



# Redes Neurais Pré-Treinadas na Classificação Automática de Sons Cardíacos

**EDUARDO JOSÉ SILVA CARVALHO VIEIRA**  
Outubro de 2020

**P.PORTO**



Instituto Superior de  
**Engenharia** do Porto



Departamento de  
**Física**

Instituto Superior de Engenharia do Porto

# **Redes Neurais Pré-Treinadas na Classificação Automática de Sons Cardíacos**

---

Eduardo José Silva Carvalho Vieira

“Dissertação apresentada no Instituto Superior de Engenharia do Porto para a obtenção do grau de Mestre em Engenharia de Computação e Instrumentação Médica”

Orientador: Luís Pinto Coelho

**2020**



## **Agradecimentos**

Agradeço à minha família a confiança e apoio.

Agradeço aos professores e amigos que passaram e compartilharam a sua experiência.



## **Resumo**

As doenças cardiovasculares são uma das principais causas de morte e hospitalização, tanto em países desenvolvidos como em desenvolvimento. O seu diagnóstico requer intervenção profissional e equipamento específico, sendo normalmente dispendioso.

O desenvolvimento de algoritmos capazes de segmentar e classificar sinais dos batimentos cardíacos beneficia esta área, uma vez que muitas doenças cardiovasculares se manifestam como irregularidades nos mesmos. Estes algoritmos servirão de apoio ao diagnóstico para os profissionais de saúde e oferecem a possibilidade de serem incorporados em dispositivos próprios para uso doméstico reduzindo a necessidade de consumo de recursos hospitalares ou de centros privados de saúde. No entanto, até ao momento, não existem implementações, clínicas ou não, destes métodos.

Nos últimos anos, vários algoritmos de classificação baseados em diferentes técnicas surgiram e bases de dados vastas e de livre acesso foram disponibilizadas procurando estabelecer um ponto de comparação da eficácia dos mesmos.

A presente dissertação explora a eficácia da utilização de redes neuronais pré-treinadas na classificação dos sinais disponibilizados no *PhysioNet/CinC Challenge 2016*, uma das maiores bases de dados de fonocardiogramas já reunida. A melhor rede gerada obteve uma precisão de classificação de 80.85%, uma sensibilidade de 79.77% e uma especificidade de 81.94%, estando em linha com resultados obtidos por métodos diferentes e recorrendo a menos pré-processamento do sinal.

**Palavras-chave:** Redes Neuronais Convolucionais, Fonocardiograma, Classificação de Sons Cardíacos



## **Abstract**

Cardiovascular diseases are the leading cause of hospitalization and death, in both developed and developing countries. Its diagnosis requires expert intervention as well as specialized equipment, being costly.

The development of algorithms capable of segmenting and classifying signals from the heartbeat benefits this field since many cardiovascular diseases manifest themselves through irregular heartbeats. These algorithms will serve as a clinical decision support system for health professionals and offer the opportunity of creating domestic devices, reducing the need for hospital and private centres resource consumption. However, at the moment, there is no clinical or otherwise implementation of such technology.

In the last years, many classification algorithms working on different techniques have emerged and vast open source databases have been made available looking to establish a comparison between those methods.

This dissertation aims to test the efficiency of pre-trained neural networks in the classification of signals retrieved from the *PhysioNet/CinC Challenge 2016*, one of the largest collection of PCG ever assembled. Our best network achieved an accuracy of 80.85%, a recall of 79.77% and a specificity of 81.94%, being competitive with other methods and requiring less signal processing.

**Keywords:** Convolutional Neural Networks, PCG, Heart Sound Classification



# Índice

1. Introdução.....	17
2. Fonocardiograma (PCG).....	19
2.1. Sons Fundamentais (S1 e S2).....	19
2.1.1. Desdobramento de S2.....	20
2.2. Sons Cardíacos Anómalos .....	20
2.2.1. Ritmo de Galope (S3 e S4) .....	20
2.2.2. Sons de ejeção (“clicks”) .....	21
2.2.3. Sopros do Coração .....	21
2.3. Tempos e Frequências no PCG .....	25
3. Redes Neurais.....	27
3.1. Estrutura Básica das ANN .....	27
3.2. Aprendizagem nas ANN .....	28
3.3. Treino de uma ANN .....	28
3.4. Redes Neurais Convolucionais.....	29
3.4.1. Camadas de uma CNN.....	30
3.4.1.1. Convolucional .....	30
3.4.1.2. <i>Pooling/Sub-sampling</i> .....	31
3.4.1.3. Não-Linearidade.....	31
3.4.1.4. Totalmente Ligadas.....	32
4. Estado da Arte.....	33
4.1. Métodos Aplicados na Classificação de PCG.....	33
4.1.1. Pré-Processamento .....	33
4.1.2. Segmentação .....	34
4.1.3. Extração de <i>Features</i> .....	34
4.1.4. Classificação.....	34
4.2. Trabalhos Desenvolvidos .....	35
4.2.1. <i>2016 PhysioNet Challenge Top 3</i> .....	35
4.2.1.1. <i>Potes et al. (2017)</i> .....	35
4.2.1.2. <i>Zabihi et al. (2017)</i> .....	35
4.2.1.3. <i>Kay &amp; Agarwal (2017)</i> .....	36
4.2.1.4. <i>Aprendizagens do Desafio</i> .....	36
4.2.2. <i>Messner et al. (2018)</i> .....	37
4.2.3. <i>Cheng et al. (2019)</i> .....	37

4.2.4.	<i>Zhang et al. (2019)</i> .....	38
4.2.5.	<i>Morales et al. (2019)</i> .....	39
4.2.6.	<b>Revisão Literária</b> .....	40
<b>5.</b>	<b>Procedimento Experimental</b> .....	45
5.1.	<i>Hardware e Software</i> .....	46
5.2.	<b>Base de Dados</b> .....	46
5.3.	<b>Pré-Processamento</b> .....	46
5.4.	<b>Extração de <i>Features</i></b> .....	47
5.4.1.	<b>Espetrograma de Mel</b> .....	47
5.4.2.	<b>MFCC</b> .....	48
5.4.3.	<b>GTCC</b> .....	49
5.4.4.	<b>PNCC</b> .....	50
5.5.	<b>Treino das CNN</b> .....	50
5.6.	<b>Métricas de Avaliação</b> .....	53
<b>6.</b>	<b>Resultados e Discussão</b> .....	55
<b>7.</b>	<b>Conclusão</b> .....	63

# Índice de Figuras

Figura 2.1.	Corte transversal do coração.....	19
Figura 2.2.	Locais de auscultação das válvulas cardíacas.....	22
Figura 2.3.	Sopros Sistólicos.....	23
Figura 2.4.	Sopros Diastólicos.....	24
Figura 3.1.	Perceptrão.....	27
Figura 3.2.	Método de <i>Back-Propagation</i> .....	29
Figura 3.3.	Visualização de diferentes representações de <i>features</i> em diferentes camadas de uma CNN.....	29
Figura 3.4.	Ilustração da operação de convolução.....	30
Figura 3.5.	Representação de dois modos de <i>pooling</i> .....	31
Figura 3.6.	Algumas das funções de ativação mais utilizadas nas CNN.....	31
Figura 4.1.	Passos envolvidos na classificação de sons cardíacos.....	33
Figura 4.2.	Diagrama de blocos do algoritmo proposto por Potes et. al.....	35
Figura 4.3.	Esquema do algoritmo proposto por Zabihi et. al.....	36
Figura 4.4.	<i>Laconic Neural Network Structure</i> .....	38
Figura 4.5.	Framework do método proposto em [64].....	38
Figura 4.6.	Diagrama de blocos da arquitetura do método proposto em [66].....	39
Figura 4.7.	<b>Mono-Aural Neuromorphic Auditory Sensor for FPGA with an I2S audio ADC and AER interface.....</b>	<b>40</b>
Figura 5.1.	Fluxograma do método proposto.....	45
Figura 5.2.	10 segundos de sinal cardíaco normal.....	47
Figura 5.3.	Espetrograma de Mel.....	48
Figura 5.4.	<i>Mel-Frequency Cepstral Coefficients</i> .....	49
Figura 5.5.	<i>Gammatone Cepstral Coefficients</i> .....	49
Figura 5.6.	<i>Power Normalized-Cepstral Coefficients</i> .....	50
Figura 5.7.	Estratégias de <i>fine-tuning</i> .....	51
Figura 5.8.	Esquema das redes geradas a partir redes pré-treinadas.....	52
Figura 5.9.	Arquitetura dos classificadores gerados.....	53
Figura 6.1.	<b>Treino da rede com a base convolucional congelada e otimizador <i>Adamax</i>.....</b>	<b>57</b>
Figura 6.2.	<b>Treino da rede com a última camada da base convolucional descongelada.....</b>	<b>58</b>



## Índice de Tabelas

<b>Tabela 2.1.</b>	<b>Fisiopatologia de Sons Cardíacos Normais e Anómalos.....</b>	<b>25</b>
<b>Tabela 4.1.</b>	<b>Métodos de extração de características e de classificação de sinais PCG.....</b>	<b>40</b>
<b>Tabela 4.2.</b>	<b>Literatura de classificação de sons cardíacos aplicando <i>Deep Learning</i>.....</b>	<b>42</b>
<b>Tabela 6.1.</b>	<b>Resultados das redes treinadas na base de dados de imagens dos MFCC.....</b>	<b>55</b>
<b>Tabela 6.2.</b>	<b>Resultados das redes treinadas na base de dados de imagens dos Espectrogramas de Mel.....</b>	<b>56</b>
<b>Tabela 6.3.</b>	<b>Resultados das redes treinadas na base de dados de imagens dos MFCC acoplados aos coeficientes delta e delta-delta.....</b>	<b>59</b>
<b>Tabela 6.4.</b>	<b>Resultados das redes treinadas na base de dados de imagens dos GTCC.....</b>	<b>59</b>
<b>Tabela 6.5.</b>	<b>Resultados das redes treinadas na base de dados de imagens dos PNCC.....</b>	<b>59</b>
<b>Tabela 6.6.</b>	<b>Resultados de precisão da rede gerada em ambas as bases de dados do grupo PASCAL.....</b>	<b>60</b>



## **Simbologia e Notações**

AER – *Adress-Event Representation*

AMDF – *Average Magnitude Difference Function*

ANN – *Artificial Neural Network*

BiGRNN – *Bidirectional Gated Recurrent Neural Networks*

CNN – *Convolutional Neural Network*

CWD – *Choi-Williams Distribution*

CWT – *Continuous Wavelet Transform*

DNN – *Deep Neural Network*

DWT – *Discrete Wavelet Transform*

EMD – *Empirical Mode Decomposition*

FPGA – *Field-Programmable Gate Array*

GTCC – *Gammatone Cepstral Coefficients*

HMM – *Hidden Markov Model*

k-NN – *k-Nearest Neighbour*

LNN – *Laconic Neural Network*

LSTM – *Long Short-Term Memory*

MAcc – *Modified Accuracy*

MFCC – *Mel-frequency Cepstral Coefficients*

NAS – *Neuromorphic Auditory Sensor*

NN – *Neural network*

PCA – *Principal Component Analysis*

PCG – *Phonocardiogram*

PNCC – *Power-Normalized Cepstral Coefficients*

ReLU – *Rectified Linear Units*

SNR – *Signal-to-Noise Ratio*

SSA – *Singular Spectral Analysis*

STFT – *Short-Time Fourier Transform*

STMHT – *Short-Time Modified Hilbert Transform*

SVM – *Support Vector Machine*

TQWT – *Tunable Q-Wavelet Transform*



## 1. Introdução

No mundo contemporâneo, doenças cardiovasculares são das principais causas de morte, tanto em países desenvolvidos como em vias de desenvolvimento, e motivo de hospitalização [1]. É estimado que 90% dos casos diagnosticados poderiam ser evitados com prevenção adequada [2].

No contexto das doenças cardiovasculares, muitas manifestam-se por alterações nos batimentos cardíacos, é o exemplo das arritmias, disfunção das válvulas, hipertensão pulmonar, etc. Estes batimentos irregulares são normalmente diagnosticados numa fase primária com recurso a eletrocardiogramas (ECG), registos da atividade elétrica do coração, ou fonocardiogramas (PCG, “*phonocardiogram*”), registo da onda sonora dos batimentos cardíacos, entre outros. Estes exigem equipamento específico e conhecimento técnico para análise dos sinais obtidos, sendo geralmente realizados em hospitais ou centros privados por profissionais de saúde.

A técnica de auscultação é rotineiramente utilizada para avaliar acusticamente a atividade mecânica do coração. Consiste no profissional de saúde escutar o batimento cardíaco através de um estetoscópio encostado em diferentes pontos da caixa torácica. Identificar irregularidades cardíacas como sopros (som de um fluxo sanguíneo turbulento), sons extra, ou outras anomalias, requer treino e está dependente da experiência do clínico. Não só requer a experiência, como em certos casos pode requerer uma monitorização prolongada para a patologia se manifestar. É estimado que clínicos em treino falham em distinguir sons normais de sopros malignos em 80% dos casos [3]. Erros de auscultação levam a que pacientes sejam muitas vezes dirigidos a exames mais dispendiosos e desnecessários [3]. No entanto, esta técnica comprovou ter uma boa relação custo-eficácia e uma série de patologias podem ser detetadas nas fases iniciais [4].

De tal modo, é de valor para o clínico a criação de um sistema de apoio ao diagnóstico por extração automática de informações relevantes do PCG. O PCG pode ser registado por um estetoscópio eletrónico no momento da auscultação. A qualidade do estetoscópio e o ambiente de aquisição influenciam a qualidade do sinal.

De facto, uma das dificuldades em estabelecer um ponto de comparação para os diferentes métodos de classificação dos batimentos cardíacos está na variância das condições de aquisição. O *2016 PhysioNet/Computing in Cardiology (CinC) Challenge* abordou este problema estabelecendo a maior base de dados de livre acesso de amostras de PCG, cedidas por diferentes equipas de investigação, em ambiente clínico e não-clínico, em diferentes locais de aquisição e com diferentes níveis de ruído. Uma descrição pormenorizada da base de dados pode ser encontrada em [5].

Geralmente, o algoritmo de classificação compreende 4 fases essenciais: O **pré-processamento**, que trata de remover ruídos inerentes à aquisição do sinal e das transformações necessárias ao original. A gama de frequências escolhida para redução do ruído, bem como as transformações do sinal (reamostragem, suavização, remoção do valor médio, etc.) variam

conforme o caso específico e necessidades do algoritmo. A **segmentação** do sinal nos seus componentes principais. O PCG de um indivíduo saudável compreende dois componentes fundamentais, S1, que corresponde ao início da sístole ventricular, e S2, que corresponde ao início da diástole ventricular. O terceiro, S3, e quarto som, S4, não sendo fundamentais, podem ou não estar presentes e ser fisiológicos ou indicadores de uma patologia. Os segmentos entre estes componentes, o período sistólico e diastólico, respetivamente, também transportam informação relevante (duração anormal de uma das fases, etc.). A **extração e selecção de características** (*features*). Este passo não é comum a todos os algoritmos de classificação, mas consiste em escolher propriedades do sinal que o diferenciem nas classes pretendidas, evitando redundância e irrelevância dos dados. Finalmente, a **classificação** dos segmentos como batimentos normais ou patológicos. Os métodos de classificação típicos podem considerar-se: redes artificiais neuronais (ANN) [6], *support vector machines* (SVM) [7], hidden Markov models (HMM) [8] e métodos de *cluster* [9].

