

Article

Evaluation of Hemodialysis Arteriovenous Bruit by Deep Learning

Keisuke Ota ^{1,*}, Yousuke Nishiura ², Saki Ishihara ², Hihoko Adachi ², Takehisa Yamamoto ² and Takayuki Hamano ³

¹ Department of Nephrology, Gamagori City Hospital, Gamagori 443-8501, Japan

² Department of Medical Engineering, Gamagori City Hospital, Gamagori 443-8501, Japan; kenkenpa221202@gmail.com (Y.N.); 1214isk@gmail.com (S.I.); toseki@city.gamagori.lg.jp (H.A.); yamamoto-takehisa@city.gamagori.lg.jp (T.Y.)

³ Department of Nephrology, Nagoya City University Graduate School of Medical Sciences, Nagoya 467-8601, Japan; hamatea@med.nagoya-cu.ac.jp

* Correspondence: keke20keke20@yahoo.co.jp; Tel.: +81-533-66-2200

Received: 29 July 2020; Accepted: 24 August 2020; Published: 27 August 2020

Abstract: Physical findings of auscultation cannot be quantified at the arteriovenous fistula examination site during daily dialysis treatment. Consequently, minute changes over time cannot be recorded based only on subjective observations. In this study, we sought to supplement the daily arteriovenous fistula consultation for hemodialysis patients by recording the sounds made by the arteriovenous fistula and evaluating the sounds using deep learning methods to provide an objective index. We sampled arteriovenous fistula auscultation sounds (192 kHz, 24 bits) recorded over 1 min from 20 patients. We also extracted arteriovenous fistula sounds for each heartbeat without environmental sound by using a convolutional neural network (CNN) model, which was made by comparing these sound patterns with 5000 environmental sounds. The extracted single-heartbeat arteriovenous fistula sounds were sent to a spectrogram and scored using a CNN learning model with bidirectional long short-term memory, in which the degree of arteriovenous fistula stenosis was assigned to one of five sound types (i.e., normal, hard, high, intermittent, and whistling). After 100 training epochs, the method exhibited an accuracy rate of 70–93%. According to the receiver operating characteristic (ROC) curve, the area under the ROC curves (AUC) was 0.75–0.92. The analysis of arteriovenous fistula sound using deep learning has the potential to be used as an objective index in daily medical care.

Keywords: hemodialysis patient; deep learning; arteriovenous fistula; shunt sound; artificial intelligence; auscultation; convolutional neural network

1. Introduction

To purify the blood of patients undergoing hemodialysis, a connection from a venous blood vessel to a machine is required. Dialysis, which requires the insertion of two needles to draw blood and two to return blood, is performed three times a week to compensate for lost kidney function. Over several years of continuing dialysis, the blood vessel pierced by the needle gradually narrows [1], a condition called stenosis. The vessel is further narrowed because of the turbulence caused by non-physiological blood flow [2–4]; eventually, it may become occluded. Hence, stenosis should be detected, prevented, and treated as early as possible so that occlusion and dialysis delays can be prevented.

The arteriovenous fistula examination is conducted for detection, which involves physical findings by palpation, auscultation, and visual inspection to find the suspected stenosis. If a site

exists where stenosis is suspected, the next step is to perform vascular ultrasound or angiography to ensure stenosis detection. After diagnosing stenosis, puncturing should be avoided, as it would further deteriorate the blood vessel from needle puncturing during dialysis. Specifically, puncturing should be avoided in that area to prevent the degree of stenosis from developing further. In addition, a treatment called percutaneous transluminal angioplasty (PTA), which inflates a stenosis site with a balloon catheter and expands it before occlusion, is performed [5–7]. To detect stenosis and prevent obstruction, daily consultations called arteriovenous fistula management are essential.

In arteriovenous fistula management, a physical examination using palpation, or palpation and auscultation, is used to check for vibration related to proper blood flow, as well as to determine whether the vessel pulsation is smooth and the pulsation rate is inflated. In auscultation, the presence or absence of constriction is determined by using the pitch difference of sound, discontinuity, spread of sound, and the sound source from other parts (e.g., whistle sound) [8]. In the medical field, this descriptive method is subjective and differs depending on the skill level of the observer and the facility. Novices may only be able to provide a simple description on the presence or absence of constriction sounds. Conversely, skilled persons may record abstract abnormal sounds, such as harmonics and intermittent sounds, in their observation records. As a patient is not assessed by the same skilled observer each day, it is not possible to apply the exact skilled assessment technique daily. Moreover, only subjective results can be communicated to the subsequent examiner. Therefore, the results of auscultation alone do not form a definitive diagnosis of stenosis. Stenosis is diagnosed after the results that have numerical certainty are obtained by performing ultrasound and invasive contrast examination, which require time for preparation and examination. Auscultation is inexpensive, simple, and convenient. However, judgment is delayed if the examiner is not skilled, and the result is abstract. Therefore, auscultation remains only a screening test.

Nevertheless, if auscultation is used as an index that can be output as an objective numerical value rather than as a subjective index, and a diagnosis result with a certainty close to that of an ultrasound examination or a contrast examination is obtained, stenosis diagnosis becomes easy in daily medical care. In addition, if the auscultation technique of a skilled person becomes mechanized, stable output results can be obtained every day, minute changes can be digitally captured and compared with past findings, and the diagnosis accuracy will be improved. The skill of an expert can be reproduced with a machine. Achieving that goal was the primary objective of this study.

There have been reports that stenosis may be estimated during auscultation by using arithmetic processing of the sound frequencies generated by arteriovenous fistulas [9–11]. Although it is a mechanical judgment, it is typically only a frequency analysis treble and bass judgment. The arteriovenous fistula sound grasped in the actual medical field is a multidimensional and abstract interpretation. Therefore, the frequency alone is considered not sufficient as a criterion. Conversely, the spectrogram representing the sound source by the frequency density per time interval is assumed to include all the sound elements [12].

To date, objectively digitizing and comparing multidimensional data in familiar environments have been difficult. However, deep learning, which is the basis of image recognition technology and has made remarkable progress in recent years, can be used on personal computers, and, thus, easily comparing multidimensional data is possible. Moreover, it is now feasible to objectively compare spectrograms with audio recordings and images. Accordingly, to evaluate the sound generated by stenosis, we applied a deep learning method using multidimensional data, similar to a spectrogram converted from a arteriovenous fistula sound, and evaluated it as an objective index. The arteriovenous fistula sound learning model is outlined in Figure 1.

