Sensors and Materials, Vol. 38, No. 9 (2026) 5175–5184
MYU Tokyo

S & M 4616

Assessment of Muscle Quality Using a Myophonogram Measurement System Combined with Deep Learning Algorithms

Tian-Hsiang Huang[1] and Wen-Hsien Ho[2,3,4,5,*]

[1]Department of Computer Science and Information Engineering,
National Penghu University of Science and Technology, Penghu 880011, Taiwan
[2]Department of Healthcare Administration and Medical Informatics,
Kaohsiung Medical University, Kaohsiung 807378, Taiwan
[3]Department of Medical Research, Kaohsiung Medical University Hospital,
Kaohsiung 807378, Taiwan
[4]Precision Sports Medicine and Health Promotion Center, Kaohsiung Medical University,
Kaohsiung 807378, Taiwan
[5]College of Professional Studies, National Pingtung University of Science and Technology,
Pingtung 912301, Taiwan

(Received March 4, 2026; accepted June 9, 2026)

Keywords: muscle quality assessment device, myophonogram, deep learning algorithm

Muscle atrophy occurs not only after prolonged postoperative recovery but also as a natural consequence of aging, with it being especially common in individuals with low early-life physical activity. Early-stage muscle loss is often overlooked, and the condition may already be severe by the time clinical signs emerge. Current muscle quality assessments substantially rely on physicians' subjective judgment, and objective measures of such quality are lacking. To address this limitation, in this study, we developed a myophonogram measurement system integrated with deep learning algorithms for the noninvasive assessment of muscle quality. This system captures low-frequency myophonogram signals (time-domain signals) from the target muscle, applies the short-time Fourier transform (STFT) to these signals to obtain spectrograms (time–frequency-domain signals), and then inputs these spectrograms to six deep learning models (LeNet-5, AlexNet, VGG-16, ResNet-50, GoogLeNet, and DenseNet-121) to classify muscle quality as healthy or unhealthy. The system can complete the entire process from signal capture to muscle quality detection within 60 s. Each model used in this system achieved an accuracy above 90% (compared with the evaluations of expert clinicians), demonstrating that the system is an intelligent, rapid, and objective tool for the convenient and reliable self-monitoring of muscle health.

1. Introduction

As public awareness regarding health and recreational activities has increased, exercise has become a crucial aspect of daily life for many people. However, improper or excessive exercise has led to a growing incidence of sports-related injuries, with ankle and knee injuries being

*Corresponding author: e-mail: whho@kmu.edu.tw
https://doi.org/10.18494/SAM6322

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

among the most common sports-related lower-limb injuries. Severe cases of lower-limb injuries, such as those involving ankle fractures or anterior cruciate ligament tears, may require surgical intervention, which often leads to the postoperative complication of muscle atrophy.(1–3)

The accurate assessment of muscle atrophy typically requires costly imaging techniques, such as magnetic resonance imaging (MRI), computed tomography (CT), and dual-energy X-ray absorptiometry. Although MRI and CT are considered gold standards for detecting sarcopenia,(4) their high cost limits their accessibility to the general population.(5,6) Moreover, these devices are not portable and require specialized facilities, imposing temporal and spatial constraints on muscle evaluation.

To address the aforementioned limitations, we developed a myophonogram measurement system integrated with deep learning algorithms for the noninvasive evaluation of muscle quality. This system first generates sound waves and then transfers them to the target muscle. After the sound waves are transmitted through and absorbed by the target muscle, they are received by a microphone for analysis. Energy attenuation occurs during wave transmission, which is mitigated by using an acoustic coupler to ensure maximal signal reception. After the system acquires the myophonogram signals (time-domain signals) transmitted through the target muscle, it converts these signals into spectrograms (time–frequency-domain signals) by using the short-time Fourier transform. Subsequently, these spectrograms are input to trained deep learning models, which analyze the spectrograms, extract crucial features, and then classify the muscle quality as healthy or unhealthy. The developed system can complete the entire process from signal capture to muscle quality detection within 60 s. The system can obtain real-time quantitative data, rapidly monitor muscle changes, and identify muscle quality; thus, it has strong potential for supporting rehabilitation and long-term health management. Figure 1 shows an image of this device.