A presente dissertação está dividida da seguinte forma: o segundo capítulo é uma familiarização com o PCG e patologias detetadas pelo mesmo. É seguido de uma abordagem às bases das redes neuronais e, mais especificamente, das redes convolucionais, no terceiro capítulo. Uma revisão literária e algoritmos do estado da arte são apresentados no quarto capítulo. O quinto capítulo apresenta o método proposto e detalhes do procedimento experimental. É seguido dos resultados obtidos por esse mesmo procedimento no sexto capítulo. Finalmente, no capítulo sete, uma reflexão sobre os dados obtidos e futuros melhoramentos à tecnologia.

## 2. Fonocardiograma (PCG)

O fonocardiograma (PCG – “*phonocardiogram*”) é o registo dos sons produzidos pelo coração, obtido através de um aparelho denominado fonocardiógrafo. Estes sons originam das vibrações criadas pelo fecho das válvulas ou fluxo turbulento do sangue.

### 2.1. Sons Fundamentais (S1 e S2)

O tom e tempo de ocorrência dos sons cardíacos seguem um padrão. Em indivíduos saudáveis compreende-se dois sons fundamentais. O primeiro som, S1, corresponde ao início da sístole ventricular e é originado pelo fecho das válvulas mitral e tricúspide (coletivamente denominadas válvulas atrioventriculares). O segundo som, S2, corresponde ao início da diástole ventricular e é originado pelo fecho das válvulas aórtica e pulmonar (coletivamente denominadas válvulas semilunares). Uma representação das válvulas mencionadas pode ser encontrada na Figura 2.1.

Normalmente, S1 é um som mais grave e de maior duração enquanto S2 é mais agudo e tem menor duração. O intervalo sistólico (S1-S2) é mais curto que o intervalo diastólico (S2-S1).

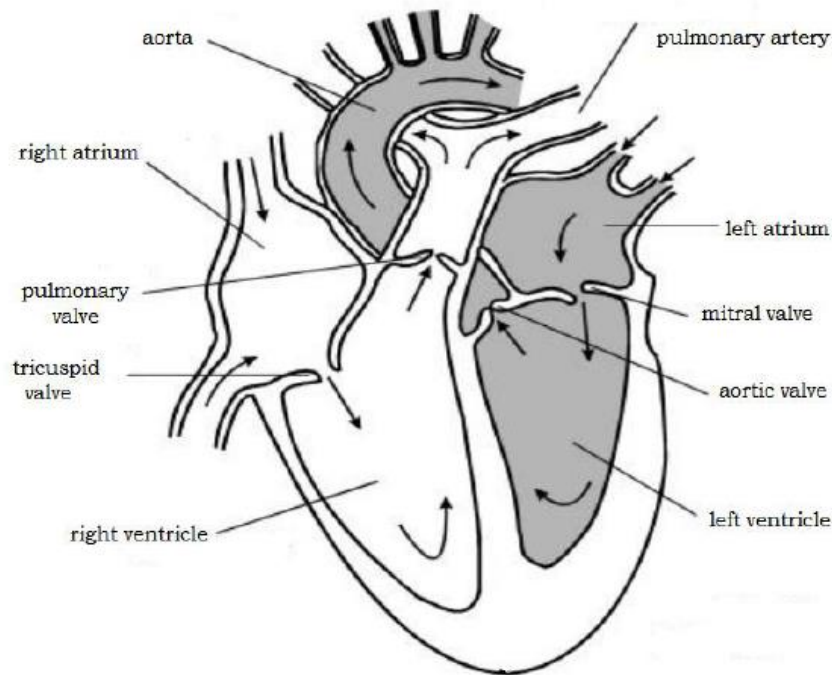


Figura 2.1 – Corte transversal do coração (disponível em: V. Elamaran et al.: *Spectral Fault Recovery Analysis Revisited With Normal and Abnormal Heart Sound Signals*, vol. 6, pág. 2).

### 2.1.1. Desdobramento de S2

O som S2 é formado por dois componentes: A2, que corresponde ao fecho da válvula aórtica e apresenta uma intensidade maior relativamente ao outro componente, P2, que corresponde ao fecho da válvula pulmonar. A avaliação clínica do som S2 foi referida como a “chave para a auscultação do coração” [10]. Envolve a determinação do desdobramento e intensidades relativas entre A2 e P2. Normalmente, o fecho da válvula aórtica, A2, antecede o da válvula pulmonar, P2 e o intervalo entre os dois (desdobramento) aumenta na inspiração e diminui na expiração. Em repouso e durante a inspiração, varia entre 20 e 80 ms e geralmente não excede os 40 ms na expiração [11]. A duração deste intervalo entre A2 e P2 é uma medida da impedância vascular. Num sentido fisiológico simples, a inspiração reduz a impedância no circuito pulmonar, prolonga o intervalo e retarda o fecho da válvula pulmonar, resultando num desdobramento mais óbvio de S2. Na expiração ocorre o inverso, o fecho da válvula pulmonar ocorre mais cedo e o intervalo entre A2 e P2 apresenta uma média entre 30 e 40 ms [11].

O desdobramento do som S2 é, portanto, um indicador do correto funcionamento do coração. Este desdobramento pode ser categorizado da seguinte forma: [12] [13] [14]

- **Fisiológico** – durante a inspiração, o fecho da válvula aórtica, A2, precede o da válvula pulmonar, P2. O desdobramento é especialmente pronunciado entre indivíduos jovens.
- **Largo** – um desdobramento exagerado; mais pronunciado durante a inspiração (A2 precede P2). Indicador de hipertensão pulmonar, embolismo pulmonar, regurgitação mitral severa, síndrome de *Wolff-Parkinson-White*, entre outros.
- **Fixo** – o desdobramento não altera em função da respiração e é audível durante a expiração. Indicador de comunicação interatrial ou falha do ventrículo direito.
- **Paradoxal (ou inverso)** – o desdobramento é audível durante a expiração mas não durante a inspiração. Indicador de estenose aórtica, bloqueio do ramo esquerdo do feixe de His (LBBB), cardiomiopatia hipertrófica obstrutiva, entre outros.
- **Ausente** – não se verifica o desdobramento em nenhuma altura. Indicador de estenose aórtica severa ou defeito do septo ventricular com síndrome de *Eisenmenger*.

## 2.2. Sons Cardíacos Anómalos

Durante o ciclo cardíaco normal, um padrão S1-S2 com um período sistólico (S1-S2) e um período diastólico (S2-S1) é claramente discernível. No entanto, é possível que outros sons extra, como S3 e S4 (ritmos de galope), sons de ejeção (“*clicks*”) ou sopros (“*murmurs*”) ocorram.

### 2.2.1. Ritmo de Galope (S3 e S4)

Sons cardíacos extra ou ritmos de galope (devido à sua semelhança com o galope de um cavalo) podem ser fisiológicos (particularmente em jovens adultos, grávidas ou idosos) ou

indicadores de uma patologia subjacente. O termo galope é normalmente utilizado para descrever sons patológicos S3 e S4. Estes são gerados maioritariamente no lado esquerdo do coração. Os gerados no lado esquerdo são menos intensos com a inspiração, enquanto os gerados do lado direito são mais intensos.

O som S3 ocorre imediatamente após S2. É gerado pelo enchimento rápido da cavidade ventricular e consequente desaceleração do sangue quando as paredes ventriculares atingem o seu limite elástico. É considerado benigno em indivíduos jovens ou grávidas mas pode ser indicador de fraca função sistólica noutros casos como regurgitação mitral crónica, falha cardíaca ou cardiomiopatia.

O som S4 ocorre no final do período diastólico, pouco antes de S1. É resultado de uma contração auricular tardia cujo fluxo sanguíneo resultante choca contra uma pressão ventricular elevada. Pode ser fisiológico em idosos devido a reduzida complacência ventricular. Em outros casos é um indicador de uma fraca função diastólica como hipertrofia ventricular, cardiomiopatia isquémica, ou enfarte agudo do miocárdio [13] [14].

### 2.2.2. Sons de ejeção (“clicks”)

Sons de ejeção, ou *clicks*, são produzidos pelo movimento de válvulas anormais e podem ser categorizados da seguinte forma:

- **Click de Ejeção Aórtica** – som agudo e de curta duração provocado pela abertura de uma válvula aórtica rígida em estenose aórtica (estreitamento da válvula aórtica resultando na obstrução do fluxo sanguíneo do ventrículo esquerdo para a aorta durante a sístole). Ocorre imediatamente após S1. É usado para diferenciar um sopro sistólico patológico de um sopro benigno.
- **Estalido de Abertura da Válvula Mitral** – som agudo causado pela abertura de uma válvula mitral em estenose mitral (obstrução do fluxo sanguíneo entre a aurícula esquerda e o ventrículo esquerdo durante a diástole). Ocorre imediatamente após S2.
- **Click do Prolapso da Válvula Mitral** – prolapso da válvula mitral para a aurícula esquerda durante a diástole.
- **Click de Válvula Mecânica** – sons S1 e S2 produzidos por válvulas prostéticas. A ausência de um *click* em pacientes com válvulas mecânicas pode ser indicadora de um problema.

### 2.2.3. Sopros do Coração

Sons produzidos por um fluxo turbulento, fruto de acelerações e desacelerações do sangue nas câmaras cardíacas, ou por estreitamento ou incapacidade das válvulas devido a regurgitação, são denominados sopros. Maior parte dos sopros são eventos intracardíacos que apresentam uma

gama de frequências entre 20 e 600 Hz. Este espectro está significativamente sobreposto a outros eventos do ciclo cardíaco como os sons fundamentais e artefactos [15].

Os sopros têm uma natureza caótica. Estes são descritos com base na sua localização e radiação, tempo, intensidade, frequência, configuração e resposta a manobras. Podem ser classificados como funcionais ou patológicos [16] [17] [18].

A localização define a região na qual o sopro se manifesta. Estas regiões podem ser observadas na Figura 2.2 e categorizadas da seguinte forma:

- **Aórtica** – indicador de estenose ou regurgitação aórtica, ou coartação da aorta.
- **Pulmonar** – indicador de estenose ou regurgitação pulmonar, ou de comunicação interatrial.
- **Mitral** – indicador de estenose ou regurgitação mitral, ou prolapso da válvula mitral.
- **Tricúspide** – indicador de estenose ou regurgitação tricúspide.
- **Ponto de Gibson** – indicador de persistência do canal arterial.
- **Ponto de Erb** – indicador de regurgitação aórtica ou pulmonar no caso de sopros diastólicos ou cardiomiopatia hipertrófica no caso de sopro sistólico.

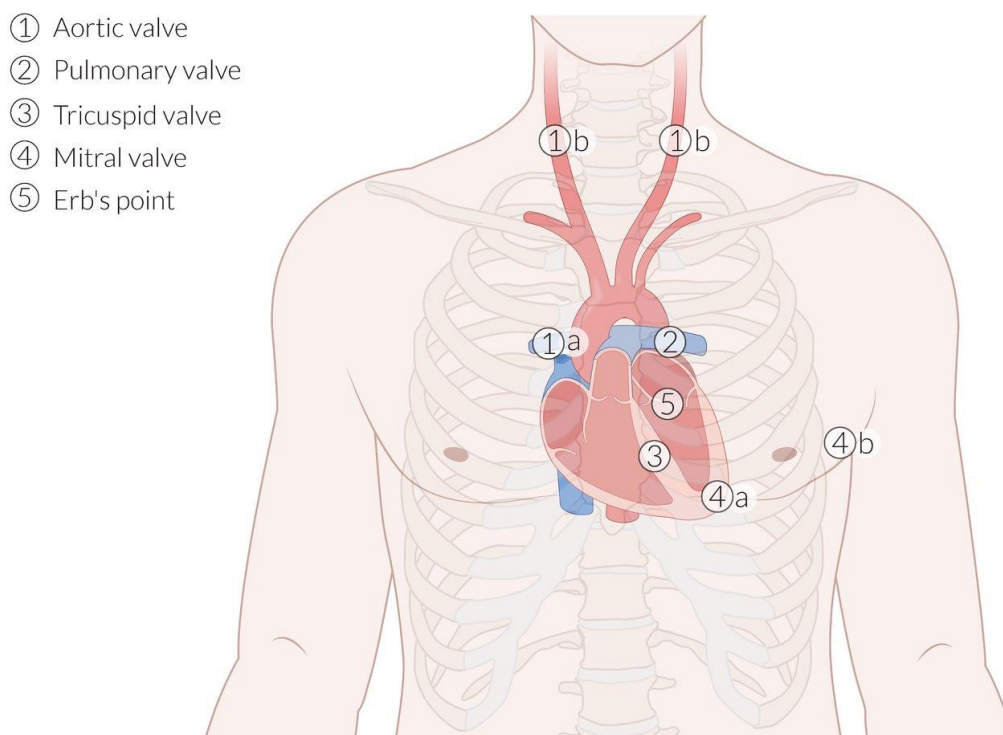


Figura 2.2 – Locais de auscultação das válvulas cardíacas. Válvula aórtica: 2º espaço intercostal direito, linha paraesternal (1a), artérias carótidas (1b); Válvula pulmonar: 2º espaço intercostal esquerdo, linha paraesternal (2); Válvula tricúspide: 4º espaço intercostal direito, linha paraesternal (3); Válvula mitral: 5º espaço intercostal esquerdo, linha médio-clavicular (4a), axila esquerda (4b); Ponto de Erb: 3º espaço intercostal esquerdo, linha paraesternal [19].

Um sopro pode ser auscultado em pontos não posicionados diretamente sobre a região do coração afetada devido à forma de propagação (radiação) da onda sonora. Por exemplo, em estenose aórtica, o sopro sistólico propaga-se para as artérias carótidas, mas no caso de uma regurgitação mitral, este radia para a axila esquerda, e no caso de uma estenose pulmonar, irradia para a região interescapular.

O tempo define a altura em que o sopro ocorre. Estes podem ser sopros sistólicos (entre S1 e S2), diastólicos (entre S2 e S1) ou contínuos (ocorre tanto na sístole como na diástole). Uma representação dos sopros sistólicos pode ser encontrada na Figura 2.3 e dos diastólicos na Figura 2.4. Dentro de cada uma destas categorias podem definir-se subcategorias conforme o sopro se encontre no início, meio, fim ou em todo o intervalo. Sopros diastólicos, ao contrário dos sopros sistólicos, são quase sempre patológicos. Sopros sistólicos podem ser funcionais em crianças, grávidas, casos de anemia e febre, entre outros [16] [17].

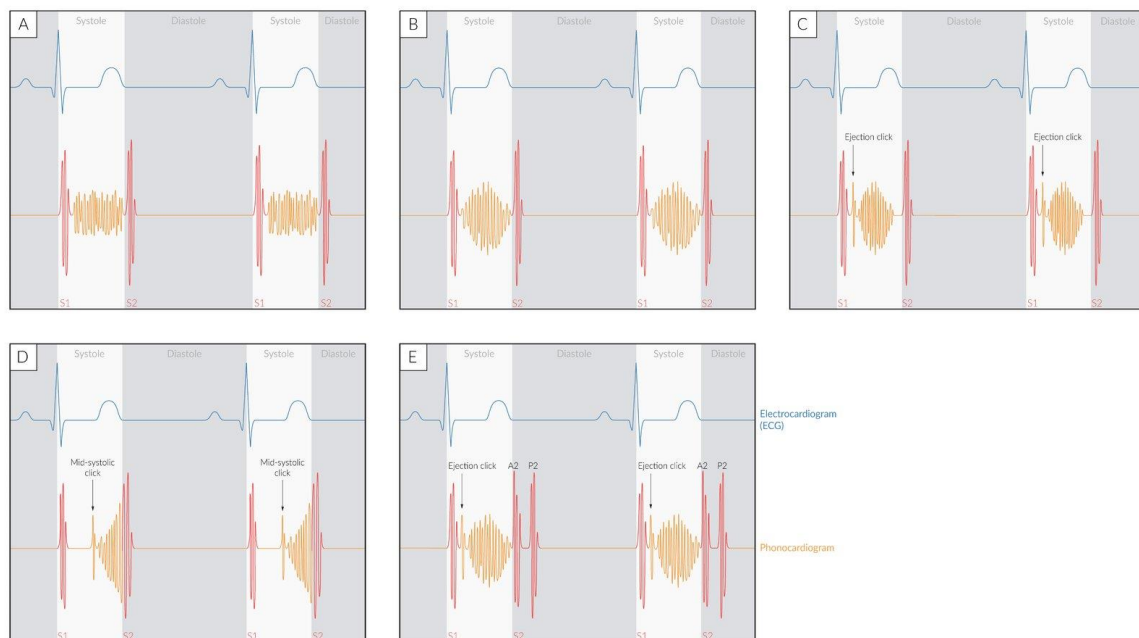


Figura 2.3 – Sopros Sistólicos. (A) Insuficiência mitral/tricúspide e/ou defeito no septo ventricular; (B) Estenose aórtica/pulmonar; (C) Estenose aórtica/pulmonar com click de ejeção; (D) Prolapso da válvula mitral; (E) Estenose pulmonar com desdobramento de S2 [19].