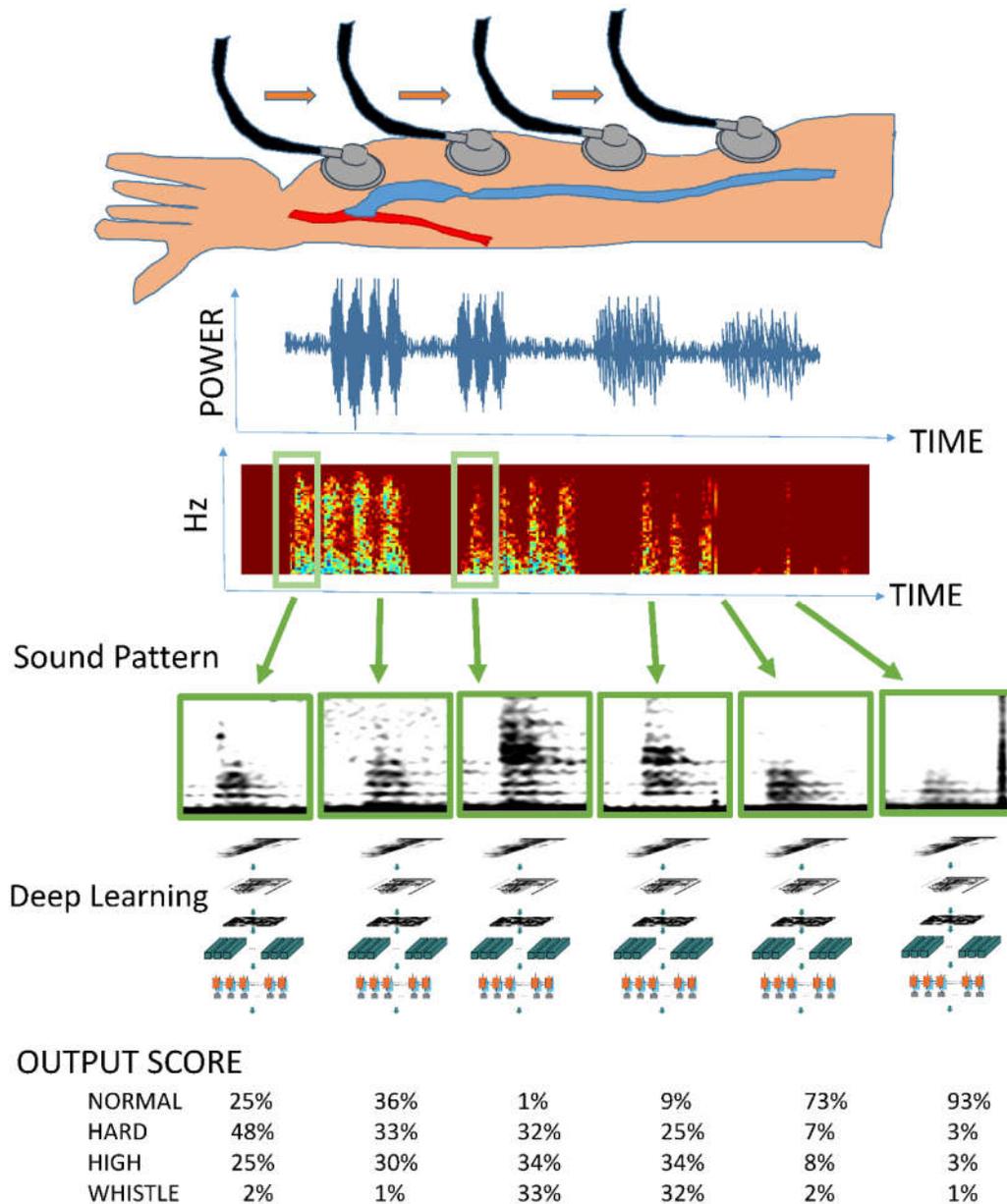


Figure 1. Outline of the arteriovenous fistula sound learning model. A series of arteriovenous fistula sounds are recorded in one medical examination. All continuous tones are converted into spectrograms. A single heartbeat is detected using the mechanism of object detection (R-CNN: regions with convolutional neural networks) similar to that used for face detection in digital cameras and smartphones. The spectrogram of one arteriovenous fistula sound is used as input data. A deep learning model is used for learning, and the importance of the sounds obtained during a general medical examination is expressed by multiclass classification.

2. Materials and Methods

The objective comparison of human consultations supplemented with deep learning requires audio monitoring, which we performed using a DR-100MK TASCAM III: 24-bit/192 kHz recording device (TEAC Corporation, Montebello, CA, USA).

2.1. Participants

The study involved 20 inpatient dialysis patients from a dialysis center (Gamagori Municipal Hospital, Gamagori, Japan). The participants were undergoing treatment in a single facility intended for patients with end-stage renal failure who were undergoing dialysis using arteriovenous fistulas with stable hemodynamic autologous blood vessels. Participating patients were hospitalized for various diseases and were continuing with maintenance dialysis. Arteriovenous fistula sounds were recorded before puncture. Table 1 presents the characteristics of the participants.

Table 1. Sociodemographic and clinical characteristics of the 20 participants. Participants included patients between the ages of 49 and 90 with various diseases.

Characteristic	Value	
Age (median (IQR), years)	73 (63–79)	
Gender (n)	Male	13
	Female	7
Dialysis duration (median (IQR), years)	0.01 (0–19)	
Cause of end-stage renal disease		
Diabetes	11	
Glomerulonephritis	5	
Others	4	
Type of arteriovenous fistula		
Radiocephalic AVF at front arm	17	
Radiocephalic AVF at mid forearm	2	
Brachiocephalic AVF at elbow	1	

Abbreviations: AVF, arteriovenous fistula; IQR, interquartile range.

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Institutional Ethics Committee of Gamagori Municipal Hospital (No. 506-2). Patients on catheter dialysis, patients on dialysis with artificial blood vessels, patients with unstable circulatory dynamics, and other patients deemed by the attending physician to be inappropriate for study participation on medical grounds were excluded.

2.2. Methods

2.2.1. Data Preprocessing: Extraction of Single Beat of the Arteriovenous Fistula Sound

The sampled 1-min arteriovenous fistula auscultation sounds also contain variable environmental sounds; therefore, to extract the arteriovenous fistula sound, the value obtained by integrating and averaging the 2–750 Hz region (frequency range characteristic of arteriovenous fistula sound) from each auscultation sound was output [13]. Spline curves were created to reduce the effect of noise, which can cause difficulty in detecting the maximum and minimum values for one beat [14]. To exclude the times during which auscultation was not performed, a convex range falling within a period of 0.5–2 s was estimated to constitute one beat [15].