Changes in muscle quantity, size, and elasticity after physical training result in detectable variations in myophonogram signals.(7) These variations can be translated into corresponding measures of muscle mass, strength, and endurance, enabling laboratory-grade rehabilitation assessment with a low-cost device. Huang *et al.*,(8) Chen *et al.*,(9) and Aslan and Akin(10) used the short-time Fourier transform (STFT) to convert audio signals into time–frequency spectrograms, which they then used as input for deep learning models.

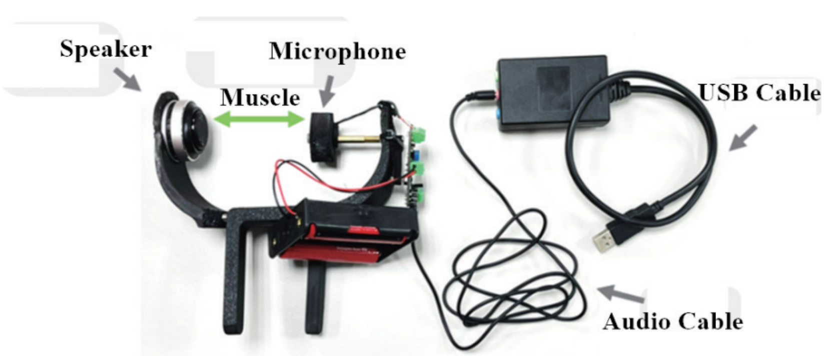

Fig. 1. (Color online) Developed muscle quality assessment device.

Deep learning has progressed substantially in terms of medical and acoustic applications, particularly with respect to image processing, speech analysis, and signal processing.[11,12] Deep learning methods can be applied to automatically extract high-level features from large-scale data, reducing reliance on manual feature engineering and enhancing model classification and prediction capabilities compared with traditional machine learning methods.

Dhar *et al.*[13] proposed a phonocardiogram (PCG) classification method based on the cross-wavelet transform and AlexNet, achieving accurate and automated PCG diagnosis. Nafiz *et al.*[14] used six convolutional neural network (CNN) architectures to classify cough sounds associated with COVID-19 in the Virufy cough dataset, converting audio signals into mel spectrograms with sizes of 32 × 32, 128 × 128, 227 × 227, and the default CNN input size. Chen *et al.*[15] investigated the use of deep learning approaches for classifying breast ultrasound images and then developed a GoogLeNet-based diagnostic method, in which feature extraction and transfer learning are employed to improve the accuracy of classifying benign and malignant tumors.

Seibold *et al.*[16] developed an acoustic sensing and deep learning approach for detecting drill breakthroughs in orthopedic surgery. They employed the STFT and log mel spectrograms for acoustic feature extraction and used ResNet-18 with focal loss optimization. Their approach achieved an accuracy of 93.64% in automated breakthrough detection, thus exhibiting the potential to support real-time error prevention in surgery. Almanifi *et al.*[17] used the PASCAL CHSC 2011 PCG database and visualized heart sound data with spectrograms. They conducted transfer learning with ResNet-50 for automatic murmur detection, using spectrograms and mel-frequency cepstral coefficients as input features. ResNet-50 achieved the highest accuracy (87.65%) with spectrogram input, outperforming VGG-16 and VGG-19 (accuracy = 83.95%), demonstrating that spectrograms are more suitable input features for CNNs.

In line with the aforementioned studies, we converted myophonogram signals into time–frequency spectrograms, which are essentially images and can be considered visual data in analyses. We then used six deep learning models, namely, LeNet-5, AlexNet, VGG-16, ResNet-50, GoogLeNet, and DenseNet-121, to analyze and classify these spectrograms as representing healthy or unhealthy muscles. The overall research workflow is illustrated in Fig. 2.