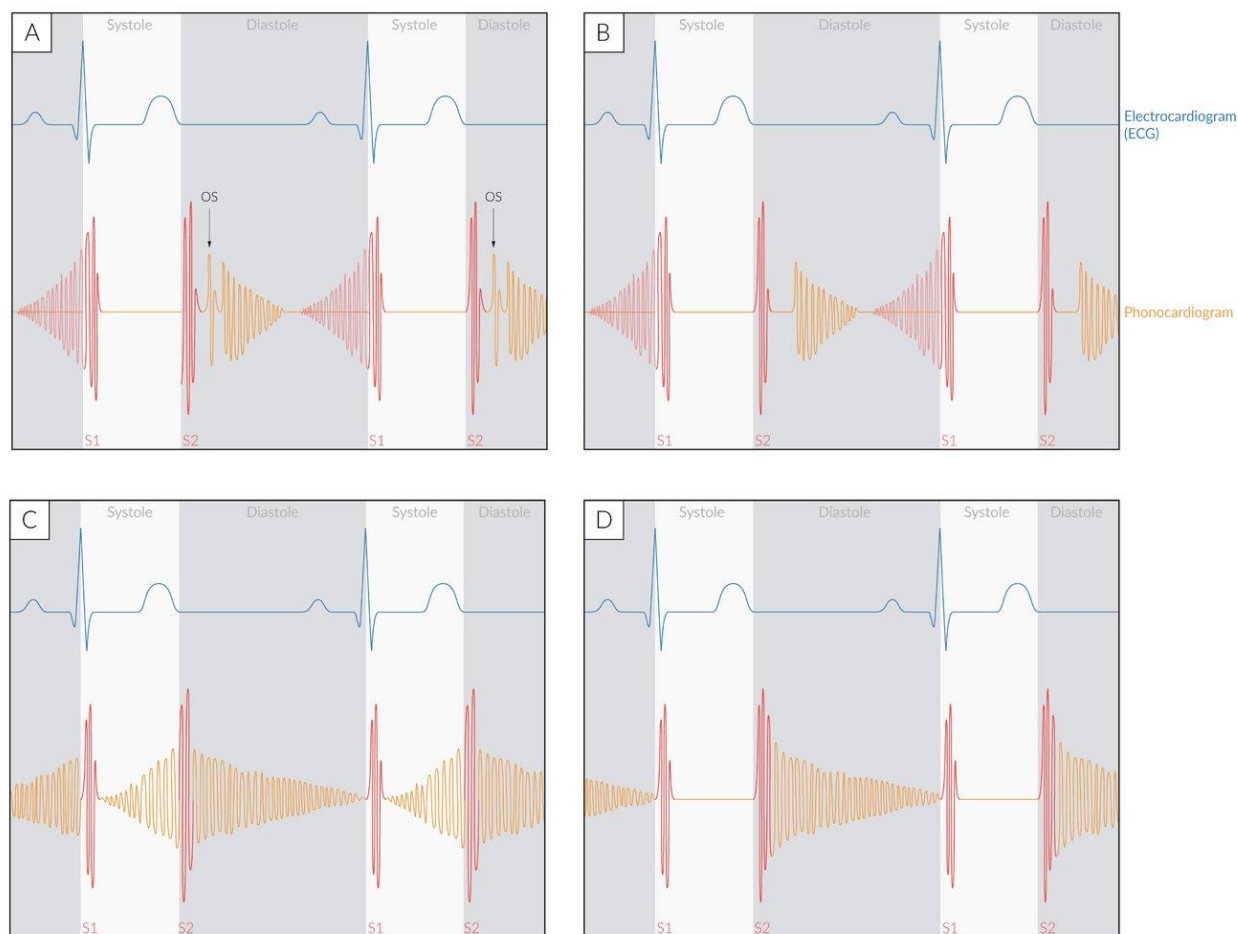


Figura 2.4 – Sopros Diastólicos. (A) Estenose mitral; (B) Estenose tricúspide; (C) Persistência do canal arterial; (D) Regurgitação aórtica e pulmonar [19].

A intensidade de um sopro pode ser categorizada em 6 níveis diferentes. Apesar de maior parte dos sopros que apresentem uma intensidade de nível III ou superior serem patológicos, uma maior intensidade não está diretamente correlacionada com a gravidade da lesão. Sopros de intensidade igual ou superior a nível II requerem uma ecocardiografia para avaliar as lesões.

A variação da intensidade de um sopro é denominada a sua configuração. Esta pode ser categorizada da seguinte forma:

- **Uniforme** – intensidade não varia.
- **Crescendo** – intensidade aumenta com o tempo.
- **Decrescendo** – intensidade diminui com o tempo.
- **Crescendo-Decrescendo** – intensidade aumenta inicialmente, mas diminui de seguida.

Manobras podem ser aplicadas para verificar uma alteração da intensidade do sopro. Estas manobras incluem inspiração, manobra de Valsalva, colocar o paciente em diferentes posições, entre outras.

### 2.3. Tempos e Frequências no PCG

Os fenômenos que originam o sinal PCG estão intrinsecamente relacionados com os que originam o sinal ECG. São os impulsos elétricos registados no ECG que originam as contrações dos músculos cardíacos e fecho das válvulas que por sua vez originam as vibrações que são sons registados no PCG.

No ECG, a ativação das aurículas corresponde à onda P; a ativação dos ventrículos corresponde ao complexo QRS; o período diastólico corresponde à onda T. O som S1 é gerado pela contração dos ventrículos e ocorre ao mesmo tempo que o complexo QRS. O som S2 é gerado pelo fecho das válvulas aórtica e pulmonar e ocorre no final da onda T. Estas propriedades foram utilizadas em diversos trabalhos para obter uma melhor segmentação e os mesmos serão aprofundados no capítulo 3.

Para processamento do PCG é útil conhecer as gamas de frequências típicas de cada componente, bem como as suas durações. A tabela 2.1 estabelece um resumo dessas mesmas durações e frequências [23].

Tabela: 2.1 – Fisiopatologia de Sons Cardíacos Normais e Anómalos [23].

Heart Sounds	Frequency Range	Qualitative sounds characteristics	Duration/location in the cardiac cycle	Cause	Description			
<i>First heart sound (S1) or lub sound</i>	10-200 Hz (lower pitch than S2)	Dull and prolonged	0.12-0.15 seconds (longer than S2)	Closure of the atrioventricular valves	Composed of M1 and T1 components			
<i>Second heart Sound (S2) or dub sound</i>	20-250 Hz (higher pitch than S1)	Sharp and short	0.08-0.12 seconds (shorter than S1)	Closure of the semilunar valves (at the end of systole)	Composed of A2 and P2 components			
<i>Third heart sound (S3)</i>	25-70 Hz (very low frequency, lower pitch than S1 and S2)	Soft and thudding quality	≈0.04 s, early-diastole (140-220 ms after S2)	Early diastolic filling of the ventricle by blood rushing in from the atria	Due to the excess blood volume in the ventricle (left ventricle failure), benign in children and in pregnancy			
<i>Fourth heart sound (S4)</i>	15-70 Hz (lower than S3, low pitch)	Weak and rumbling, less loud than S1 or S2	Late-diastolic/pre-systolic, slightly before S1	Diastolic dysfunction because of the stiff ventricle	Due to atrial contraction, manifests coronary heart disease			
<i>Gallop</i>	<i>Ventricular (S3 gallop rhythm)</i>	15-50 Hz (very low frequency, low pitch, short and faint)	Galloping rhythm, lilt, trot or canter quality	≈ 0.15 s after S2, early-diastole	During S3 due to rapid deceleration of blood flow into the ventricle	Gallop rhythms indicate serious myocardial dysfunction		
	<i>Atrial (S4 gallop rhythm)</i>			≈ 0.08 – 0.20 s just before S1			During S4, due to decreased ventricular distensibility	<i>S4-S1 may be confused with a split S1</i>
	<i>Summing (both S3 and S4 gallops)</i>			Quadruple rhythm, loud sound			During diastole period	Both S3 and S4 are superimposed

<i>Murmurs</i>	<i>IM</i>	120-250 Hz (mid-range frequency, high pitch)	Whooshing, roaring, turbulent fluid noise	Mostly in early-systole, short duration	Due to the turbulent flow of the blood which may occur inside or outside the heart	Frequently heard in children and during exercise, position dependent
	<i>SM</i>	Up to 600 Hz (frequency range usually lower than DM, high pitch)	Rasping and blowing, usually crescendo-decrescendo	Early-, mid-, late- or holo-systolic	Mechanical systolic and ventricular ejection	Includes PS, AS, ASD, HOCM, MI, TI, MVP, PDA and VSD
	<i>DM</i>	Up to 600 Hz (high or low pitch)	Puffing and rumbling quality, usually decrescendo	Early-, mid- or late-diastolic	Ventricular relaxation and filling	Difficult to hear, Includes PI, AI, MS, TS
<i>Opening snaps</i>		100-800 Hz (high pitch)	Snapping sound	During diastole period	Inspissating of valve leaflets	The sudden opening of the stiff mitral valve
<i>Rubs</i>			Scratching, harsh, creaking sound	Loudest in systole and can be heard in the beginning and end of diastole	Friction between layers, abrasion of pericardial surfaces	Depends on the body position and breathing
<i>Clicks</i>			Short and loud	Very early systole, mid-systolic click, may be immediately after S <sub>1</sub>	Due to the opening of a rigid and calcifies aortic or pulmonary valve	Includes AEC, SEC, mechanical valve click, or prosthetic valve click

\*Information regarding frequency range and timings from [23].

Abbreviation: IM: innocent murmurs, SM: systolic murmurs, DM: diastolic murmurs, PS: pulmonary stenosis, AS: aortic stenosis, ASD: atrial septal defect, HOCM: hypertrophic obstructive cardiomyopathy, MI: mitral insufficiency, TI: tricuspid insufficiency, MVP: mitral valve prolapse, PDA: patent ductus arteriosus, VSD: ventricular septal defect, PI: pulmonary insufficiency, AI: aortic insufficiency; TS: tricuspid stenosis, MS: mitral stenosis; AEC: aortic ejection click, SEC: systolic ejection click.

### 3. Redes Neurais

Entre os métodos aplicados na classificação automática dos PCG encontrados na literatura, destacam-se as redes neurais (*artificial neural network*, ANN). As redes neurais são sistemas inspirados no sistema nervoso humano que procuram replicar o modo como os humanos aprendem através do reconhecimento de padrões.

Estes sistemas obtiveram bons resultados quando aplicados em diferentes problemas como regressão e classificação, visão por computador, análise de séries temporais, extração de características, detecção de padrões, etc.

#### 3.1. Estrutura Básica das ANN

As ANN são compostas por múltiplos nodos inspirados no neurónio biológico. Por semelhança a um neurónio, um nodo é ativado quando recebe estímulo suficiente. Um nodo combina os dados de entrada com uma série de coeficientes, ou pesos, que ampliam ou atenuam essa entrada em função da sua significância face à tarefa para a qual o algoritmo está a ser treinado.

Uma única camada de uma ANN é designada perceptrão (Fig. 3.1) e origina uma única saída. Esta é uma combinação linear dos valores das entradas com os pesos a não ser que seja adicionada uma função de ativação, que tornará a saída não-linear.

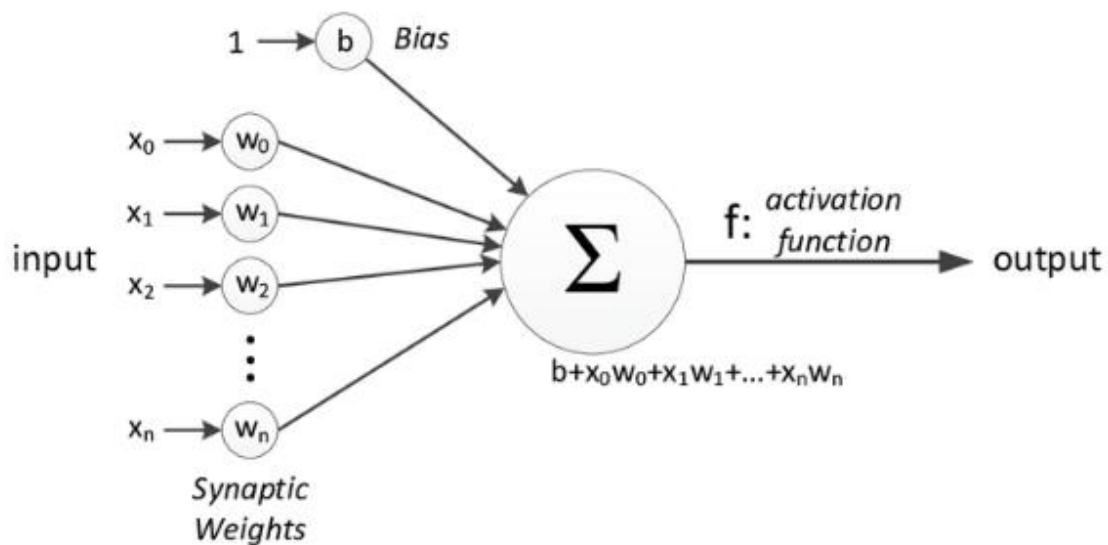


Figura 3.1 – Perceptrão (adaptado de D. Heger, 2015 [90])

Na figura acima,  $x_0, x_1, x_2, \dots, x(n)$  representam diversas entradas (variáveis independentes). Cada entrada é multiplicada por um peso ou sinapse  $w_0, w_1, w_2, \dots, w(n)$ . Os pesos refletem a influência de uma certa entrada.

$b$  é o valor de *bias*. Este valor permite mover a função de ativação.

Nos casos mais simples, estas entradas são somadas e alimentadas à função de ativação que gera a saída. A função de ativação determina se um neurónio deve ou não ser ativado, introduzindo não-linearidade ao sistema. Sem a função de ativação a saída seria determinada por uma função linear o que não satisfaz certas exigências de complexidade de diversos problemas.

A fórmula matemática de um perceptrão é dada por:

$$y = \varphi \left( \sum_{i=1}^n w_i x_i + b \right) = \varphi ( w^T x + b ) \quad (3.1)$$

Onde  $w$  é o vetor de pesos,  $x$  o vetor de entradas,  $b$  o *bias* e  $\varphi$  a função de ativação.