The sounds determined using a deep learning classifier as being produced by arteriovenous fistulas with a probability exceeding 50% were extracted as the sounds for one arteriovenous fistula beat. Individual arteriovenous fistula sounds recorded for 10,000 beats were classified into one of five categories of sound audible to the human ear (i.e., normal sound, hard sound, high sound, intermittent sound, and whistling). The arteriovenous fistula sounds classified by the human ear had minimal abnormal sounds, including whistles and intermittent sounds. The sound of the whistle sounds like a whistle, which refers to the turbulence in blood flow due to a sharp decrease in the diameter of blood vessels. Intermittent sounds indicate a discontinuity in vascular noise during diastole. It shows a strong obstruction that can cause a complete interruption of blood flow during

diastole. High sounds, hard sounds, and normal sounds are often encountered during regular medical examinations. The high sound indicates high-frequency blood vessel noise. It is believed to denote that the diameter of the shunt vessel is thin over long distances. A hard sound is produced by a substantial increase in the amount of vascular noise at the peak, and is considered to indicate arteriovenous fistula vascular resistance physiologically. The collected sound sources were divided into two groups: Test data and training data. We ensured that sources from patients included in the training data were not included in the test data group. Although the training data should have had a large number of samples, the total number of training data was adjusted to be small because the number of abnormal sounds was minimal.

In addition, we used class weights that change the weights during learning so that the adverse effects of imbalanced data were minimized. Table 2 presents the contents of the data classified by the human ear. A dataset of 4000 sounds, containing minimal background noise and exhibiting representative characteristics, was used as the learning source. Figure 2 illustrates an outline of the learning preprocessing.

Table 2. Content of the data classified by the human ear and input data used for learning and training. Training and test data were categorized to avoid similar input data from the same patient.

	Training	Test	Total
Normal	394	485	879
Hard	578	901	1479
High	670	563	1233
Intermittent	91	4	95
Whistle	91	217	308
Total	1824	2170	3994

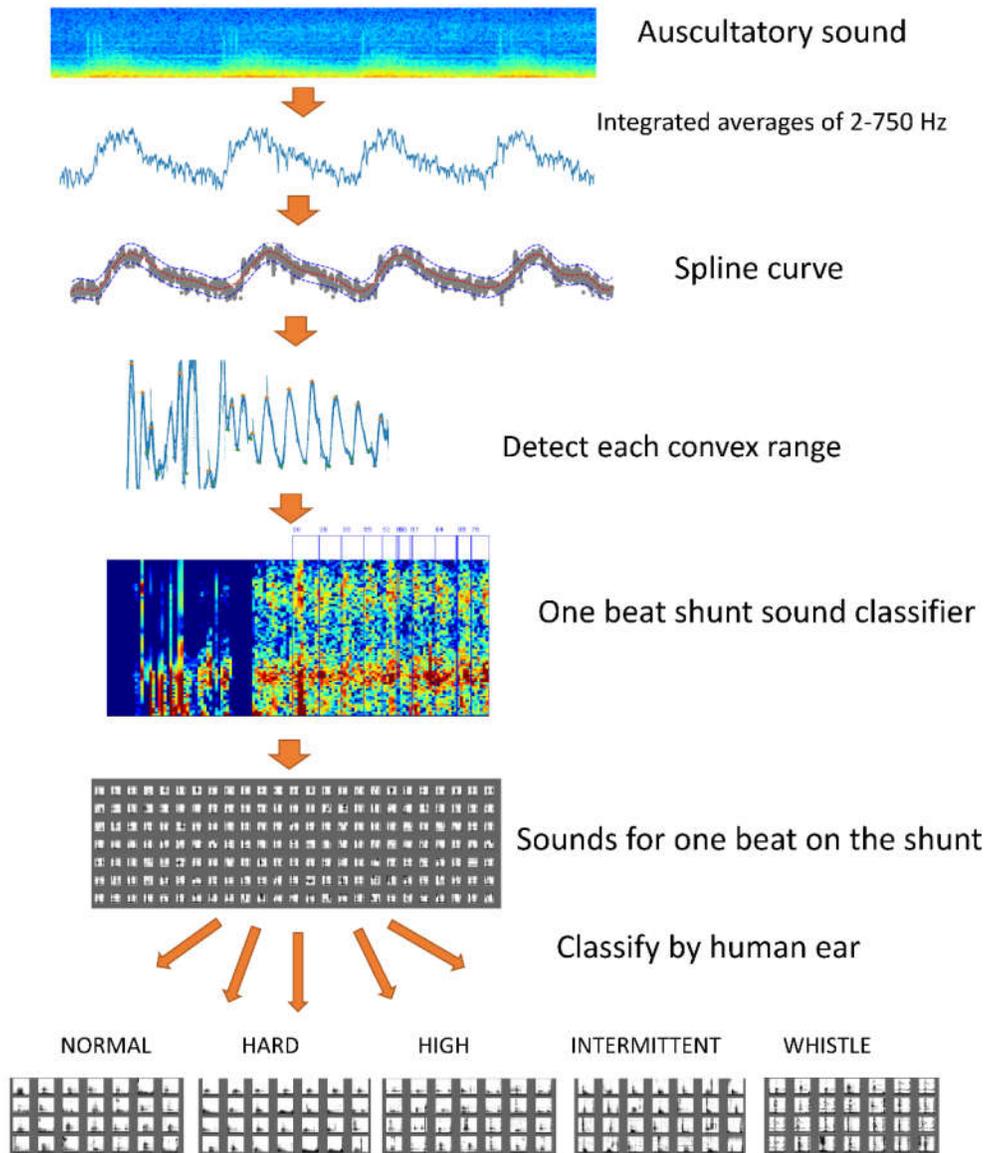


Figure 2. Learning preprocessing. A spline curve was created to eliminate noise from one auscultation sound. We extracted a number of convex curve ranges from the created spline curve with a duration of 0.5–2.0 s. The result was estimated to include an arteriovenous fistula sound equivalent to one heartbeat. The sound of one beat of the arteriovenous fistula was extracted by a deep learning classifier. Arteriovenous fistula sounds of 10,000 beats were classified into one of five types (i.e., normal sound, hard sound, high sound, intermittent sound, and whistling).