2. Materials

The data used to train and validate the adopted deep learning models were obtained from five community care centers affiliated with E-Da Hospital Medical Center in Taiwan (IRB No.: EMRP23112N). Myophonogram data and corresponding muscle quality assessment results were obtained from 116 participants aged 65 years and above. An audio sampling frequency of 44100 Hz was adopted in this study. A total of 100 recordings were used for training and validating the deep learning models. On the basis of professional clinicians' evaluations, the collected data were categorized into groups representing healthy and unhealthy muscle quality. A total of 50 audio recordings with a duration of 57.2 s each were obtained for each group. A total of 80% of the collected data were used for model training, whereas the remaining 20% of the data were used for validation.

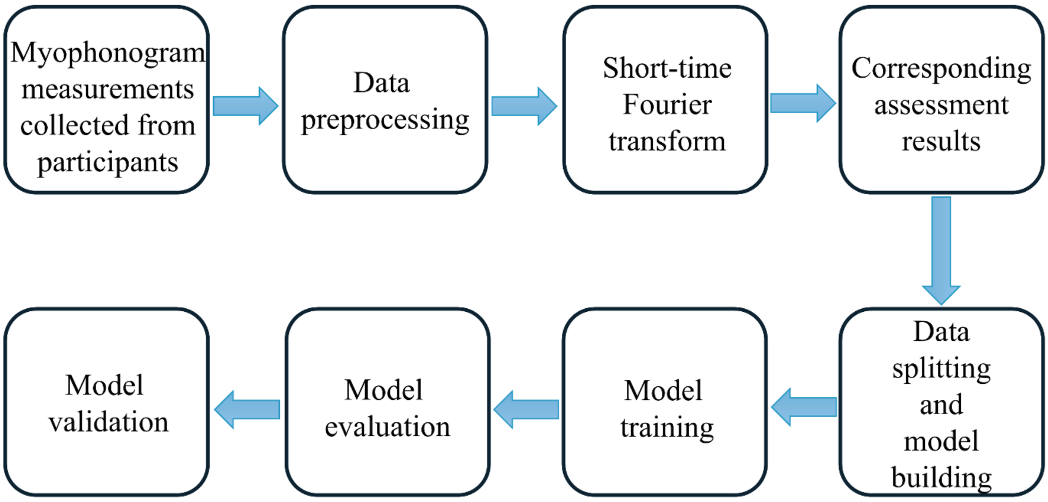

Fig. 2. (Color online) Schematic of research workflow.

To preserve detailed sound and frequency characteristics, each audio frame was set to have a length of 2 s, with a 50% overlap between frames (Fig. 3). This approach increased the effective sample size while maintaining signal continuity. Each audio recording was segmented into 56 audio clips, which were then converted into corresponding time–frequency spectrograms by using the STFT. Consequently, 2800 spectrograms were generated for each group, and these spectrograms visually represented the temporal and frequency characteristics of the collected audio signals.

The total number of audio frames was calculated as

$$Number\ of\ \ frames = \frac{Audio\,length - Frame\,size}{Overlap} + 1\,, \tag{1}$$

where *Number of frames* denotes the total number of audio frames, *Audio length* is the duration of the audio file, *Frame size* represents the length of each frame, and *Overlap* refers to the duration of overlap between consecutive frames.

Figures 4 and 5 present sample time–frequency spectrograms of myophonogram signals obtained for healthy and unhealthy muscles, respectively, after STFT conversion. The horizontal axis represents time (in seconds), the vertical axis represents frequency (in Hertz), and the color indicates signal intensity (in decibels).

To evaluate the adopted models' generalization capability and performance on unseen data, an external testing dataset was collected from 16 participants who were not included in the training or validation set. Each audio recording from these participants was segmented into 56 audio clips, which were sequentially input into the deep learning models for classification. A majority voting scheme was then applied. If more than 50% of the 56 predictions indicated unhealthy muscle quality, the participant was classified as having unhealthy muscle quality; otherwise, they were classified as having healthy muscle quality.

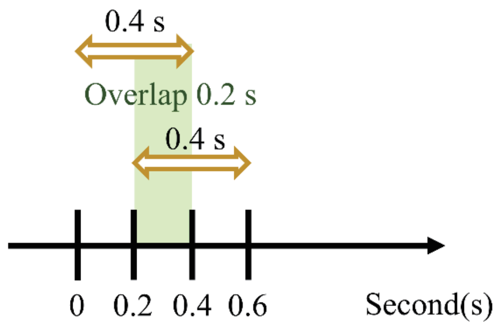

Fig. 3. (Color online) Schematic of audio frame segmentation.