### 3.2. Aprendizagem nas ANN

As ANN podem agrupar e classificar as suas entradas de diferentes maneiras. Existem três principais tipos de aprendizagem:

- **Supervisionada** – é apresentada à rede a correspondência entre uma entrada e uma saída. A rede calcula a diferença entre o sinal gerado e o desejado para ajustar os pesos da rede.
- **Não-Supervisionada** – a rede é treinada de modo a responder a agrupamentos de padrões de entrada (ou “*clusters*”). É necessário quando não exista uma base de dados com as respostas desejadas.
- **Reforço** – é uma estratégia construída a partir da observação. Se o ambiente responder negativamente os pesos serão alterados de acordo. A saída só é conhecida a partir de certa altura do treino.

A aprendizagem supervisionada é a aplicada no caso de estudo e será a discutida no resto do documento até indicação em contrário.

### 3.3. Treino de uma ANN

O treino de uma ANN é feito através do ajuste dos pesos e *bias*. Este é calculado da diferença entre o valor gerado e o desejado, ou seja, é um valor de erro conhecido como **função custo** (metade do quadrado da diferença entre o valor gerado e o esperado). Este valor é computado e enviado de volta ao sistema. O objetivo do ajuste dos pesos é minimizar a função de custo para que o valor gerado e desejado se aproximem.

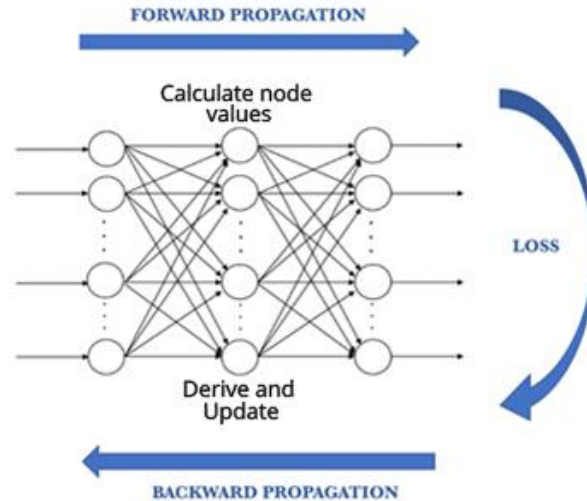


Figura 3.2 – Método de Back-propagation (adaptado de D. Heger, 2015 [90])

O processo descrito é denominado propagação retroativa ou *back-propagation* (Fig. 3.2) e é aplicado continuamente até à função custo atingir um mínimo.

Os métodos acima descritos estão extensivamente debatidos e podem ser encontrados na literatura [24] [25].

### 3.4. Redes Neurais Convolucionais

As redes neuronais convolucionais (*convolutional neural networks*, CNN) são principalmente utilizadas em reconhecimento de objetos, identificação e agrupamento de imagens. Partindo do pressuposto que a entrada é uma imagem (ou vídeo), a arquitetura deste tipo de redes é adaptada para reduzir o número de parâmetros, permitindo lidar com maiores bases de dados, e manter a dependência espacial e temporal da entrada.

A função de uma CNN é estabelecer automaticamente uma hierarquia de características relevantes, extrair a sua representação e integrá-la no problema de classificação. As camadas iniciais são de baixo nível, ou seja, captam características mais generalizáveis (Fig. 3.3). As camadas intermédias e finais tratam de características de alto nível, sendo que as representações destes níveis possuem maior especificidade.

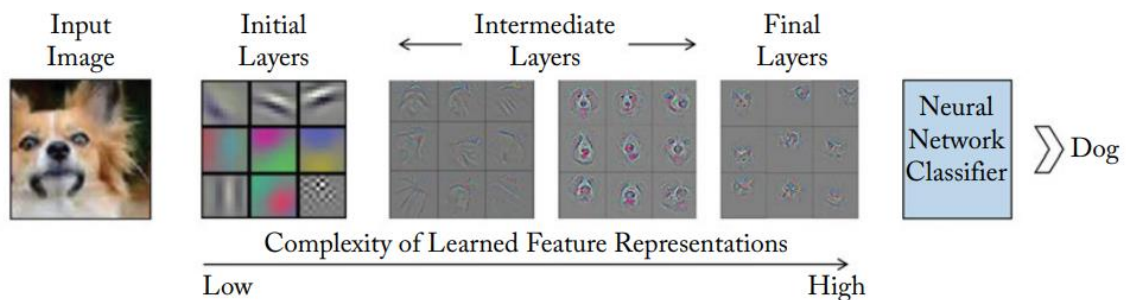


Figura 3.3 – Visualização de diferentes representações de características em diferentes camadas de uma CNN. Adaptado de Zeiler and Fergus [26].

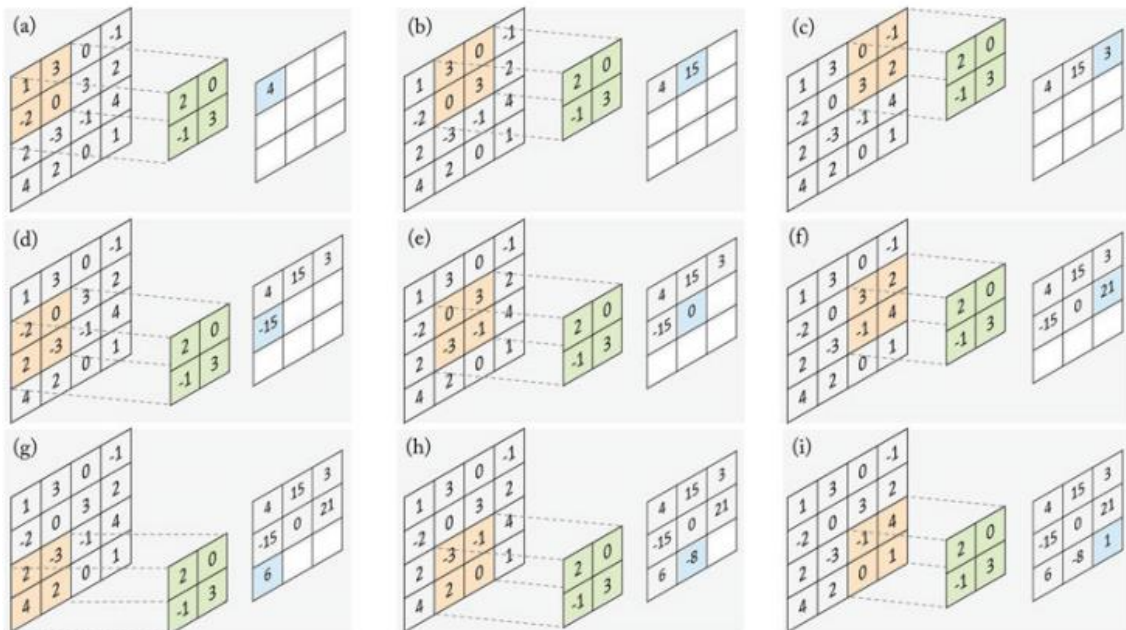
### 3.4.1. Camadas de uma CNN

Ao utilizar uma série de camadas diferentes, arquiteturas complexas são desenvolvidas para problemas de classificação. Os quatro tipos de camadas mais comuns são: convolucionais, *pooling/sub-sampling*, camadas não-lineares e totalmente ligadas.

#### 3.4.1.1. Convolutacional

A camada convolutacional é o componente mais importante de uma CNN. É composta por uma série de filtros ou *kernels* que serão utilizados para realizar a convolução com a entrada. Cada filtro contém números discretos (ou pesos) organizados numa grelha que serão atualizados com o treino da rede.

De modo a calcular o mapa de saída, o filtro é movido ao longo de uma coluna ou linha da entrada como representado na Figura 3.4. O número de linhas ou colunas que o filtro avança em cada iteração é denominado *stride*. Este é um hiperparâmetro que deve ser definido pelo arquiteto da rede. Além do *stride*, são também definidos pelo arquiteto o *padding* (atribui valores ao limiar da entrada para evitar perdas de informação), a profundidade (número de filtros utilizados na convolução) e a taxa de dilatação (usada para expandir o campo recetivo) [27].



O caso representado acima a 2D é o mais simples, onde o filtro tem um único canal representado com uma matriz. Quando a entrada for um tensor (p.ex. volumes 3D), os filtros serão também formas a três dimensões que geram um mapa de saída, por sua vez, também de três dimensões. As propriedades discutidas são aplicáveis a casos de dimensão superior [27].

### 3.4.1.2. Pooling/Sub-sampling

Estas camadas operam nos mapas de entrada e reduzem a sua resolução. Isto torna os mapas mais robustos contra ruídos e distorções [28] e a representação compacta é invariante face a alterações como a escala, sentido ou posição de um objeto numa imagem [29]. A operação de pooling pode ser feita a partir da média ou do máximo de uma entrada (Fig 3.5).

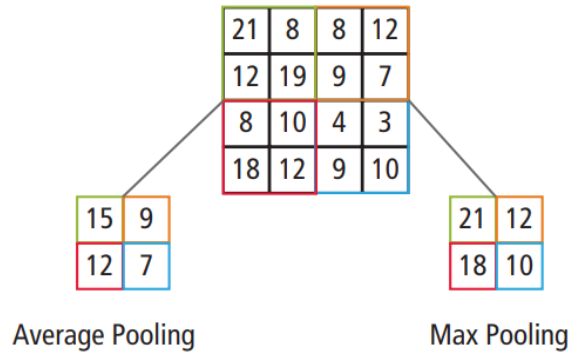


Figura 3.5 – Representação dos dois modos de pooling [28].

### 3.4.1.3. Não-Linearidade

As camadas com pesos numa CNN (camadas convolucionais e totalmente ligadas) são frequentemente seguidas por uma função de ativação não-linear. A função de ativação toma valores e distribui-os numa escala menor como  $[0,1]$  e  $[-1,1]$ . Além disso, atribui não-linearidade à rede.

Estas funções atuam como mecanismos de interruptor, decidem se um nodo deve ser ativado em função das suas entradas. As funções utilizadas são diferenciáveis de modo a permitir a propagação retroativa ou *back-propagation*. Na Fig. 3.6 podem ser encontradas as funções mais utilizadas.

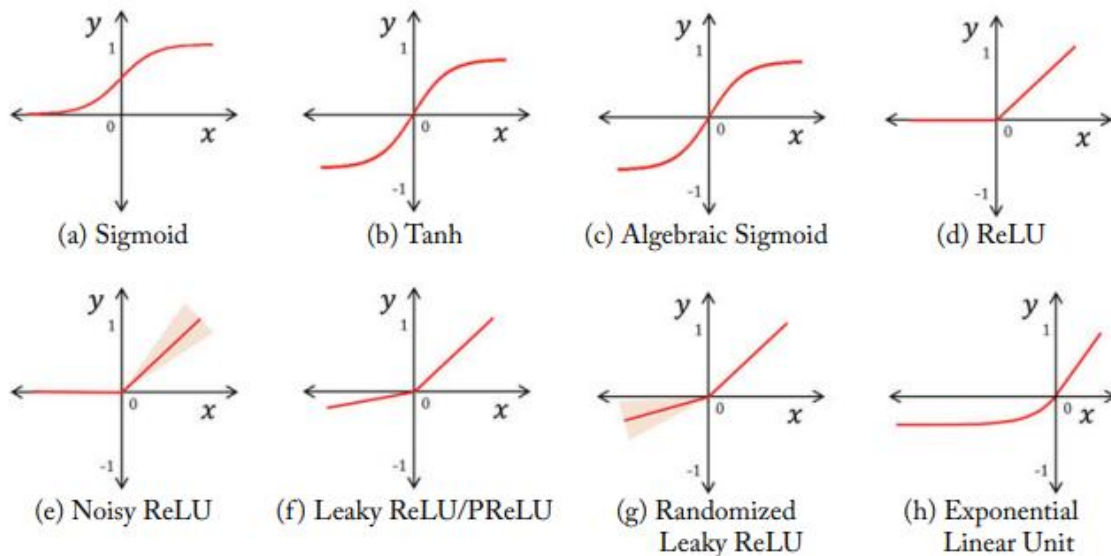


Figura 3.6 – Algumas das funções de ativação mais utilizadas nas CNN [27].

#### 3.4.1.4. Totalmente Ligadas

São, normalmente, as camadas finais de uma CNN, apesar de existirem casos na literatura onde a localização desta é intermédia. Correspondem essencialmente a filtros convolucionais  $1 \times 1$ . Cada unidade está ligada a todas as unidades da camada anterior. A sua operação pode ser representada como uma matriz de multiplicação seguido da soma de um vetor de *bias* e aplicação de uma função não-linear:

$$y = f(w^T x + b) \quad (3.2)$$

Onde  $x$  e  $y$  são os vetores de entrada e saída, respetivamente,  $w$  denota a matriz que contém os pesos das conexões entre unidades das camadas, e  $b$  representa o termo de *bias*.

Uma explicação mais detalhada destes tópicos pode ser encontrada em [28].

## 4. Estado da Arte

Algoritmos de classificação automática dos batimentos cardíacos já são estudados há cerca de 50 anos. No entanto, a carência de bases de dados de livre acesso, de qualidade, rigorosamente validadas e estandardizadas, dificulta o desenvolvimento de (e comparação entre) algoritmos eficientes e aplicáveis à variedade de sinais PCG existente.

O *2016 PhysioNet/Computing in Cardiology (CinC) Challenge* abordou este problema reunindo bases de dados de diferentes equipas de investigação, em ambiente clínico e não-clínico, em diferentes locais de aquisição e com diferentes níveis de ruído [5]. É até hoje a maior base de dados de livre acesso de sinais PCG [30]. O desafio consistia em desenvolver um algoritmo de classificação de sinais PCG, testá-lo na base de dados reunida e obter a maior percentagem de classificação correta possível. Uma descrição do desafio pode ser encontrada em [31].

Desde então, esta base de dados tem sido utilizada por diversas equipas de investigação como ponto de comparação entre algoritmos. Não obstante, existem outras bases de dados de menor dimensão utilizadas para o mesmo efeito como é o caso da *PASCAL, Open Michigan Heart Sound & Murmur Library (OMHSML), Cardiac Auscultation of Heart Murmurs database (eGeneralMedical)*, entre outras.

Os algoritmos desenvolvidos distinguem-se principalmente pelos métodos utilizados no pré-processamento, segmentação, seleção e extração de *features* e classificador utilizado (ilustrado na Figura 4.1). O seguinte subcapítulo, 4.1, apresenta alguns dos métodos mais utilizados nas diferentes fases dos algoritmos que surgiram nos últimos anos. Uma comparação mais detalhada dos métodos apresentados pode ser encontrada em [23].

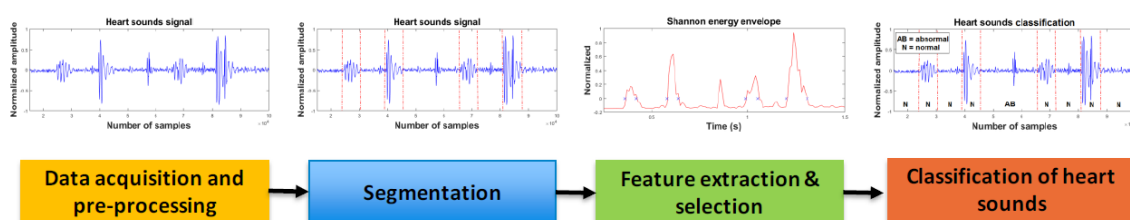


Figura 4.1 – Passos envolvidos na classificação de sons cardíacos [23].

### 4.1. Métodos Aplicados na Classificação de PCG

#### 4.1.1. Pré-Processamento

Esta fase trata principalmente de separar o sinal PCG de todos os ruídos inerentes à sua aquisição e posteriormente realizar as transformações necessárias ao sinal para o processamento.