2.2.2. Data Analysis

In the field of deep learning, the Kaggle Competition is an online platform in which statisticians and data analysts from around the world compete to determine who has developed the best models. A competition involving audio tagging took place in 2018. Participants were challenged to develop a machine that can identify common sounds, such as a dog barking, telephone ringing, or guitar being strummed. The Surrey CVSSP DCASE 2018 Task 2 system (an open-source program published on GitHub, MIT licensee) ranked third in the competition and was used in this study for sound classification, as we are familiar with Keras, i.e., the programming language used to develop the program, and the ability of the model to evaluate changes in the time series of sounds [8]. During the

actual competition at Kaggle, the sound source provided was 44 kHz; however, this time, we recorded at 192 kHz. The learning effect was expected by capturing fine rumble sounds with a high-resolution spectrogram. According to the product standard, the frequency that the stethoscope can collect is limited by the diaphragm to approximately 700 Hz at the maximum. However, when comparing the auscultation sounds recorded at 192 kHz, there was a wide range of frequencies that changed the sound pattern in synchronization with the beat of the arteriovenous fistula. These included a spiked short keystroke sound up to 3000 Hz, and the arteriovenous fistula sound, which converged to a constant frequency, was 1500 Hz maximum.

Therefore, the frequency axis of the spectrogram used for learning was set to 2000 Hz. The sample size in the Fourier transform was 192 KHz, the window function was 4096, and the hop size was 2048. Feature extraction from the spectrogram was performed by the Mel filter bank. The number of filter banks was increased from the standard 64–1024 to ensure that the effect could be confirmed by a high-resolution conversion, and a comparison was conducted. The collected sound source had a long sound when the heartbeat was slow and a short sound when the heartbeat was fast. Therefore, blanks were filled with blanks for short beats so that the feature amount on the time axis was constant. Gated recurrent unit (GRU) and long short-term memory (LSTM), which use time series data, were used for learning. A small number of features on the input time axis led to a small number of features on the convoluted time axis. The number of features on the time axis before and after the convolution was 400 and 12, respectively. Table 3 shows the learning models used in this system.

Table 3. Various learning models used for the input source. The basic structure of all learning models comprises a neuron model called VGG13. First, the input data (Log Mel spectrogram) to VGG13 is convolved with 64 filters of 3×3 size, and the same convolution is repeated. Next, a rectified linear unit (ReLU) is used as the activation function, and Batch normalization (BN) is performed on the obtained data. Subsequently, the information is compressed using a 2×2 size filter called max pooling. Then, the same operation is repeated by gradually increasing the number of filters so that fine features can be captured effectively. Finally, all the obtained data are combined, the activation function (softmax) is used to stimulate 5 neurons, and the output value represents classification into 5 classes. When using convolutional recurrent neural network (CRNN), we use data convoluted by VGG13 as the input data before fully combining them. The output results using the bidirectional gated recurrent units (Bi-GRU) or bidirectional long short-term memory (Bi-LSTM) of the neuron model, which can evaluate the time axis, are connected to the fully connected layer and classified into five classes.

VGG13	CRNN (Bi-GRU)	CRNN (Bi-LSTM)
Log mel spectrogram	Log mel spectrogram	Log mel spectrogram
3×3 , 64, BN, ReLU	3×3 , 64, BN, ReLU	3×3 , 64, BN, ReLU
3×3 , 64, BN, ReLU	3×3 , 64, BN, ReLU	3×3 , 64, BN, ReLU
2×2 Max Pooling	2×2 Max Pooling	2×2 Max Pooling
$3 \times 3,128$, BN, ReLU	$3 \times 3,128$, BN, ReLU	$3 \times 3,128$, BN, ReLU
$3 \times 3,128$, BN, ReLU	$3 \times 3,128$, BN, ReLU	$3 \times 3,128$, BN, ReLU
2×2 Max Pooling	2×2 Max Pooling	2×2 Max Pooling
$3 \times 3,256$, BN, ReLU	$3 \times 3,256$, BN, ReLU	$3 \times 3,256$, BN, ReLU
$3 \times 3,256$, BN, ReLU	$3 \times 3,256$, BN, ReLU	$3 \times 3,256$, BN, ReLU
2×2 Max Pooling	2×2 Max Pooling	2×2 Max Pooling
$3 \times 3,512$, BN, ReLU	$3 \times 3,512$, BN, ReLU	$3 \times 3,512$, BN, ReLU
$3 \times 3,512$, BN, ReLU	$3 \times 3,512$, BN, ReLU	$3 \times 3,512$, BN, ReLU
2×2 Max Pooling	2×2 Max Pooling	2×2 Max Pooling
$3 \times 3,512$, BN, ReLU	$3 \times 3,512$, BN, ReLU	$3 \times 3,512$, BN, ReLU
$3 \times 3,512$, BN, ReLU	$3 \times 3,512$, BN, ReLU	$3 \times 3,512$, BN, ReLU
	Bi-GRU, 512, ReLU	Bi-LSTM, 512, ReLU
	Global average pooling	
	Softmax (5 classes)	

3. Results

3.1. Learning Curve

We compared various learning methods, such as AlexNet, VGG13, a deep residual network (ResNet), and a convolutional recurrent neural network (CRNN), with a bidirectional gated recurrent unit and bidirectional long short-term memory (Bi-GRU, Bi-LSTM). We also compared the mel-frequency cepstrum, mel-frequency log spectrogram, and simple spectrogram with the input signal. Reasonable learning accuracy rates were achieved with the CRNN:Bi-GRU model, which proved capable of learning from sources that included a time series as the data source and the mel-frequency log spectrogram as the input source. Figure 3 shows a graph of the learning process.