Fig. 4. (Color online) Sample time–frequency spectrogram of myophonogram signals for healthy muscle quality.

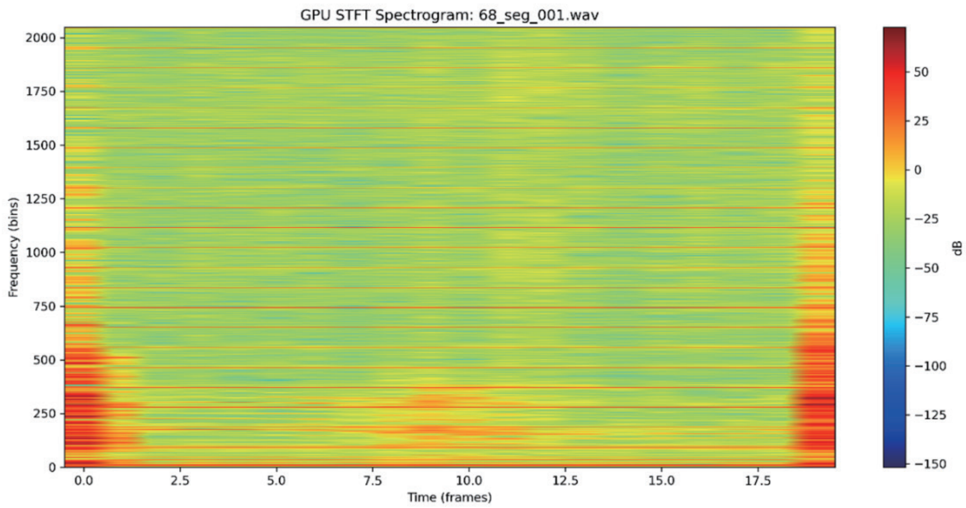

Fig. 5. (Color online) Sample time–frequency spectrogram of myophonogram signals for unhealthy muscle quality.

3. Methods

Because spectrograms can be considered visual data, we adopted CNN-based models for learning and classification. After preliminary experimentation with various CNN architectures, six deep learning models, namely, LeNet-5, AlexNet, VGG-16, ResNet-50, GoogLeNet, and DenseNet-121, were selected for model training and comparison. These six models represent a wide range of architectures, from early classical designs to modern deep residual structures.

LeNet, which was proposed by LeCun *et al.* in 1998, is one of the earliest CNNs and was originally used for recognizing handwritten digits.[18] This model combines convolutional, pooling, and fully connected layers to automatically extract image features and perform classification.[18] Several LeNet models have been developed, including LeNet-1, LeNet-4, and LeNet-5; in this study, we selected LeNet-5.

AlexNet, which was developed by Krizhevsky *et al.* in 2012, exhibited excellent performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012, reducing the classification error rate from 26 to 15% and sparking a deep learning revolution in computer vision.[19]

The VGG model, which was introduced by Simonyan and Zisserman in 2014, is a deep CNN architecture with 3 × 3 convolutional filters and max-pooling layers.[20] This model has two commonly used variants, namely, VGG-16 and VGG-19, which contain 16 and 19 layers, respectively.[20]

ResNet, which was developed by He *et al.* in 2015, won the ILSVRC 2015 competition.[21] It includes residual blocks with skip or shortcut connections, enabling feature maps to bypass certain layers. This residual mapping mechanism effectively mitigates the vanishing gradient problem, enabling the successful training of CNNs with more than 100 layers.

GoogLeNet (Inception V1), which was proposed by Szegedy *et al.* in 2014, won the ILSVRC 2014 competition. Its key innovation lies in the Inception module, which substantially improves CNN performance while reducing the parameter count.[22]

DenseNet, which was introduced by Huang *et al.* in 2017, was inspired by the ResNet architecture.[23] Compared with ResNet, DenseNet has achieved higher performance and parameter efficiency across various image classification tasks on datasets such as CIFAR-10 and ImageNet. The core innovation in DenseNet is the dense connectivity mechanism, which enhances information flow, promotes feature reuse, and mitigates vanishing gradients and parameter redundancy.[23]

In this study, we assessed the performance of the aforementioned deep learning models in terms of training accuracy, validation accuracy, training loss, and validation loss. Loss indicates the deviation between predicted and actual labels, directly reflecting model prediction error, and it was used to guide parameter updates during backpropagation.

