Further Reading

Definitions:

Heart Arrhythmia:
An arrhythmia is an abnormality of the heart's rhythm.

Heart Murmur:
Heart murmurs are sounds — such as whooshing or swishing — made by rapid,
choppy (turbulent) blood flow through the heart.

Extrasystoles/Premature ventricular contractions:
Extra heartbeats that begin in one of the heart's two lower pumping chambers
(ventricles).

Zero-Crossing Rate (ZCR):
A measurement in audio processing that counts the number of times a signal crosses
zero from positive to negative or vice versa within a specific time window. It's a
simple way to characterise the "brightness" or "harshness" of a sound.

Chroma Short-Time Fourier Transform (Chroma STFT):
A signal processing technique used to analyse audio. It breaks down the sound into
its component frequencies over short periods of time. Chroma STFT focuses
specifically on the pitch information, representing the signal in terms of musical notes
(chroma).

Mel-Frequency Cepstrum Coefficient (MFCC):
A feature extraction technique commonly used in speech and audio recognition. It
mimics how the human auditory system perceives sound by converting the sound's
frequency spectrum into a representation on the Mel scale, which approximates
human hearing. MFCCs capture the spectral envelope of the sound, making them
useful for tasks like speaker identification and speech recognition.

Mel-Spectrogram:
A visual representation of a sound's frequency content over time. It uses colour to
represent the intensity of different frequencies at different time points. Mel
spectrograms are often used in conjunction with MFCCs, as they provide a more
intuitive way to understand the spectral information captured by MFCCs.

Deep Neural Network (DNN):
A type of artificial neural network with multiple layers of interconnected nodes. DNNs
can learn complex patterns from data and are powerful tools for tasks like image
recognition, natural language processing, and speech recognition.

Convolutional Neural Network (CNN):
A powerful type of artificial neural network used to classify sounds. Similar to image
processing, CNNs excel at finding patterns in audio data.

Further Reading

Trainable Parameters:
The numerical values within a deep neural network that are adjusted during the training process. These parameters determine how the network maps input data to output predictions. By adjusting these parameters, the network learns to perform a specific task.

1-Dimensional Convolutional Layers:
A specific type of layer in a deep neural network used for processing sequential data like audio or text. These layers apply a filter (like a small window) that slides across the input data, extracting features based on local patterns. By stacking multiple convolutional layers, the network can learn increasingly complex features.

Linear Layers:
Layers in a deep neural network that perform a weighted sum of their inputs. These layers are often used at the end of a network to combine the learned features and produce the final output prediction.

ReLU Activation Function:
A popular activation function used in artificial neural networks. It adds a non-linearity to the network, which is crucial for its ability to learn complex patterns from data.

Cross-Entropy Loss Function:
A common function used to measure the error between a neural network's predictions and the true labels of the data. It's particularly useful for classification tasks where the network outputs probabilities for different categories. The loss function helps the network learn by indicating how much its predictions deviate from the desired outcome.

AdamW Optimizer:
An optimization algorithm used to train deep neural networks. It efficiently adjusts the network's trainable parameters based on the calculated loss function. AdamW is a variant of the Adam optimizer that addresses certain stability issues.

Overfitting:
A situation where a deep neural network performs well on the training data but poorly on unseen data. This occurs when the network memorises specific details of the training examples rather than learning generalisable patterns. Weight decay and other regularisation techniques can help prevent overfitting.

Weight Decay:
A technique used during training to prevent a deep neural network from overfitting to the training data. It penalises the network for having large weights, encouraging it to learn more generalisable features.

Further Reading

Dropout Layers:
A technique used in artificial neural networks to improve their performance, especially to address overfitting, by randomly deactivating a certain percentage of neurons in a layer during training. This forces the network to learn to use different combinations of neurons each time, preventing it from relying too heavily on any specific neuron or connection.

Upsampling:
An operation that increases the resolution of an image or signal. In audio processing, it might involve interpolating new data points between existing ones to create a higher sampling rate.

Minority Classes:
In a classification task with multiple categories, the classes with the fewest data points are referred to as minority classes.

Class Imbalance:
A situation in a classification dataset where some classes have significantly fewer data points than others. This can pose challenges for training a deep neural network, as the model might prioritise learning patterns from the majority classes and perform poorly on the minority classes.

Noise (in reference to audio):
Unwanted sound that disrupts the desired audio signal. Noise can come from various sources and can significantly impact the quality and clarity of audio recordings or playback.

MAX9814:
An integrated circuit (IC) that combines a microphone amplifier with automatic gain control (AGC). It's commonly used in audio applications to boost weak microphone signals to a usable level. AGC helps ensure the amplified signal stays within a certain range, preventing distortion.

MCP3008:
An analogue-to-digital converter (ADC) IC that converts analogue voltage signals from sensors or other circuits into digital data that can be processed by microcontrollers or computers. It has 8 channels, meaning it can convert signals from up to 8 analogue inputs simultaneously. The MCP3008 typically uses a serial communication protocol (SPI) to communicate with the microcontroller.

Links to helpful videos:
- ▶ But what is a neural network? | Chapter 1, Deep learning by 3Blue1Brown
- ▶ But what is a convolution? by 3Blue1Brown
- ▶ But what is the Fourier Transform? A visual introduction. by 3Blue1Brown