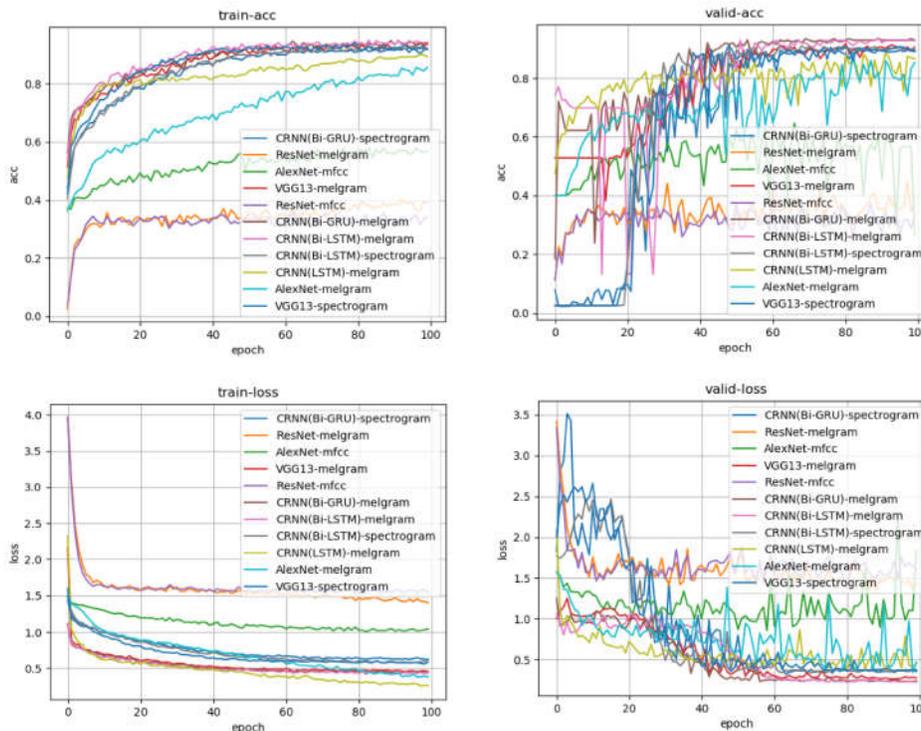


Figure 3. Learning curve for each learning model and each input source. Accuracy is presented in the upper row, whereas the loss is in the lower row. The left side is the transition based on the training data, and the right side is the transition based on the verification data. The horizontal axis indicates the number of times of learning, and accuracy increases as learning progress. Loss represents the difference between the answer of the input data predicted by the model during the learning process (e.g., the degree of firing that is a high tone) and the teacher's answer to the actual input data (the high tone is the correct answer). It can be observed that the difference between the answer and answer from the learning model obtained during learning has decreased. The CRNN:Bi-GRU model, which had the mel-frequency log spectrogram as input, was the learning model with good accuracy and loss in both training data and verification data.

3.2. Postprocessing

When the five types of sounds (i.e., normal, hard, high, intermittent, and whistling) were individually analyzed, a model suitable for each sound was found. Receiver operating characteristic (ROC) curves were created using the model that best fits each sound. Comparing the area under the ROC curves (AUCs), the model best fitting high and intermittent sound was the GRU with short-term memory; high sounds were best classified using a CNN that could identify rumble sounds. Table 4 lists AUCs classified by each learning model and their corresponding input data features.

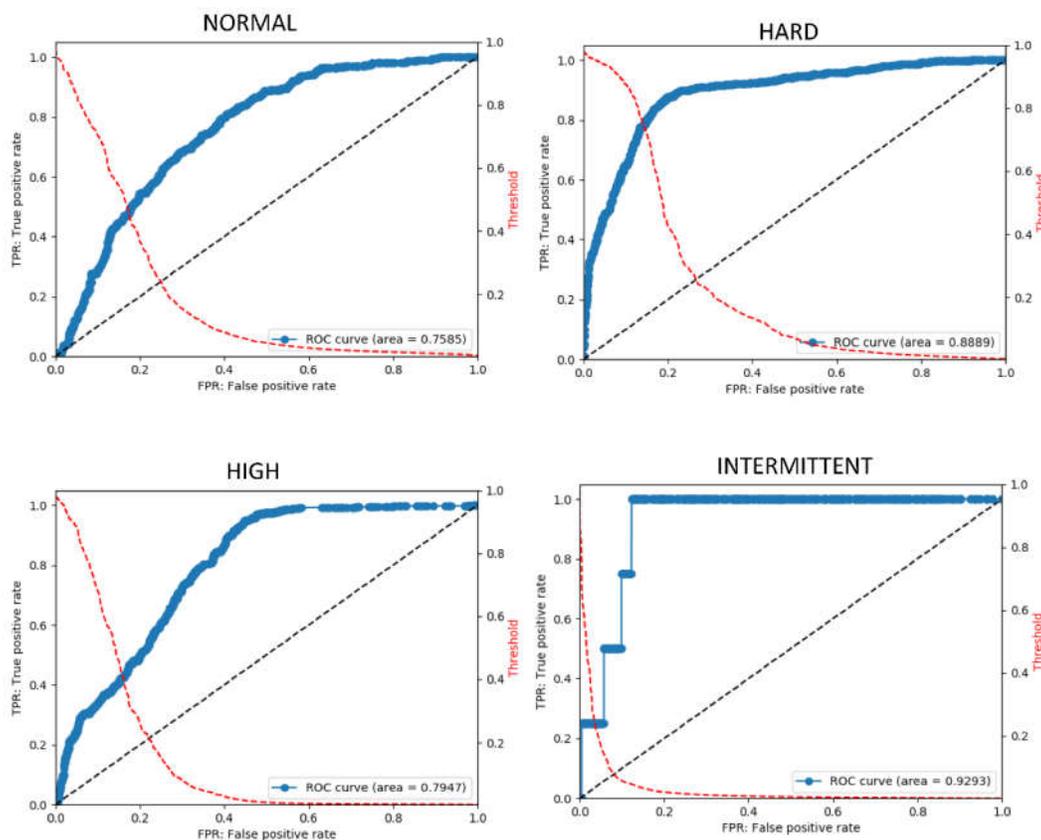
Table 4. Comparison of the areas under the receiver operating characteristic (ROC) curves (AUCs). A list of AUCs classified by each learning model, each input feature, and each sound is presented. For each sound, the item with the maximum AUC is highlighted. Hard has the best score in the GRU model using 1024 features. High is the model using 128 features, intermittent is the GRU model using 1024 features, and whistle is the CNN and GRU model using 64–256 features.