We also employed 10-fold cross-validation to assess model stability and generalization, effectively reducing overfitting risk resulting from fixed data partitions and improving model robustness across different data distributions. The evaluation metrics used in this cross-validation process were accuracy (the proportion of correct classifications), precision (the proportion of correct positive predictions), recall (the proportion of correctly identified positive cases), and F1 score (the harmonic mean of precision and recall). These metrics are defined as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{2}$$

$$precision = \frac{TP}{TP + FP}, \tag{3}$$

$$recall = \frac{TP}{TP + FN}, \tag{4}$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{5}$$

where *TP*, *TN*, *FP*, and *FN* denote the numbers of true positives, true negatives, false positives, and false negatives, respectively.[24]

4. Results and Discussion

Through their automatic feature learning capabilities, the adopted models captured numerous high-level discriminative features, which enhanced their predictive performance for muscle quality. Table 1 presents the detailed results of model training and validation. These results indicate that all models achieved high accuracy and low loss. Table 2 presents the results of the 10-fold cross-validation, which indicate that all models exhibited high accuracy and strong generalization capability in muscle quality classification.

To evaluate the generalization capability of the models, they were tested on the collected external dataset. All models exhibited strong performance on this external dataset. For example, GoogLeNet correctly classified 15 out of 16 cases, achieving an accuracy of 93.75%. This result indicates the high reliability and stability of GoogLeNet, highlighting its strong potential for use in muscle quality assessment. The required hyperparameter configurations of GoogLeNet are listed in Table 3.

Table 1
Training and validation results obtained for the adopted deep learning models.

Model	Dataset	Accuracy	Loss
LeNet-5	Training	0.98	0.04
	Validation	0.99	0.02
AlexNet	Training	1.00	≈0.00
	Validation	0.97	0.06
VGG-16	Training	0.92	0.23
	Validation	0.89	0.41
ResNet-50	Training	1.00	≈0.00
	Validation	1.00	≈0.00
GoogLeNet	Training	1.00	≈0.00
	Validation	1.00	≈0.00
DenseNet-121	Training	0.97	0.19
	Validation	0.91	0.63

Note: The symbol "≈0.00" indicates that the value obtained is very close to zero.

Table 2
Results of 10-fold cross-validation for the adopted deep learning models.

Model	Accuracy	Precision	Recall	F1 score
LeNet-5	0.81 ± 0.17	0.73 ± 0.28	0.84 ± 0.30	0.78 ± 0.28
AlexNet	0.94 ± 0.01	0.92 ± 0.02	0.97 ± 0.02	0.94 ± 0.01
VGG-16	0.88 ± 0.02	0.84 ± 0.02	0.93 ± 0.02	0.88 ± 0.02
ResNet-50	0.95 ± 0.01	0.93 ± 0.01	0.98 ± 0.01	0.95 ± 0.01
GoogLeNet	1.00 ± 0.00	1.00 ± 0.01	1.00 ± 0.00	1.00 ± 0.00
DenseNet-121	0.98 ± 0.01	0.97 ± 0.02	0.99 ± 0.01	0.98 ± 0.01

Table 3
Required hyperparameter configurations of GoogLeNet.

Convolutional Layer, kernel size 7 × 7, stride size 2, 64 kernels, ReLU activation
Max Pooling Layer, kernel size 3 × 3, stride size 2
Convolutional Layer, kernel size 3 × 3, stride size 1, 192 kernels, ReLU activation
Max Pooling Layer, kernel size 3 × 3, stride size 2
Inception Module, 128 neurons
Inception Module, 192 neurons
Max Pooling Layer, kernel size 3 × 3, stride size 2
Inception Module, 208 neurons
Inception Module, 224 neurons
Max Pooling Layer, kernel size 3 × 3, stride size 2
Global Average Pooling Layer
Dropout layer omitting 50% of neurons
Fully Connected Layer, 128 neurons, ReLU activation
Dropout layer omitting 50% of neurons
Fully Connected Layer, 1 neuron, softmax activation