O PCG é contaminado por fontes fisiológicas como a respiração e atividade muscular. Além destes ruídos também se manifestam os inerentes ao equipamento e ao ambiente como a interferência eletromagnética, ruído ambiental, artefactos, entre outros.

Vários trabalhos têm sido desenvolvidos de modo a aumentar o SNR (*signal-to-noise ratio*). De entre os inúmeros métodos de redução de ruído destacam-se métodos de detecção de ruído por similaridade espectral (SSA) [32], por periodicidade do sinal [33], *principal component analysis* (PCA) associado a estruturas de aprendizagem [34], métodos *wavelet* (*discrete wavelet transform* (DWT) [35], *tunable Q-wavelet transform* (TQWT) [36], entre outras variantes), *empirical mode decomposition* (EMD) [37] e *Mel-frequency cepstral coefficients* (MFCC) [38].

#### 4.1.2. Segmentação

O objetivo da segmentação do PCG é identificar os sons fundamentais (S1 e S2) e os componentes S3 e S4, se presentes. De um modo geral, os métodos de segmentação podem ser englobados em métodos de envelope [38]–[41], ECG ou pulso de referência [20][21][42], modelos probabilísticos [43][44], modelos baseados em *features* [37][45][46], análise nos domínios tempo-frequência [46][48] e métodos de aprendizagem [49][50].

*Springer et. al* [51] abordou o problema de segmentação de sinais PCG ruidosos utilizando um modelo de regressão logística de semi-*Markov* oculto (*logistic regression hidden semi-Markov model* - HSMM) e obteve uma percentagem de classificação de  $95.63 \pm 0.85\%$  sendo considerado o estado da arte. No entanto, o algoritmo não contempla as componentes S3 e S4 e o facto de incorporar informação *a priori* de durações expectáveis do ciclo cardíaco torna-o suscetível a casos de arritmia cardíaca.

#### 4.1.3. Extração de *Features*

A representação de sinais cardíacos em diferentes domínios revela características diferenciadoras entre sinais fisiológicos e patológicos e permite posterior extração das mesmas. Representações do sinal no domínio tempo-frequência têm sido preferidas às que contemplam apenas um dos domínios. [23] É o exemplo de representações que fazem uso da transformada S [52], transformada de *Fourier* de curto tempo (STFT) [53], distribuição *Choi-Williams* (CWD) [54], transformada modificada de *Hilbert* de tempo curto (STMHT) [40], transformada *wavelet* (*discrete wavelet transform* (DWT) [21], *continuous wavelet transform* (CWT) [55], *Mel-scaled wavelet transform* [56]).

#### 4.1.4. Classificação

O passo final consiste em escolher um classificador e introduzir-lhe as *features* selecionadas pelo passo anterior. Não existe um consenso acerca de qual o melhor classificador. As diferentes abordagens de classificação incluem *support vector machines* (SVM) [41][57], *Hidden Markov Models* (HMMs), *k-Nearest Neighbors* (k-NN) [9], redes neuronais (NN) [52],

árvores de decisão [59], *BayesNet* [60], métodos de *machine learning* [20][61][62], *random forest* [63], entre outros.

## 4.2. Trabalhos Desenvolvidos

Os próximos subcapítulos apresentam os trabalhos desenvolvidos pelos primeiros três classificados no *2016 PhysioNet/Computing in Cardiology (CinC) Challenge*, bem como outros que foram surgindo das aprendizagens retiradas do mesmo.

### 4.2.1. 2016 PhysioNet Challenge Top 3

#### 4.2.1.1. Potes et al. (2017)

*Potes et. al* (2017) propuseram um algoritmo que classifica batimentos normais/anormais através de uma votação pesada de duas redes neuronais diferentes (ilustrado na Figura 4.2). O algoritmo obteve um valor de *MAcc* (*modified accuracy*) de 0.8602 sendo o melhor classificado do desafio.

Para o primeiro classificador, cada amostra de PCG foi reamostrada para uma frequência de 1000Hz e filtrada entre 25 e 400Hz. Os sons cardíacos foram segmentados recorrendo ao algoritmo desenvolvido por *Springer et al.* [51]. Features no domínio tempo-frequência foram alimentadas ao classificador *AdaBoost*.

Para o segundo classificador, baseado em CNN, cada PCG foi decomposto em quatro bandas de frequência (20-45, 45-80, 80-200, 200-400 Hz) e segmentado nas diferentes fases do ciclo cardíaco. Cada CNN era composto por 3 camadas: a camada de *input* e duas convolucionais.

O algoritmo de votação final consistia em estabelecer um limiar para os dois classificadores, fosse um deles ou ambos ultrapassado e o sinal seria considerado anormal.

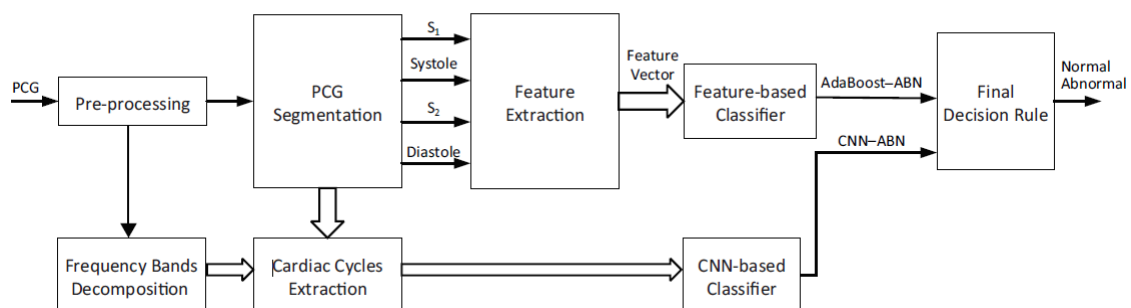


Figura 4.2 – Diagrama de blocos do algoritmo proposto por Potes et. al [64].

#### 4.2.1.2. Zabihi et al. (2017)

*Zabihi et. al* (2017) propuseram um algoritmo de classificação de sinais PCG sem recorrer a segmentação (janelas assíncronas). A classificação não procurava apenas distinguir

PCG normais de anómalos, como também os de boa e fraca qualidade. O algoritmo obteve um valor de MAcc de 0.8590 sendo o segundo melhor classificado do desafio.

Um conjunto de 18 *features*, de um total de 40, contidas no domínio do tempo, frequência e tempo-frequência, foram selecionadas e alimentadas a um aglomerado de 20 redes neurais artificiais que produzem uma decisão baseada num algoritmo de votação [65] como representado na Figura 4.3.

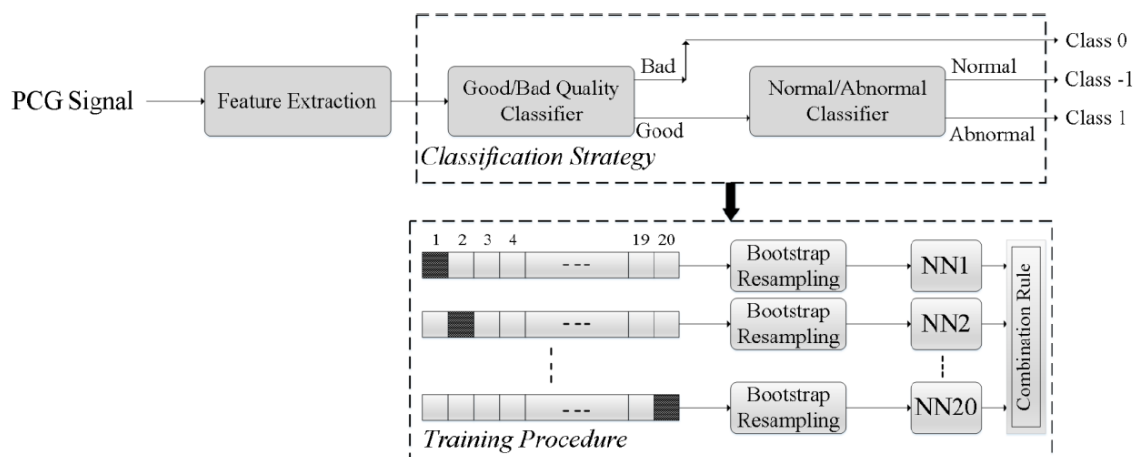


Figura 4.3 – Esquema do algoritmo de classificação proposto por Zabihi et. al. [59]

#### 4.2.1.3. Kay & Agarwal (2017)

Kay & Agarwal (2017) propuseram um algoritmo que utiliza uma rede neuronal *DropConnected* [66] treinada com *features* contidas no domínio tempo-frequência e outras inter-batimentos. O algoritmo atingiu um MAcc de 0.8520 e foi o terceiro classificado do desafio.

Os sons cardíacos foram segmentados recorrendo ao algoritmo desenvolvido por Springer et. al [51]. Um total de 675 *features* foram retiradas de uma análise CWT (220), MFCC (400), inter-batimentos (20) e medidas de complexidade (35). As *features* extraídas foram depois normalizadas e a dimensionalidade reduzida para 50 usando uma técnica de PCA. Finalmente, as *features* foram alimentadas a uma rede neuronal com duas camadas ocultas, treinadas por alimentação inversa (*backpropagation*) e regularizadas através do *DropConnect* [67]. O algoritmo demonstrou ter dificuldades em generalizar de uma base de dados para outra [30].

#### 4.2.1.4. Aprendizagens do Desafio

Os algoritmos desenvolvidos apresentaram piores resultados quando testados em bases de dados novas, o que indica que a capacidade de generalização é sensível a diferentes métodos de aquisição e deve ser melhorada ou que os classificadores devem ser treinados de novo sempre que um novo método seja introduzido [30].

Os resultados apresentados pelos melhores classificados diferem apenas ligeiramente, o que indica que não existirá um melhor classificador. Métodos que recorrem a diferentes

classificadores conjuntos (*ensemble classifiers*) como utilizado por *Zabihi et al.*, podem aumentar a percentagem de acerto [30].

A inabilidade de classificadores mais elaborados (como é o caso das CNN) ultrapassarem resultados obtidos por classificadores mais simples e *features* seleccionadas indica que o foco deverá ser no pré-processamento e segmentação do sinal. Não significando, porém, que um classificador superior não melhore os resultados, mas que a escolha cuidada de *features* e o tratamento do sinal aparentam uma melhoria marginal mais significativa [30].

#### **4.2.2. Messner et al. (2018)**

Após os resultados obtidos, certos trabalhos redirecionaram a sua atenção para as fases de segmentação, remoção de ruído e extração de *features*. No primeiro caso, temos o exemplo de *Messner et al.* [68] que procuraram desenvolver um algoritmo de segmentação que não recorresse a informação *a priori* (como é o caso do algoritmo do estado da arte, *Springer et al.*). Retiraram *features* espectrais e de envelope dos PCG e investigaram a eficácia de diferentes *deep recurrent neural networks* (DRNN).

Os PCG foram reamostrados para uma frequência de 1000 Hz e filtrados por um passa-alto de frequência de corte de 10 Hz. Para as *features* espectrais, os sinais foram processados com uma *Short-Time Fourier Transform* (STFT) recorrendo a uma janela de *Hamming* com 80 ms e 75% de sobreposição. Daí foram retiradas *features* do espetograma (41) e dos MFCC (60). Todas as *features*, incluindo as de envelope, foram normalizadas.

Das redes investigadas, a que obteve o melhor resultado foi a *bidirectional gated recurrent neural network* (BiGRNN) com um valor  $F1 = 95,6\%$ . O método proposto é competitivo com o do estado da arte e apresenta melhores resultados especialmente nos casos de arritmia cardíaca. Mais informação acerca dos testes realizados pode ser encontrada em [68].

#### **4.2.3. Cheng et al. (2019)**

*Cheng et al.* procuraram desenvolver uma abordagem com baixa exigência de hardware e que pudesse ser estendida a terminais móveis através de redes neuronais lacónicas (LNN – *laconic neural network*) para classificação dos PCG. Os seus resultados foram testados na base de dados *Physionet* e obtiveram um  $Macc = 0.8950$ .

Primeiro, era analisado um espetograma do PCG e determinado o tamanho do mesmo. Em função do obtido, era escolhida uma rede neuronal baseada em três parâmetros: menor número de conexões (*sparse connection*), envelope e distribuição de frequência e energia quasi-periódica do PCG. As redes escolhidas procuravam otimizar o número de camadas ocultas necessárias para a rede convolucional. Uma descrição do método desenvolvido pode ser encontrado em [69] e uma ilustração das redes desenvolvidas está presente na Figura 4.4.

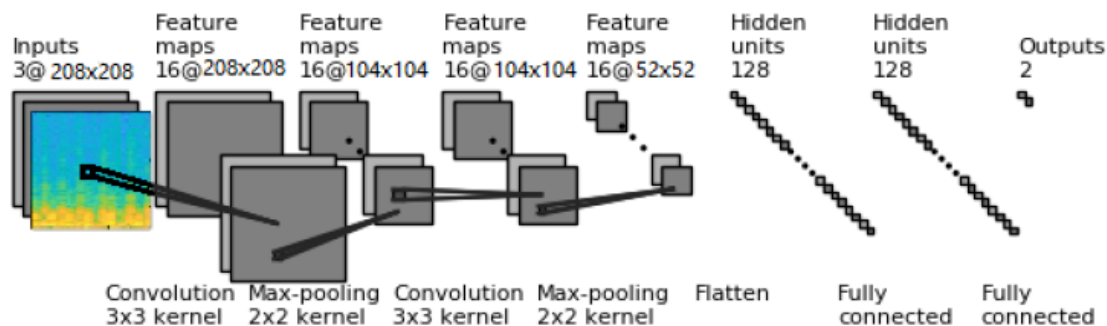


Figura 4.4 – Laconic Neural Network Structure [63].

#### 4.2.4. Zhang et al. (2019)

Zhang et al. desenvolveram um algoritmo que, por semelhança ao de Zabihi et al., não requer segmentação do PCG. Basearam-se em *features* quase-periódicas e *long short-term memory* (LSTM) [70]. O método foi testado na base de dados *Physionet* e obtiveram um  $Macc = 0.9008$ .

Primeiro, o espectrograma do PCG era obtido através de uma STFT. A seguir, *features* quase-periódicas eram calculadas através de uma função de diferença média de magnitudes (AMDF - *average magnitude difference function*). Das *features* retiradas era estudada a relação de dependência de curto e longo tempo através da LSTM (ilustrado na Figura 4.5).

As relações de dependência extraídas eram alimentadas a uma rede neuronal de duas camadas [71].

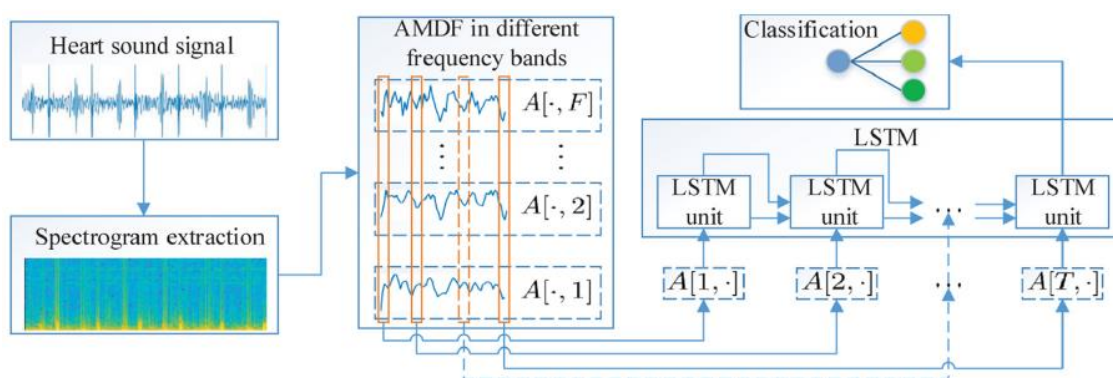


Figura 4.5 – Framework do método proposto em [64].