Feature	64			128			256			512			1024		
	CNN	GRU	LSTM	CNN	GRU	LSTM	CNN	GRU	LSTM	CNN	GRU	LSTM	CNN	GRU	LSTM
Normal	0.58	0.59	0.60	0.70	0.72	0.64	0.69	0.73	0.66	0.69	0.72	0.66	0.70	0.72	0.65
Hard	0.70	0.70	0.68	0.81	0.81	0.69	0.84	0.87	0.78	0.83	0.90	0.85	0.83	0.91	0.73
High	0.77	0.77	0.76	0.80	0.77	0.78	0.80	0.80	0.80	0.78	0.80	0.80	0.79	0.76	0.77
Intermittent	0.83	0.88	0.77	0.78	0.85	0.82	0.83	0.78	0.77	0.87	0.87	0.82	0.84	0.94	0.92
Whistle	0.89	0.85	0.85	0.89	0.89	0.86	0.89	0.89	0.89	0.87	0.87	0.87	0.88	0.88	0.86

An ensemble method combining judgment outputs was developed using a model advantageous for each sound. The results of the five classifications of unknown data performed by the created classifier are as follows. For each sound type listed above, the accuracy was 75–93%, the sensitivity was 46–86%, and the specificity was 81–93%. According to the ROC curve, the AUC was 0.75–0.92. These results are reasonably reliable for evaluating arteriovenous fistula sounds. Table 5 summarizes the final scores. Figure 4 shows the final ROC curve for each sound type.

Table 5. Final score. The final score is obtained by ensembling a plurality of models in which the best score is recorded for each sound using a technique called stacking and using the final stage model. The correct answer rate for the test data was unknown 72–93% (average 82%), and the AUC was 0.75–0.92% (average 0.83%).

	NORMAL	HARD	HIGH	INTERMITTENT	WHISTLE	MEAN
accuracy	0.753	0.833	0.729	0.935	0.855	0.821
precision	0.45	0.766	0.478	0.014	0.36	0.414
recall	0.468	0.862	0.478	0.5	0.581	0.578
specificity	0.836	0.812	0.817	0.936	0.885	0.857
f1	0.459	0.811	0.478	0.028	0.444	0.444
AUC	0.759	0.889	0.795	0.929	0.845	0.843



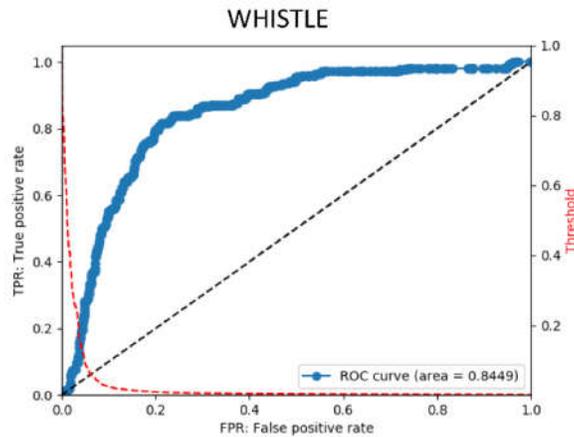


Figure 4. Final ROC curve obtained using the final stage model. The ROC curve is obtained by changing the threshold and plotting the true positive rate (TPR) at each threshold on the vertical axis and the false positive rate (FPR) on the horizontal axis. The ROC curve is on the blue line. The red dotted line shows the curve for obtaining the FPF at each threshold. It can be observed that, by changing the threshold, the threshold can be adjusted to lower FPR to detect rare diseases.

3.3. Clinical Application

The analysis of arteriovenous fistula sounds in the two cases that followed the developmental progress starting from arteriovenous fistula creation and continuing for one month is discussed in this section. Figure 5 shows the output results. The first case is that of arteriovenous fistula construction resulting from acute renal failure, with the puncture starting two weeks later. In the second case, the arteriovenous fistula was constructed after a long history of diabetes, and the puncture was initiated after three weeks. The latter artery was a multilayered, highly calcified vessel. The former was affected by vasospasm immediately after arteriovenous fistula creation [16], so the initial sound had a small component of hardness. Thereafter, in both cases, it was observed that a hard sound was conspicuous at an early stage, even when the arteriovenous fistula sound first developed. Moreover, the ratio of random intermittent and harmonic sounds increased after the start of puncturing. This may have indicated an increased risk of arteriovenous fistula stenosis after puncture [17].

There have been reports on the use of deep learning for digitizing visual examination results in dermatology, as well as in analyzing heart sounds for the estimation of heart disease [21–23]. However, this approach requires that only useful heart sounds should be extracted for a consultation record, and the preprocessing required for a trained classifier also hinders use in clinical settings. In this study, we used a region-detection classifier that detects a single beat from the preprocessing stage for the deep learning of arteriovenous fistula sounds. The required arteriovenous fistula beat sound was detected from the stored long-term auscultation sound without prior knowledge of the record, and the sound could be connected to the stenosis judgment classifier as carefully selected judgment data material. Deep learning using a carefully selected one-beat arteriovenous fistula sound has facilitated the objective digitization of the subjective evaluation performed by the main examiner.

In addition, by converting simple scientific findings into objective numerical values, it becomes possible to obtain frequent objective medical data. In the area of arteriovenous fistula examination, frequent observation may allow stenosis caused by a puncture to be found in the early stages of arteriovenous fistula development. Multidimensional data that were originally difficult to compare and evaluate are not limited to storage as images but could be digitized for medical examination evaluation. Thus, this practice may be applied to other medical fields if an input device is developed.

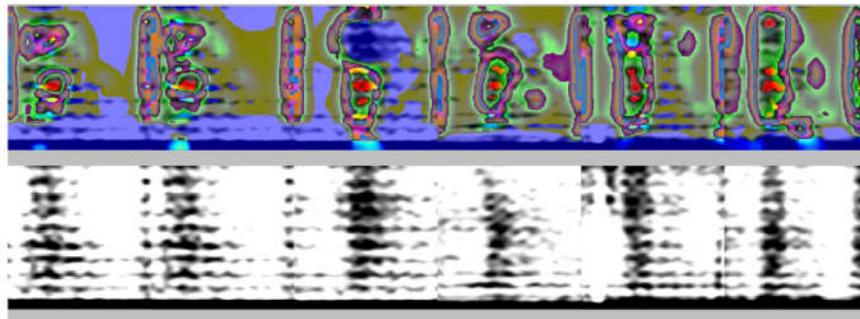
However, classifications based on black-box artificial intelligence methods remain unclear. An example of this is Grad-CAM, which identifies the feature points that CNN selects, focuses deep learning on the sound pattern image [24], and highlights these features in a heat map. In a different application, using a “dog” image classifier, color is applied to the characteristic portion of test images determined to be those of a dog. Figure 6 shows a concrete image output by Grad-CAM. In our case, when the characteristic portion of the sound pattern image that felt “high” was visualized, the 250–750 Hz region in the systole was emphasized in the heat map. When the characteristic part of the sound pattern image that felt “hard” was visualized, the silent region in the diastolic area was emphasized. It can be observed that the rumble region in the high-frequency region is used as an index, which is consistent with a previous report on the features in the spectrogram of stenosis sounds [25]. In the intermittent periods, the classifier focuses on the silence, and intermittent sound can be indexed as an element that allows a person to recognize a hard sound.