The aforementioned experimental results indicate that applying deep learning models for the classification of muscle quality is highly feasible. All adopted models achieved classification accuracies exceeding 90% and very low loss values on the test dataset, thus demonstrating excellent classification performance. Among the six models, GoogLeNet achieved the highest accuracy, precision, recall, and F1 score in 10-fold cross-validation. Furthermore, it exhibited an accuracy of 94% on the external dataset, confirming its strong generalization ability and practical value for muscle quality evaluation in clinical applications.

The core of GoogLeNet is the Inception module, which uses convolutional kernels of different sizes in parallel within the same layer.[22] It has multi-scale feature extraction capabilities and can more effectively capture the characteristics of muscle sound signals across different frequency bands. In the time-frequency spectrograms converted from muscle sound signals, the energy distribution representing muscle features has different scales in both time and frequency domains (some features are very localized, while others span wider frequency bands). Therefore, GoogLeNet can simultaneously capture these spatial features at different scales, providing a more complete description of muscle sound characteristics compared with models with single-size convolutional kernels (such as VGG-16).

5. Conclusions

Muscle atrophy is difficult to identify in its initial stages, and if left untreated, it can lead to considerable muscle weakness, which may impair quality of life, increase the risk of falls, and cause adverse health outcomes, including death. To address the limitations of traditional muscle quality measurement methods, we developed a noninvasive muscle quality assessment system. This system converts myophonogram signals into time–frequency spectrograms through the application of the STFT and then uses deep learning models to classify muscle quality automatically on the basis of acoustic muscle features. All deep learning models used in this study (LeNet-5, AlexNet, VGG-16, ResNet-50, GoogLeNet, and DenseNet-121) achieved an accuracy exceeding 90%, with GoogLeNet achieving the best performance under 10-fold cross-validation. These results highlight the strong potential of using the developed system for noninvasive and automated muscle quality analysis.

By integrating acoustic technology with deep learning, the developed system overcomes various challenges associated with conventional muscle assessment methods, such as invasiveness, operational complexity, high knowledge requirements, long examination times, limited state-specific detection, and subjectivity. Moreover, the developed system enables muscle quality assessment within only 60 s. Thus, this system can serve as a fast, objective, and user-friendly diagnostic support tool for clinicians in muscle assessment, minimizing patient discomfort and eliminating the need for extensive medical expertise. It has high potential for adoption in diverse health-care applications, such as sarcopenia prevention in older adults, muscle rehabilitation, sports training, and routine health examinations.

Furthermore, owing to the device's portability, and with further optimization of its mechanical/structural design, it can be easily integrated into a cloud system for home monitoring, allowing fitness enthusiasts to conveniently assess their muscle training progress—all of which are affordable for the average family.

Acknowledgments

This study was supported by the National Science and Technology Council of Taiwan under grant number NSTC 112-2221-E-037-004-MY3 and by Kaohsiung Medical University (KMU-TC115A06-1), Kaohsiung 807, Taiwan.

References

1 M. J. Toth, P. D. Savage, T. B. Voigt, B. M. Anair, J. Y. Bunn, I. B. Smith, T. W. Tourville, M. Blankstein, J. Stevens-Lapsley, and N. J. Nelms: J. Appl. Physiol. **133** (1985) 647. https://doi.org/10.1152/japplphysiol.00323.2022

2 R. L. Mizner, S. C. Petterson, J. E. Stevens, K. Vandenborne, and L. Snyder-Mackler: J. Bone and Joint Surgery **87** (2005) 1047. https://doi.org/10.2106/JBJS.D.01992

3 L. K. Lepley, S. M. Davi, J. P. Burland, and A. S. Lepley: Sports Health **12** (2020) 579. https://doi.org/10.1177/1941738120944256

4 S. Y. Salim, O. Al-Khathiri, P. Tandon, V. E. Baracos, T. A. Churchill, L. M. Warkentin, and R. G. Khadaroo: J. Surg. Res. **256** (2020) 422. https://doi.org/10.1016/j.jss.2020.06.043