#### 4.2.5. Morales et al. (2019)

Morales et al. propuseram um algoritmo (cujo esquema está representado na Figura 4.6) baseado em CNN combinada com um sensor auditivo neuromórfico (NAS – *neuromorphic auditory sensor*) que decompõe o PCG em bandas de frequência [72]. A rede com melhor classificação é uma adaptação da rede *AlexNet* e atingiu um  $Macc = 0.9416$ .

O sinal PCG é segmentado em janelas temporais não sobrepostas de 1, 1.25 e 1.5 s gerando 77573, 61518 e 51009 amostras (CNN requerem um número vasto de amostras para serem robustas). As amostras são enviadas para uma plataforma de nodos-AER (ilustrado na Figura 3.7). O NAS está implementado na FPGA (*field-programmable gate array*) Spartan-6 dessa plataforma e decompõe o sinal em bandas de frequência. Estas são condensadas em pacotes de informação seguindo o protocolo AER (*Address-Event Representation*). Os ficheiros são enviados para um computador onde um *script* MATLAB os recebe e gera ficheiros AEDAT. Destes ficheiros são gerados sonogramas, cocleagramas, histogramas, entre outras *features* relevantes.

Várias redes foram treinadas recorrendo ao *Caffe* (*Convolutional Architecture for Fast Feature Embedding*) sendo que a arquitetura modificada da rede *AlexNet* foi a que obteve os melhores resultados. O autor salienta, no entanto, que esta requereu um tempo dezoito vezes superior a uma rede modificada LeNet-5 (de arquitetura mais simples) apresentando uma melhoria de classificação de apenas 3.5%.

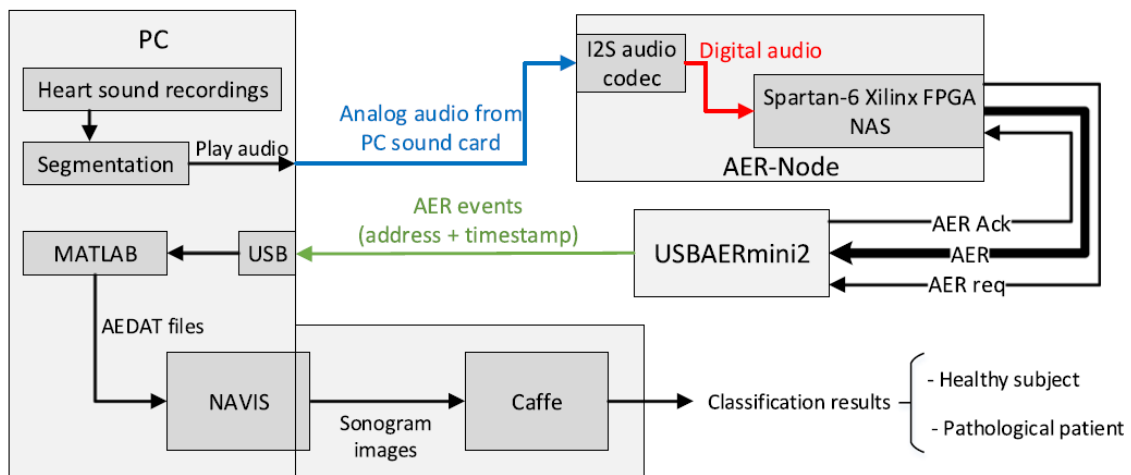


Figura 4.6 – Diagrama de blocos da arquitetura do sistema proposto em [66].

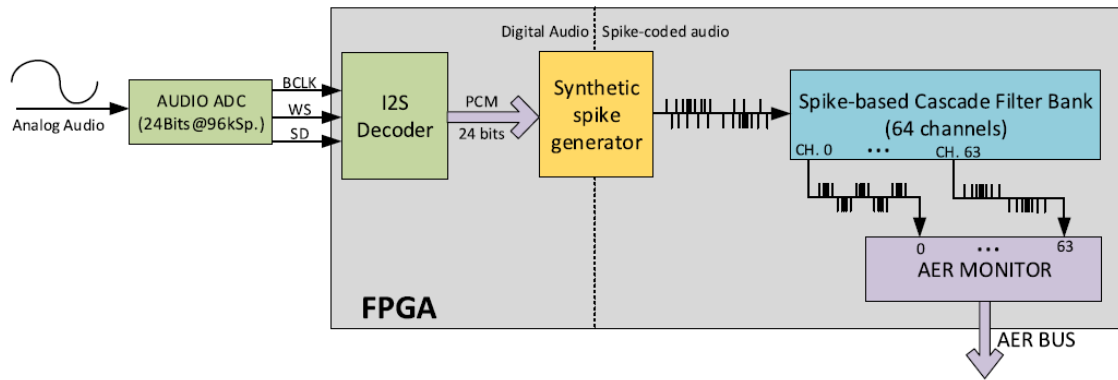


Figura 4.7 – Mono-Aural Neuromorphic Auditory Sensor for FPGA with an I2S audio ADC and AER interface [66].

#### 4.2.6. Revisão Literária

Além dos trabalhos supramencionados, várias equipas de investigação contribuíram para o melhoramento desta tecnologia com diferentes abordagens em cada uma das diferentes fases envolvidas no problema de classificação de sons cardíacos. As tabelas seguintes foram retiradas de um artigo de revisão literária publicado em Janeiro do ano de 2020 [73]. A primeira tabela, 4.1, compara a eficiência de diferentes métodos em função da extração e classificação de características do PCG. A tabela 4.2, por sua vez, estabelece uma comparação entre diferentes métodos que abordam o problema de classificação utilizando redes neuronais profundas (*Deep Neural Networks*).

Tabela: 4.1 – Métodos de extração de características e de classificação de sinais PCG [73].

Year	Author	Feature extraction methods	Classifier	Database	Result		
2019	Shi et al.	Feature extraction algorithm of Springer	AdaBoost	PhysioNet and PASCAL	ACC: 96.36%		
2019	Nogueira et al.	MFCC	SVM	PhysioNet	Sensitivity 91.87%	Specificity 82.05%	Accuracy 97%
2019	Cheng (without segmentation)	Envelope autocorrelation	SVM	HSCT11 dataset	Accuracy all could reach to 100%		
2018	Meintjes et al.	CWT	SVM, kNN	PhysioNet	MAcc: 86%		
				Dataset A from PhysioNet	MAcc: 92%		
2018	Hamidi et al.	Curve fitting, MFCC	Euclidean distance	Dataset B from PhysioNet	MAcc: 81%		
				Dataset C from PhysioNet	MAcc: 98%		

2018	Juniati et al.	DWT	kNN, Fuzzy c-means clustering	40 normal heart sounds, 40 extra systole, 40 murmurs	MAcc: 86.17%		
2017	Kay et al.	CWT, MFCC	BP neural networks	PhysioNet	MAcc: 85.2%		
2017	Karar et al.	DWT	Rule-based classification tree	22 sets of heart sounds and noise data from the public database of the CliniSurf medical school	MAcc: 95.5%		
2017	Zhang et al.	Tensor decomposition	SVM	Dataset A: normal heart sounds, extra systole, murmurs, artificial heart sounds	MAcc: 76%		
				Dataset B: normal heart sounds, extra systole, murmurs	MAcc: 83%		
2017	Langley and Murray (without segmentation)	/	Wavelet entropy	PhysioNet	Sensitivity	Specificity	Accuracy
					94%	65%	80%
2017	Whitaker et al.	Sparse coding	SVM	PhysioNet	Sensitivity	Specificity	MAcc
2017	Li et al.	FFT	BP neural networks	PhysioNet	84.3%	77.2%	80.7%
					68.36%	94.01%	88.56%
2017	Li et al.	FFT	Logistic regression	PhysioNet	Sensitivity	Specificity	MAcc
					75.68%	87.71%	72.56%
2016	Deng and Han (without segmentation)	DWT	SVM-DM	Dataset A from PASCAL	The highest total precision of 3.17		
				Dataset B from PASCAL	The highest total precision of 2.03		
2015	Zheng et al.	EMD	SVM	A dataset collected from the healthy volunteers and CHF patients	Sensitivity	Specificity	Accuracy
2015	Safara	Wavelet packet tree	Higher-order cumulants (HOC)	A set of 59 heart sounds from different categories: normal heart sounds, mitral regurgitation, aortic stenosis, and aortic regurgitation.	Best classification accuracies: 99.39%		
2011	Yuenyong et al. (without segmentation)	DWT	Neural network	Several on-line databases and recorded with an	Tenfold cross-validation: 0.92 for noise free case, 0.90 under white noise with 10 dB signal-to-noise ratio (SNR), and		

electronic  
stethoscope

0.90 under impulse noise up to 0.3 s  
duration

Tabela: 4.2 – Literatura de classificação de sons cardíacos aplicando Deep Learning [73].

Year	Author	Segmentation method	Dataset	Performance			
				Hold out testing			
				Sensitivity	Specificity	Accuracy	
2019	Wu et al.	CNN	PhysioNet (2575 normal heart sounds and 665 abnormal heart sounds)	86.46%	85.63%	86.0%	
				Ten-fold cross validation			
				Sensitivity	Specificity	Accuracy	
				91.73%	87.91%	89.81%	
2019	Abduh et al.	DNN	PhysioNet	Sensitivity	Specificity	Accuracy	
				89.30%	97%	95.50%	
2018	Gharehbaghi and Lindén	DTGNN	130 recordings of the heart sound signal	Sensitivity	Specificity	CR	
				83.9%	86%	85.5%	
2018	Chen et al.	DNN	PASCAL	Sensitivity	Specificity	Accuracy	Precision
				98%	88.5%	93%	89.1%
2018	Yaseen et al.	DNN	5 categories of heart sound signal, 200 per class (N, AS, MR, MS, MVP)	Sensitivity	Specificity		
				94.5%	98.2%		
2018	Han et al.	CNN	2575 normal recordings and 665 abnormal recordings	MAcc	Sensitivity	Specificity	
				91.50%	98.33%	84.67%	
2018	Ren et al.	CNN	PhysioNet	19.8% higher than the baseline accuracy obtained using traditional audio processing functions and support vector machines.			
2018	Morales et al.	CNN	PhysioNet	Accuracy	Sensitivity	Specificity	
				97%	93.20%	95.12%	
2018	Baris et al.	CNN	UoC-murmur database (innocent murmur versus pathological Murmur) and PhysioNet-2016 database (normal versus pathological)	MAcc	Specificity	Sensitivity	
				81.5%	78.5%	84.5%	
2018	Messner et al.	DNN	PhysioNet	F1 ≈ 96%			

2017	Ghaemmaghami et al.	DNN	128 recordings from male and female subjects with healthy hearts	Accuracy	Sensitivity	Specificity
				95.8%	83.2%	99.2%
					Accuracy	Precision
				RNN 4 layer	53.8%	55.8%
			Dataset A from PhysioNet	LSTM 4 layer	76.9%	83.3%
				GRU 4 layer	75.3%	78.2%
2017	Sujadevi et al.	RNN & LSTM&GRU			Accuracy	Precision
				RNN 4 layer	65.2%	68.1%
			Dataset B from PhysioNet	LSTM 4 layer	74.7%	94.5%
				GRU 4 layer	74.4%	69.7%
2017	Chen et al.	DNN	311 S1 and 313 S2 from 16 people (11 males and 5 females)		Accuracy: 91.12%	
2017	Yang and Hsieh	RNN	PhysioNet		MAcc: 84%	
			Dataset A from PASCAL		Normalized precision: 0.77	
2017	Zhang and Han	CNN	Dataset B from PASCAL		Normalized precision: 0.71	
2017	Faturrahman et al.	DBN	MITHSDB		Accuracy: 84.89%	
			AADHSDB		Accuracy: 86.15%	
					Train accuracy: 99.7%	
2017	Maknickas and Maknickas	CNN	PhysioNet		Validation accuracy: 95.2%	
2016	Thomae et al.	DNN	PhysioNet	Sensitivity	Specificity	Score
				96%	83%	0.89
2016	Tschannen and Dominik	CNN	PhysioNet	Sensitivity	Specificity	Score
				84.8%	77.6%	0.812
2016	Potes et al.	AdaBoost & CNN	PhysioNet	Sensitivity	Specificity	MAcc
				94.24%	77.81%	86.02%

Da revisão literária supramencionada, podemos concluir que maior parte dos trabalhos utiliza a totalidade ou parte dos sinais disponibilizados pelo *Physionet*, sendo esta a base de dados com maior número de amostras até à data. No entanto, como não existe uma base estabelecida como padrão, a comparação objetiva entre métodos torna-se impossível.

De entre os últimos trabalhos desenvolvidos, os que têm obtido melhores classificações, não obstante a base de dados utilizada, são os que utilizam CNN.

Tendo em conta as considerações anteriores, foram treinadas CNN recorrendo à base de dados *Physionet* e numa tentativa de reduzir o tempo de treino, bem como solucionar um problema das CNN que é o elevado número de amostras necessárias para o seu treino, importar bases convolucionais de redes de maior dimensão treinadas noutra problema de classificação, um método denominado *transfer learning*. As redes criadas são então treinadas em bases de dados de imagens de *features* retiradas dos PCG, sendo estas espectrogramas de Mel, MFCC (uma *feature* popular entre os trabalhos de classificação de sons cardíacos e largamente utilizada em processamento de áudio), bem como duas outras *features* que surgiram recentemente na literatura, os GTCC (*Gamma-tone-Cepstral Coefficients*) e PNCC (*Power-Normalized Cepstral Coefficients*).

## 5. Procedimento Experimental

O objetivo é implementar um algoritmo que processe os PCG, os converta em informação que possa ser alimentada a uma rede neuronal convolucional e obter uma classificação binária (batimento normal ou patológico).

A base de dados de PCGs disponibilizada pelo *PhysioNet* [75] foi equilibrada, ou seja, removeram-se sinais da classe em maior número até este corresponder ao da classe em minoria. Estes sinais foram filtrados e segmentados assincronamente para a fase seguinte.

Dos PCGs tratados foram criadas bases de dados de imagens, cada correspondendo a uma única *feature*. Estas podem ser: espectrogramas, MFCC, GTCC ou PNCC. As *features* foram guardadas como *heatmaps* em formato .png.

Após criadas as bases, elas são divididas aleatoriamente em grupo de treino e de validação. Estes são então usados para treinar uma rede que consiste numa base convolucional importada de uma rede pré-treinada acoplada a um classificador. Os pesos da base convolucional importada são mantidos enquanto o classificador tem os seus hiper-parâmetros alterados em cada teste.

De entre os classificadores testados, aquele que apresentar melhores resultados é mantido para uma segunda fase que envolve descongelar gradualmente as últimas camadas da base convolucional e deixar os pesos importados serem atualizados com o treino. O resultado deste treino é então comparado com as restantes redes criadas e a melhor combinação de *features*, base convolucional e classificador é escolhida.

Finalmente, a rede que gerar os melhores resultados de classificação é testada numa nova base de dados, cedida publicamente pelo grupo PASCAL [89].

As diferentes fases do procedimento experimental são descritas em pormenor nos subcapítulos seguintes. A figura 5.1 ilustra a metodologia proposta.