One limitation of this study is that the number of cases used was small; thus, the number of abnormal sounds collected was also small. During the learning phase, learning was performed by weighting the class of abnormal sounds; however, the extent of the performance handicap was unknown. In clinical practice, the disease frequency is low, approximately once every several thousand times. Regardless of how effective the classifier is, the number of judgment results will be false positive or false negative. Either can be prioritized by changing the cutoff value of the classifier; nonetheless, the effect is a tradeoff. As a countermeasure, in the medical field, a population with a higher pre-test probability in other items can be extracted and applied to the judgment machine. In this arteriovenous fistula sound classification approach, a large classifier that combines the classical statistical method based on physical sound features [26] and a learning model that detects abnormal sounds by unsupervised learning with this learning model is considered to be an effective classification method [27–29].

Dog image classifier and Cat image classifier with Grad-CAM



Sound pattern image that felt "HIGH"



Sound pattern image that felt "HARD"

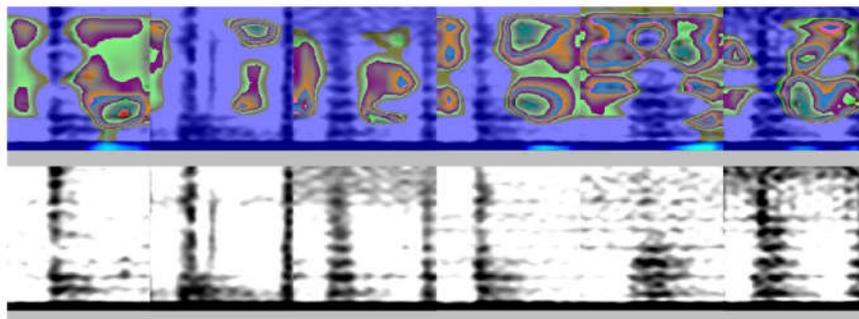


Figure 6. Grad-CAM for visualization of characteristic sites. With Grad-CAM, if the image is judged to be a dog using a dog image classifier, the image is colored. The same applies to a cat image classifier. When the characteristic portion of the sound pattern image that felt "high" is visualized, the 250–750 Hz region in the systole is emphasized in the heat map. In the case of a sound pattern image that felt "hard," the silent region in the diastolic area was emphasized.

The system used in this case was a multi-class classification system that distinguishes the sound from everyday sounds. The arteriovenous fistula sound may have harmonic, intermittent, and whistle sounds that simultaneously overlap. Thus, we should have changed the system to multi-label classification. Moreover, the diagnostic accuracy did not increase although it was a normal sound with a high proportion. The concept of normal sound is an excluded item that is not abnormal, and the feature amount is difficult to extract. The arteriovenous fistula sound should be a regression problem that numerically outputs the degree of stenosis, not a classification problem. The numerical value output by the system based on this classification problem is not proportional to the degree of stenosis. For example, the number output in the item of harmonics is a percentage representing the possibility of being the average harmonics collected this time. If the value is high, it does not mean that the constriction is a strong harmonic. When the average harmonics collected by

the training data are compared with the other groups, the ratio of the average harmonics is shown. As the teacher data determines the presence of harmonics, realistic recursive numerical values are difficult to output.

In future examinations of the risk of stenosis, arteriovenous fistula auscultation should not be studied in only one time period. However, the hard sound before stenosis, the high and whistle sounds that signify stenosis, and the normal sound after stenosis is resolved should all be studied [30]. It will likely prove necessary to construct a model for arteriovenous fistula stenosis diagnosis that uses a higher-dimensional time-series learning model, instead of the single-beat model. In addition, the accuracy rate for the existing model was high. However, it decreased when presented with new data, indicating inadequate generalization. Additional data collection periods may be required until a small number of abnormal sounds have sufficient data to facilitate classification. Finally, evidence confirming that this method can realize the ultimate goal of assisting with diagnoses is insufficient; consequently, comparisons with existing ultrasound resistance indices and a comparison before and after PTA remain necessary in clinical practice [31].

5. Conclusions

Arteriovenous fistula auscultation can be substituted using deep learning methods with high accuracy. However, arteriovenous fistula consultations require careful evaluation in conjunction with visual inspection, palpation, and other types of information. As the training data used for deep learning are also influenced by the audio of various devices and environmental characteristics of the facility, assessing and improving the reliability, efficacy, and cut-off scores of these methods require multi-center studies with larger populations.

Author Contributions: Conceptualization, K.O. and T.H.; Methodology, K.O.; Software, K.O.; Investigation, K.O., Y.N., S.I., H.A., and T.Y.; Writing-Original Draft Preparation, K.O.; Writing-Review & Editing, K.O. and T.H.; Supervision, K.O. and T.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the staff of Nagoya City University and Gamagori Municipal Hospital for their support and cooperation.

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

References

1. Lee, T.; Barker, J.; Allon, M. Needle Infiltration of Arteriovenous Fistulae in Hemodialysis: Risk Factors and Consequences. *Am. J. Kidney Dis.* **2006**, *47*, 1020–1026.
2. Remuzzi, A.; Ene-Iordache, B. Novel Paradigms for Dialysis Vascular Access: Upstream Hemodynamics and Vascular Remodeling in Dialysis Access Stenosis. *Clin. J. Am. Soc. Nephrol.* **2013**, *8*, 2186–2193.
3. Brahmabhatt, A.; Remuzzi, A.; Franzoni, M.; Misra, S. The molecular mechanisms of hemodialysis vascular access failure. *Kidney Int.* **2016**, *89*, 303–316.
4. Badero, O.J.; Salifu, M.O.; Wasse, H.; Work, J. Frequency of Swing-Segment Stenosis in Referred Dialysis Patients With Angiographically Documented Lesions. *Am. J. Kidney Dis.* **2008**, *51*, 93–98.
5. Schmidli, J.; Widmer, M.K.; Basile, C.; De Donato, G.; Gallieni, M.; Gibbons, C.P.; Haage, P.; Hamilton, G.; Hedin, U.; Kamper, L.; et al. Editor's Choice—Vascular Access: 2018 Clinical Practice Guidelines of the European Society for Vascular Surgery (ESVS). *Eur. J. Vasc. Endovasc. Surg.* **2018**, *55*, 757–818.
6. Vascular Access 2006 Work Group. National Kidney Foundation Vascular Access 2006 Work Group. KDOQI. Clinical practice guidelines for vascular access. *Am. J. Kidney Dis.* **2006**, *48* (Suppl. 1), S176–S247.
7. Al-Jaishi, A.A.; Oliver, M.J.; Thomas, S.M.; Lok, C.E.; Zhang, J.C.; Garg, A.X.; Kosa, S.D.; Quinn, R.R.; Moist, L.M. Patency Rates of the Arteriovenous Fistula for Hemodialysis: A Systematic Review and Meta-analysis. *Am. J. Kidney Dis.* **2014**, *63*, 464–478.
8. Salman, L.; Beathard, G. Interventional Nephrology: Physical Examination as a Tool for Surveillance for the Hemodialysis Arteriovenous Access. *Clin. J. Am. Soc. Nephrol.* **2013**, *8*, 1220–1227.
9. Sato, T. New diagnostic method according to the acoustic analysis of the shunt blood vessel noise. *Toin Univ. Yokohama Eng. Jpn. Soc. Dial. Ther. J.* **2005**, *2*, 332–341.