5 E. J. R. van Beek, C. Kuhl, Y. Anzai, P. Desmond, R. L. Ehman, Q. Gong, G. Gold, V. Gulani, M. Hall-Craggs, T. Leiner, C. C. T. Lim, J. G. Pipe, S. Reeder, C. Reinhold, M. Smits, D. K. Sodickson, C. Tempany, H. A. Vargas, and M. Wang: J. Magn. Resonan. Imaging **49** (2019) e14. https://doi.org/10.1002/jmri.26211
6 D. Baiguissova, A. Laghi, A. Rakhimbekova, I. Fakhradiyev, A. Mukhamejanova, G. Battalova, S. Tanabayeva, S. Zharmenov, T. Saliev, and G. Kausova: Health Science Reports **6** (2023) e1102. https:// doi.org/10.1002/hsr2.1102
7 M. S. Goldenberg, H. J. Yack, F. J. Cerny, and H. W. Burton: J. Appl. Physiol. **70** (1991) 87. https://doi.org/10.1152/jappl.1991.70.1.87
8 J. Huang, B. Chen, B. Yao, and W. He: IEEE Access **7** (2019) 92871. https://doi.org/10.1109/ACCESS.2019.2928017
9 J. Chen, Z. Guo, X. Xu, L.-b. Zhang, Y. Teng, Y. Chen, M. Woźniak, and W. Wang: IEEE/ACM Trans. Comput. Biol. Bioinf. **21** (2023) 936. https://doi.org/10.1109/TCBB.2023.3247433
10 Z. Aslan and M. Akin: Traitement du Signal **37** (2020) 235. https://doi.org/10.18280/ts.370209
11 M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C.-A. Deledalle: J. Acoustical Society of America **146** (2019) 3590. https://doi.org/10.1121/1.5133944
12 M. Li, Y. Jiang, Y. Zhang, and H. Zhu: Front. Public Health **11** (2023) 1273253. https://doi.org/10.3389/fpubh.2023.1273253
13 P. Dhar, S. Dutta, and V. Mukherjee: Biomed. Signal Process. Control **63** (2021) 102142. https://doi.org/10.1016/j.bspc.2020.102142
14 M. F. Nafiz, D. Kartini, M. R. Faisal, F. Indriani, and T. Hamonangan: J. Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI) **9** (2023) 535. https://doi.org/10.26555/jiteki.v9i3.26374
15 S.-H. Chen, Y.-L. Wu, C.-Y. Pan, L.-Y. Lian, and Q.-C. Su: J. Radiat. Res. Appl. Sci. **16** (2023) 100628. https://doi.org/10.1016/j.jrras.2023.100628
16 M. Seibold, S. Maurer, A. Hoch, P. Zingg, M. Farshad, N. Navab, and P. Fürnstahl: Sci. Rep. **11** (2021) 3993. https://doi.org/10.1038/s41598-021-83506-4
17 O. R. A. Almanifi, A. F. Ab Nasir, M. A. M. Razman, R. M. Musa, and A. P. A. Majeed: Alexandria Eng. J. **61** (2022) 10995. https://doi.org/10.1016/j.aej.2022.04.031
18 Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner: Proc. IEEE **86** (1998) 2278. https://doi.org/10.1109/5.726791
19 A. Krizhevsky, I. Sutskever, and G. E. Hinton: Commun. ACM **60** (2017) 84. https://doi.org/10.1145/3065386
20 K. Simonyan and A. Zisserman: arXiv preprint arXiv:1409.1556v6 (2014). https://doi.org/10.48550/arXiv.1409.1556
21 K. He, X. Zhang, S. Ren, and J. Sun: 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (2016) 770. https://doi.org/10.1109/CVPR.2016.90
22 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich: 2015 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (2015) 1. https://doi.org/10.1109/CVPR.2015.7298594
23 G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger: 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (2017) 4700. https://doi.org/10.1109/CVPR.2017.243
24 M.-K. Chen, C.-L. Chang, C.-J. Lin, and W.-J. Chen: Sens. Mater. **36** (2024) 1605. https://doi.org/10.18494/SAM4718