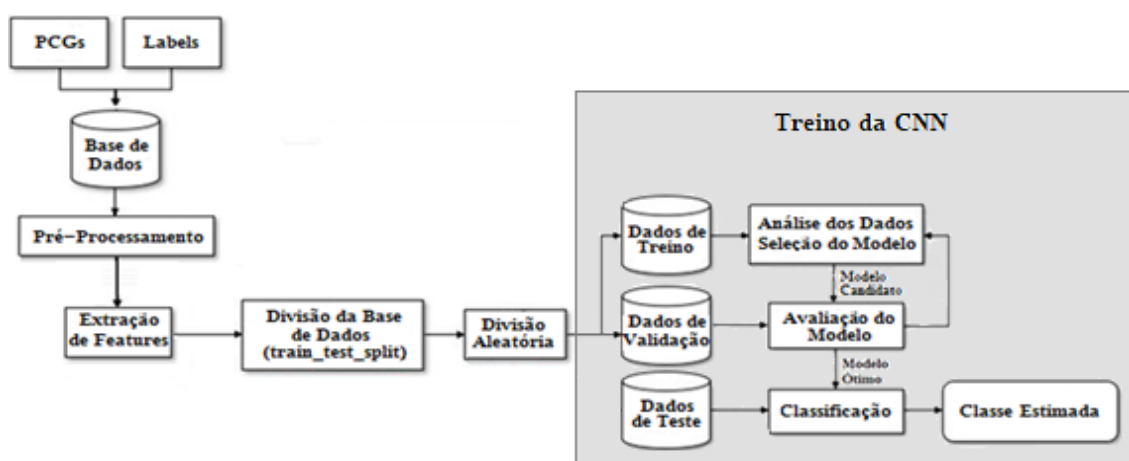


Figura 5.1 – Fluxograma do método proposto (adaptado de Braga et. al. [91]).

### **5.1. Hardware e Software**

Todo o processamento de dados e testes foram realizados num ambiente *Python* versão 3.6. O processamento de sinal foi facilitado por diversas bibliotecas das quais se destaca a maioritariamente aplicada no pré-processamento e análise dos sons, *Librosa* [68], versão 0.7.1, e para construção da CNN, *Tensorflow* [77], versão 1.4, em conjunto com a interface de programação de aplicações *Keras* [78], versão 2.0.

O tratamento do PCG bem como os primeiros testes com a CNN foram corridos utilizando um processador Intel(R) Core(TM) i5-3337U CPU@ 1.80GHz com 8GB de RAM. A dimensão da base de dados e o número de operações que decorrem na rede neuronal tornam este tipo de testes especialmente morosos. De modo a poder correr os testes na CNN num prazo menor, utilizou-se o *Google Colaboratory*, um serviço *online* que permite aceder a uma GPU Tesla K80 e escrever código *Python* em *Jupyter notebooks*.

### **5.2. Base de Dados**

Os PCG utilizados para criação das bases de dados foram disponibilizados publicamente pelo *Physionet* [75]. Os sons cardíacos foram recolhidos por diferentes equipas de investigação, em diferentes locais de aquisição, com diferentes níveis de ruído bem como equipamentos distintos. Isto confere à base de dados uma grande variabilidade dentro da mesma classe (sinal normal ou patológico). No decorrer da dissertação e até indicação em contrário, quando for referida a classe de um PCG será sempre sinónimo da sua classificação como normal ou patológico.

Os ficheiros estão divididos por sete pastas, cada contendo os PCG cedidos por uma única equipa de investigação. Uma descrição pormenorizada da base de dados e da origem dos sinais pode ser encontrada em [5].

A base de dados não é equilibrada, ou seja, o número de amostras normais não corresponde ao número de amostras patológicas. Isto é um fator indesejado nas redes neuronais convolucionais e portanto, o número de ficheiros foi reduzido para que os números de cada classe coincidissem.

Determinou-se qual a classe predominante em cada pasta e esse mesmo número foi reduzido até ser equivalente ao da classe em menor número. Os PCG removidos foram selecionados aleatoriamente.

Após equilíbrio da base de dados, obtiveram-se 904 sinais dos 3240 inicialmente presentes, correspondendo 50% destes a sinais patológicos e os restantes a sinais saudáveis.

### **5.3. Pré-Processamento**

Após as amostras de som estarem devidamente etiquetadas e em igual número entre classes, procedeu-se à limpeza e tratamento do sinal. Os PCG foram normalizados, filtrados entre os 10 e

os 400Hz utilizando um filtro passa-banda *Butterworth* de quarta ordem e foram removidos períodos de silêncio ou sem sinal do início e final de cada amostra.

Após limpeza, os sinais foram segmentados em intervalos de 1 segundo (intervalo que gera mais amostras mas registrando um batimento cardíaco completo [72]) com sobreposição de 50%, ou seja, cada sinal contém meio segundo do sinal anterior e seguinte. Isto permite não perder eventos aquando da passagem para o domínio das frequências [79]. Um exemplo do sinal filtrado e segmentado pode ser encontrado na Fig.5.2.

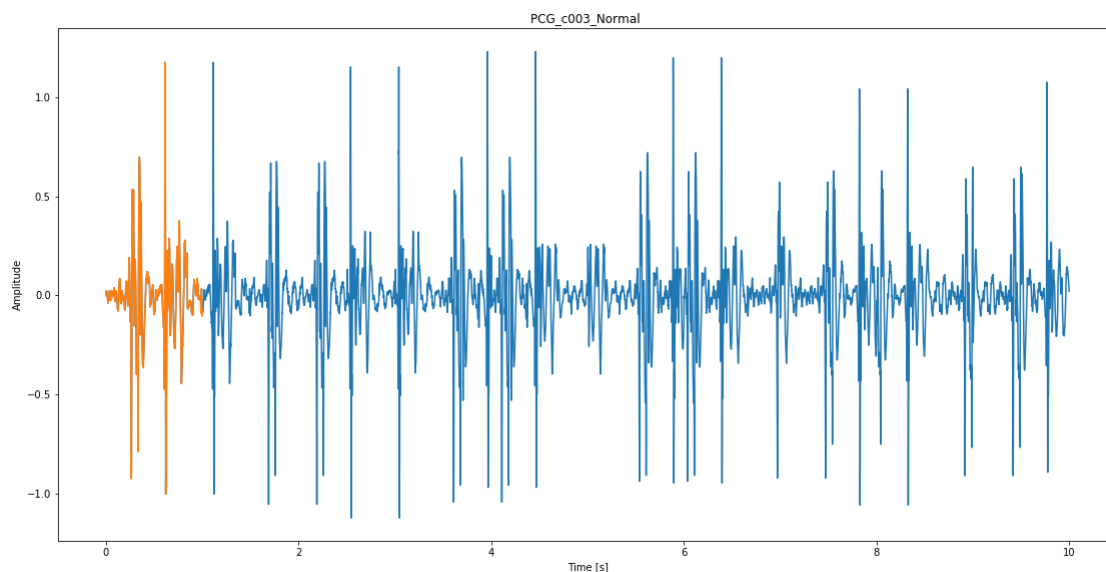


Figura 5.2 – 10 segundos de sinal cardíaco normal (segmento de 1 segundo representado a laranja).

## 5.4. Extração de *Features*

Como discutido no subcapítulo 3.4, as redes convolucionais têm uma arquitetura adaptada para que as suas entradas sejam imagens ou vídeos. Os PCG tiveram, portanto, que ser transformados em dados que pudessem ser alimentados à rede. Isto pode ser feito através duma representação gráfica da amplitude da onda sonora. Métodos mais favorecidos procuram uma representação que contenha informação do domínio das frequências. De entre os diversos métodos presentes na literatura, foram utilizados para esta experiência o espectrograma de Mel, os MFCC (*Mel-frequency cepstral coefficients*) [80], os GTCC (*Gammatone Cepstral Coefficients*) [81] e PNCC (*Power-Normalized Cepstral Coefficients*) [82]. Estas *features* foram representadas sob a forma de *heatmaps* e guardadas em formato .png.

### 5.4.1. Espectrograma de Mel

O espectrograma é uma representação visual do espectro de frequência de um sinal à medida que este varia com o tempo. Torna-se um espectrograma de Mel quando é desenhado à escala de

Mel através do uso de filtros triangulares aplicados ao sinal original. Uma descrição pormenorizada deste método pode ser encontrada em [80].

Para criação dos espectrogramas foi utilizada a biblioteca *Librosa* [68]. Para a STFT (*short-time Fourier transform*) foi escolhida uma janela de *Tukey* ( $\alpha=0.8$ ) por apresentar melhor desempenho em sinais cardíacos como demonstrado em [83]. Cada janela abrange 25ms e tem um passo de 10ms. Isto originou uma série de espectrogramas, guardados sob o formato .png de tamanho 224×224, com uma resolução temporal de 10ms e uma resolução de frequência de 1Hz (Fig.5.3).

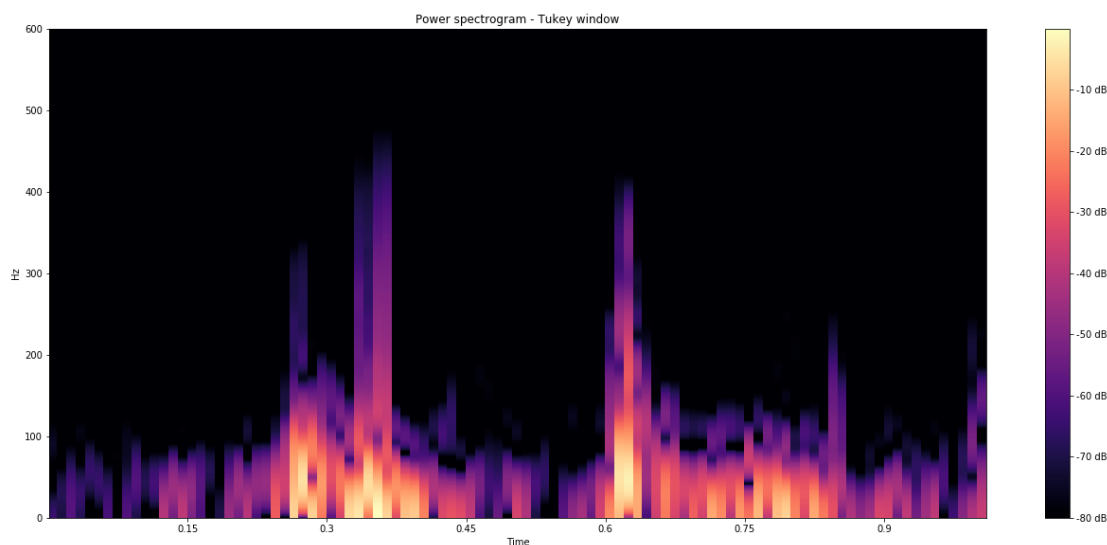


Figura 5.3 – Espectrograma de Mel

#### 5.4.2. MFCC

Os MFCC são coeficientes retirados da transformada discreta de cosseno aplicada ao logaritmo do espectro de Mel. Uma descrição dos MFCC pode ser encontrada em [80]. São largamente utilizados em reconhecimento de voz, processamento da fala, codificação vocal, etc. Foram criados a partir dos espectrogramas apresentados acima pelo que os parâmetros indicados se mantêm. Foram retirados os 13 primeiros coeficientes como representado na Fig.5.4.

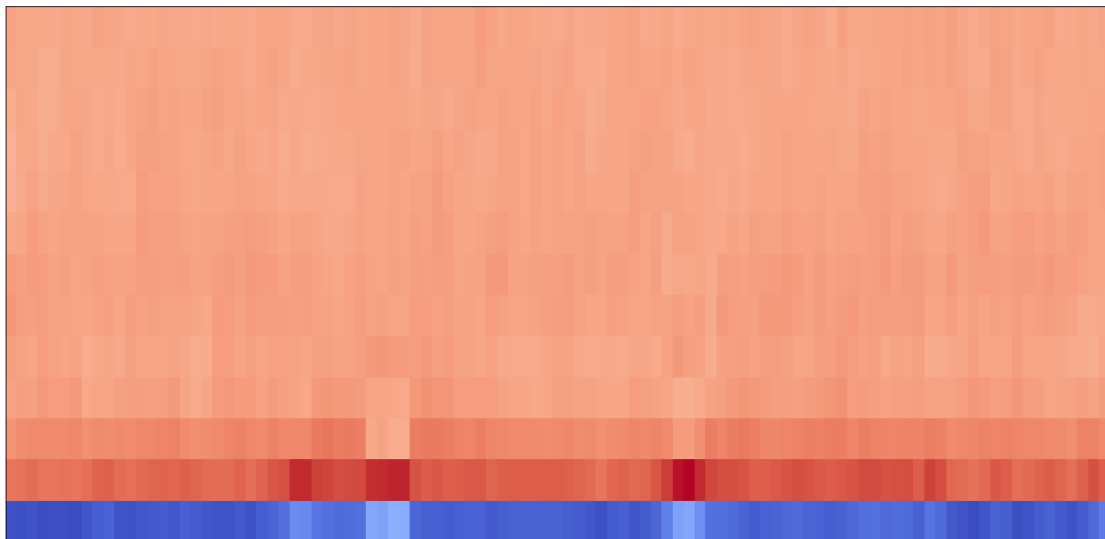


Figura 5.4 – Mel-Frequency Cepstral Coefficients

### 5.4.3. GTCC

Os GTCC (*Gammatone Cepstral Coefficients*) são uma variação dos MFCC. Consistem nos coeficientes retirados de filtros lineares descritos por uma resposta a impulso que é o produto de uma distribuição gama e um senoide. São largamente utilizados em modelos de reconhecimento da fala e em certos casos têm apresentado melhor desempenho que os MFCC [81]. A sua representação utilizada para criação da base de dados encontra-se ilustrada na Fig.5.5.

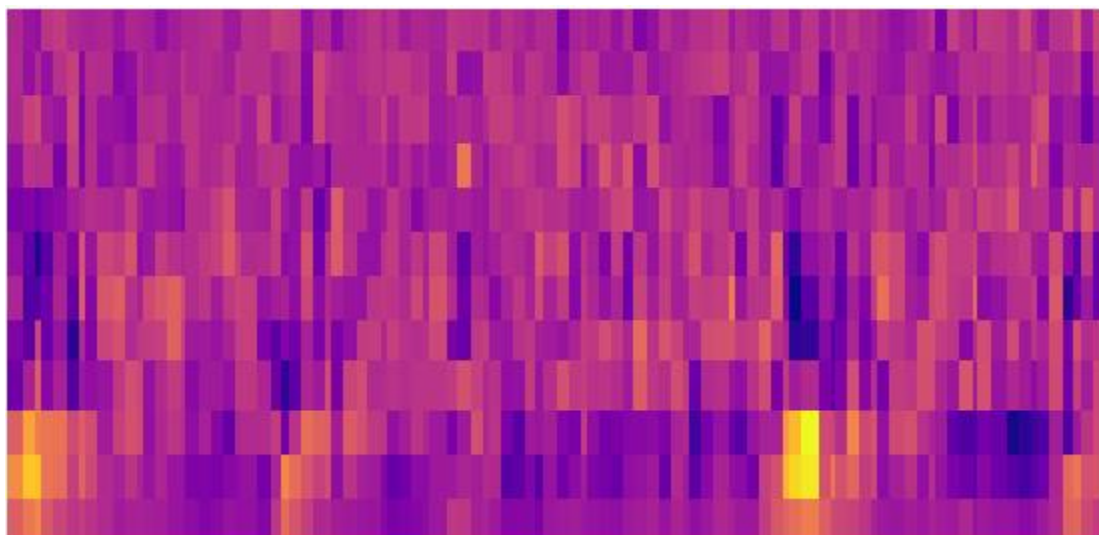


Figura 5.5 – Gammatone Cepstral Coefficients

#### 5.4.4. PNCC

Os PNCC (*Power-Normalized Cepstral Coefficients*) têm como principais características diferenciadoras dos coeficientes supramencionados a não-linearidade baseada numa lei de potência, ao invés de não-linearidade logarítmica, e um algoritmo de redução de ruído baseado em filtragem assimétrica. Uma descrição do processamento dos PNCC pode ser encontrada em [82]. A sua representação utilizada para criação da base de dados encontra-se ilustrada na Fig.5.6.