10. Kokorozashi, N. Analysis of the shunt sound frequency characteristic changes associated with shunt stenosis. *Jpn. Soc. Dial. Ther. J.* **2010**, *3*, 287–295.
11. Todo, A.; Kadonaka, T.; Yoshioka, M.; Ueno, A.; Mitani, M.; Katsurao, H. Frequency Analysis of Shunt Sounds in the Arteriovenous Fistula on Hemodialysis Patients. In Proceedings of the 6th International Conference on Soft Computing and Intelligent Systems, and the 13th International Symposium on Advanced Intelligence Systems, Kobe, Japan, 20–24 November 2012; doi:10.1109/scis-isis.2012.6505044.
12. Wang, H.-Y.; Wu, C.-H.; Chen, C.-Y.; Lin, B.-S. Novel Noninvasive Approach for Detecting Arteriovenous Fistula Stenosis. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 1851–1857.
13. Kamijo, Y.; Kanda, E.; Horiuchi, H.; Kounoue, N.; Ono, K.; Maeda, K.; Yanai, A.; Honda, K.; Tsujimoto, R.; Yanagi, M.; et al. Continuous monitoring of blood pressure by analyzing the blood flow sound of arteriovenous fistula in hemodialysis patients. *Clin. Exp. Nephrol.* **2017**, *22*, 677–683.
14. Serven, D.; Brummit, C. pyGAM: Generalized Additive Models in Python. 2018. Available online: <https://zenodo.org/record/1476122> (accessed on 26 August 2020).
15. Iqbal, T.; Kong, Q.; Plumbley, M.D.; Wang, W. *Stacked Convolutional Neural Networks for General-Purpose Audio Tagging*; University of Surrey: Guildford, UK, 2018.
16. Owada, A.; Saito, H.; Mochizuki, T.; Tanaka, M.; Otsuka, M.; Yokoyama, Y.; Fukuda, S. Radial Arterial Spasm in Uremic Patients Undergoing Construction of Arteriovenous Hemodialysis Fistulas: Diagnosis and Prophylaxis with Intravenous Nicardipine. *Nephron* **1993**, *64*, 501–504, doi:10.1159/000187390.
17. Keisuke, O. Cannulation should be more than 3 weeks after creation of a radial-cephalic arterio-venous fistula. *Clin. Surg.* **2019**, *4*, 2299.
18. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88.
19. Akatsuka, J.; Yamamoto, Y.; Sekine, T.; Numata, Y.; Morikawa, H.; Tsutsumi, K.; Yanagi, M.; Endo, Y.; Takeda, H.; Hayashi, T.; et al. Illuminating Clues of Cancer Buried in Prostate MR Image: Deep Learning and Expert Approaches. *Biomolecules* **2019**, *9*, 673.
20. Asif, T.; Mohiuddin, A.; Hasan, B.; Pauly, R.R. Importance of Thorough Physical Examination: A Lost Art. *Cureus* **2017**, *9*, 1212.
21. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118.
22. Messner, E.; Zöhrer, M.; Pernkopf, F. Heart Sound Segmentation—An Event Detection Approach Using Deep Recurrent Neural Networks. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1964–1974.
23. Raza, A.; Mehmood, A.; Ullah, S.; Ahmad, M.; Choi, G.S.; On, B.-W. Heartbeat Sound Signal Classification Using Deep Learning. *Sensors* **2019**, *19*, 4819.
24. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359.
25. Wakisaka, Y.; Miura, T. A new method for estimating blood flow through arteriovenous fistulas and grafts in patients undergoing hemodialysis. *Jpn. Soc. Dial. Ther.* **2019**, *52*, 569–575.
26. Higashi, D.; Nishijima, K. Classification of Shunt Murmurs for Diagnosis of Arteriovenous Fistula Stenosis. In Proceedings of the 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Honolulu, HI, USA, 12–15 November 2018; doi:10.23919/APSIPA.2018.8659641.
27. Chalapathy, R.; Chawla, S. Deep learning for anomaly detection: A survey. *arXiv* **2019**, arXiv:1901.03407.
28. Vununu, C.; Moon, K.-S.; Lee, S.-H.; Kwon, K.-R. A Deep Feature Learning Method for Drill Bits Monitoring Using the Spectral Analysis of the Acoustic Signals. *Sensors* **2018**, *18*, 2634, doi:10.3390/s18082634.
29. Heo, S.-J.; Kim, Y.; Yun, S.; Lim, S.-S.; Kim, J.; Nam, C.M.; Park, E.-C.; Jung, I.; Yoon, J.-H. Deep Learning Algorithms with Demographic Information Help to Detect Tuberculosis in Chest Radiographs in Annual Workers’ Health Examination Data. *Int. J. Environ. Res. Public Health* **2019**, *16*, 250, doi:10.3390/ijerph16020250.
30. Matsuura, K.; Gotoh, Y.; Sadaoka, S.; Takase, K.; Narimatsu, Y. Guidelines for basic techniques in vascular access intervention therapy (VAIVT). *Interv. Radiol.* **2018**, *3*, 28–43, doi:10.22575/interventionalradiology.3-1g2.

31. Murakami, K.; Inomata, F.; Takeda, T.; Utino, J.; Sakai, T.; Kouno, T.; Sigematu, T.; Yoshida, T. Usefulness of pulsed doppler ultrasonography to manage internal AV shunt. *Kidney Dial.* **2003**, *55*, 39–43.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).