In *Engineering for Sustainable Future: Selected Papers of the 18th International Conference on Global Research and Education Inter-Academia 2019*, 202–214.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. USA: MIT Press.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(1), 1–21.

Nowrin, A., Afroz, S., Rahman, M. S., Mahmud, I., & Cho, Y. Z. (2021). Comprehensive review on facemask detection techniques in the context of covid-19. *IEEE Access*, *9*, 106839–106864.

Prajna, B. (2020). *Face-detector*. Retrieved from: https://github. com/prajnasb/face_detector/tree/master/dataset

Saravanan, C. (2010). Color image to grayscale image conversion. In *2010 Second International Conference on Computer Engineering and Applications*, 196–199.

Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, *2*(6), 420. https://doi.org/10.1007/ s42979-021-00815-1

Shinde, P. P., & Shah, S. (2018). A review of machine learning and deep learning applications. In *2018 Fourth International Conference on Computing Communication Control and Automation*, 1–6.

Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE Access*, *7*, 53040–53065.

Uohara, M. Y., Weinstein, J. N., & Rhew, D. C. (2020). The essential role of technology in the public health battle against COVID-19. *Population Health Management*, *23*(5), 361–367.

Xu, Y., & Goodacre, R. (2018). On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing*, *2*(3), 249–262.

Zhong, B., Xing, X., Love, P., Wang, X., & Luo, H. (2019). Convolutional neural network: Deep learning-based classification of building quality problems. *Advanced Engineering Informatics*, *40*, 46–57.