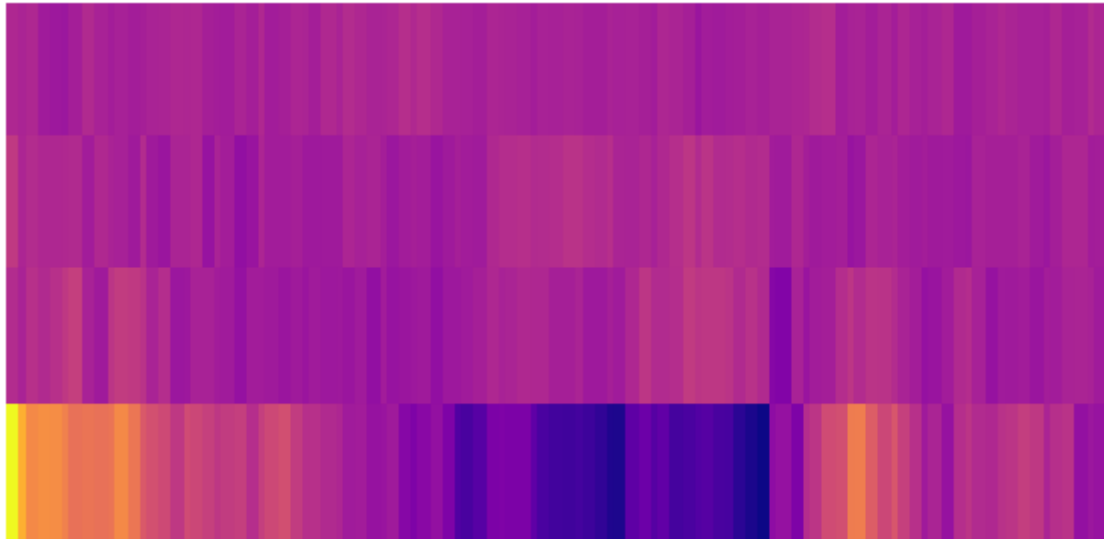


Figura 5.6 – Power Normalized-Cepstral Coefficients.

#### 5.5. Treino das CNN

Transferência de conhecimento no âmbito das redes neuronais envolve utilizar um modelo que tenha sido treinado numa tarefa diferente e numa outra base de dados, e aplicá-lo a um novo problema. A utilização de uma rede neuronal pré-treinada, ou método de *transfer learning*, permite a construção de modelos precisos com um tempo de treino menor requerido. [74] O processo de aprendizagem da rede neuronal em vez de ser iniciado do zero pode usufruir de padrões aprendidos pela rede pré-treinada num problema diferente. Por exemplo, a rede LeNet-5 utilizada por Morales et. al. foi inicialmente treinada para classificar dígitos escritos manualmente em cheques bancários. As redes neuronais convolucionais são o estado da arte nos problemas de classificação de imagem (subcapítulo 3.4).

Como discutido no subcapítulo 3.4, as camadas iniciais de uma rede neuronal convolucional capturam características mais gerais e a especificidade do problema é representada nas intermédias e últimas camadas da base convolucional. Normalmente, é desejável não alterar os pesos das camadas iniciais para que as características mais abrangentes e que sejam comuns a ambas as bases de dados sejam refletidas.

Uma abordagem é manter a estrutura e pesos aprendidos pela rede, alterando apenas as camadas finais que correspondem ao classificador de modo a produzir uma resposta adequada ao

novo problema (estratégia 3, ilustrada na Fig. 5.7). Por exemplo, as redes VGG foram treinadas para classificar as suas entradas em 1 de 1000 objetos diferentes, pelo que para utilizar a sua arquitetura para um problema de classificação binária (como é o caso desta dissertação) as suas camadas finais, ou de classificação, têm de ser alteradas para devolver 1 objeto de 2 possíveis.

Uma segunda fase de aprimoramento, comumente denominada de *fine-tuning*, é gradualmente “descongelar” as últimas camadas da base convolucional (estratégia 2, ilustrada na Fig 5.7), ou seja, permitir que os pesos aprendidos pela rede sejam alterados durante o treino com a nova base de dados. Este processo deve ser feito lentamente para não destabilizar os pesos da rede.

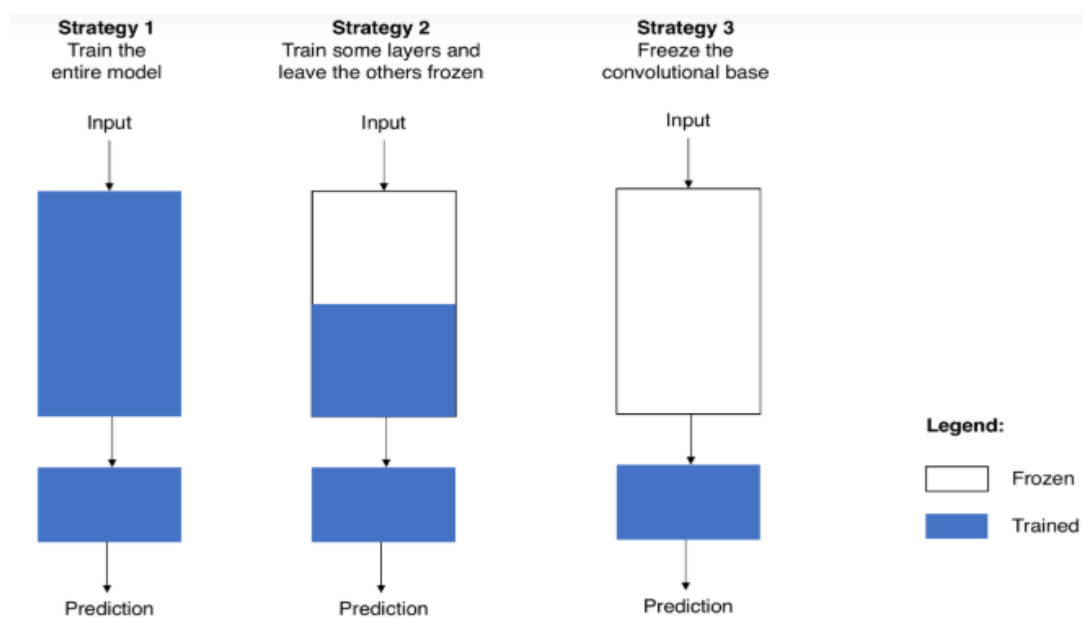


Figura 5.7 – Estratégias de fine-tuning. O primeiro bloco que recebe o input trata-se da base convolucional e o segundo bloco do classificador (adaptado de Marcelino P. [92]).

Existe uma série de hiperparâmetros que o utilizador pode recorrer para adaptar o modelo treinado ao seu problema tendo em atenção fatores referidos como o congelamento de certas camadas e a aceleração/desaceleração do processo de aprendizagem, entre outros. A biblioteca *Tensorflow* [77] em conjunto com a interface de programação de aplicações *Keras* [78] dispõe de funções que permitem alterar esses mesmos hiperparâmetros de entre os quais se destaca:

- **Learning Rate** – é um hiperparâmetro que controla a mudança do modelo em resposta ao erro estimado de cada vez que os pesos são atualizados. Um valor muito pequeno leva a um longo processo de treino que pode congelar, enquanto um valor muito alto pode resultar num treino instável e aproximação de um valor subótimo.
- **Otimizadores** – algoritmos que tratam de minimizar a função custo (subcapítulo 3.3). Uma comparação de diferentes tipos de otimizadores pode ser encontrada em [83].

- **Regularizadores L1 ou L2** – aplicam penalizações a parâmetros de uma camada durante a otimização. Estas penalizações são incorporadas na função custo. São utilizados para prevenir o sobreajuste (modelo adapta-se aos dados nos quais foi treinado mas falha ao generalizar).

As redes treinadas para esta experiência utilizaram as bases convolucionais de outras redes pré-treinadas: as redes VGG16 e VGG19 [85], *DenseNet169* [86] e *Inception* [87]. Estas apresentam diferentes profundidades, arquiteturas e foram treinadas em diferentes bases de dados. Foram escolhidas de modo a verificar se as diferentes arquiteturas originam resultados de classificação significativamente diferentes.

As redes têm a sua base convolucional congelada e é-lhes acoplado o classificador como ilustrado na Fig. 5.8. Os hiperparâmetros do classificador são alterados em diferentes testes, mantendo sempre os pesos importados da base convolucional. O classificador que obtiver os melhores resultados é mantido para uma segunda fase que envolve descongelar as últimas camadas da base convolucional e permitir o treino dos seus pesos, como ilustrado pela estratégia 2 da Fig.5.7.

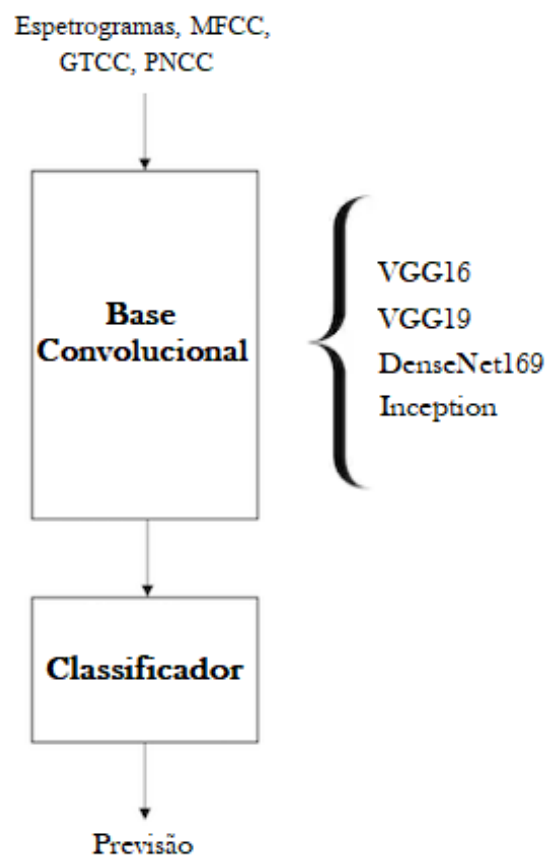


Figura 5.8 – Esquema das redes geradas a partir redes pré-treinadas.

Os hiperparâmetros dos classificadores foram adaptados a cada rede para obter a melhor classificação possível. No entanto, partilham uma arquitetura base relativamente simples. Os classificadores são compostos, sequencialmente, por uma camada *Flatten* que permite tornar a saída da base convolucional num vetor, uma camada convolucional com regularizadores L2 (o valor varia conforme a rede como representado nas tabelas 6.1 e 6.2) e função de activação *swish*, uma camada de *dropout* (0.5) e a camada de saída que toma a decisão final através de uma função de activação sigmoide. Uma representação da arquitetura dos classificadores pode ser observada na Fig. 5.9. As redes são treinadas usando uma função de entropia cruzada como função custo, e foram testados diferentes optimizadores (tabelas 6.1 e 6.2). Os resultados obtidos podem ser verificados no capítulo seguinte.

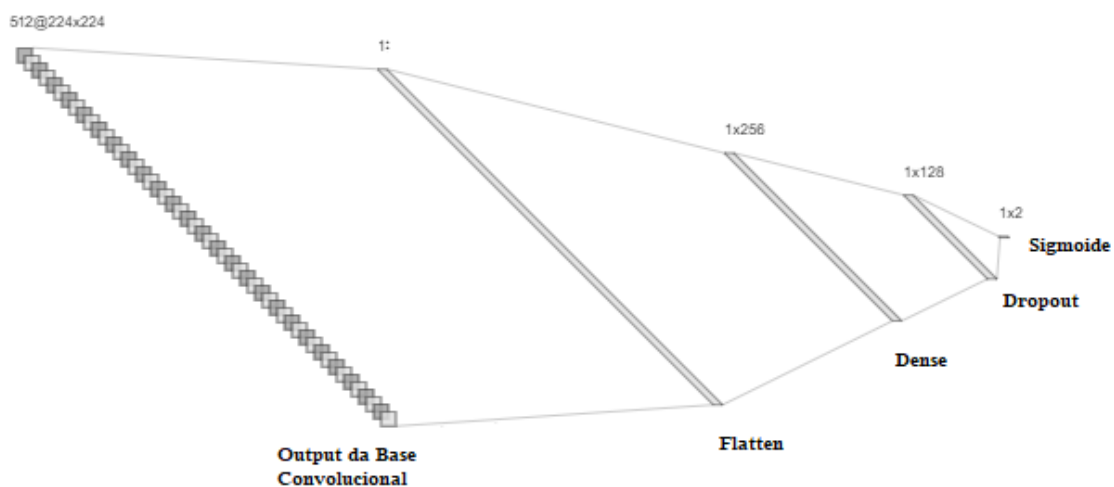


Figura 5.9 – Arquitetura dos classificadores gerados.

## 5.6. Métricas de Avaliação

Avaliar o desempenho de uma rede é um passo crucial para a comparação de resultados entre diferentes modelos. A métrica mais simples e utilizada é o resultado de precisão de classificação. Não é, no entanto, uma métrica suficiente por si só. Não dá indicações dos tipos de erros que estão a ser cometidos pelo classificador. Existem muitas outras métricas, mas nesta dissertação serão apenas utilizadas a precisão de classificação, a perda logarítmica, a sensibilidade e a especificidade.

- **Precisão** – corresponde ao número de previsões corretas de entre todas as previsões realizadas.

$$Precisão = \frac{Número\ de\ Previsões\ Corretas}{Número\ Total\ de\ Previsões} \quad (5.1)$$

- **Perda Logarítmica** – penaliza as classificações incorretas. Quanto mais próximo de 0 for o seu valor, maior será a sua precisão de classificação. O classificador deve

atribuir uma probabilidade de uma amostra pertencer a uma classe. Seja  $N$  o número de amostras e  $M$  o número de classes, onde  $y_{ij}$  indica se a amostra  $i$  pertence à classe  $j$ , e  $p_{ij}$  indica a probabilidade da amostra  $i$  pertencer à classe  $j$ , a perda logarítmica é dada por:

$$LogLoss = \frac{-1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} * \log(p_{ij}) \quad (5.2)$$

- **Sensibilidade** – proporção de amostras de uma classe que são identificadas como pertencentes a essa mesma classe, ou positivos verdadeiros. Isto pressupõe que existirá um número de amostras incorretamente classificadas como falsas, ou falsos negativos. Em contexto clínico traduz-se como a probabilidade de um indivíduo doente ter o seu teste positivo.

$$= \frac{\textit{Positivos Verdadeiros}}{\textit{Positivos Verdadeiros} + \textit{Falsos Negativos}} \quad (5.3)$$

- **Especificidade** – proporção de amostras não pertencentes a uma classe que são corretamente identificadas como não pertencentes, ou negativos verdadeiros. Isto pressupõe que existirá um número de amostras incorretamente classificadas como positivas, ou falsos positivos. Em contexto clínico traduz-se como a probabilidade de um indivíduo saudável ter o seu teste negativo.

$$= \frac{\textit{Negativos Verdadeiros}}{\textit{Negativos Verdadeiros} + \textit{Falsos Positivos}} \quad (5.4)$$

A sensibilidade é uma métrica que no campo da saúde é preferida à especificidade. Caso tenha de se escolher um de dois sistemas com resultados de precisão semelhantes, a métrica que determinará o melhor sistema neste contexto é a sensibilidade. Isto porque é preferível enviar um paciente saudável realizar testes do que não realizar os mesmos testes num paciente doente que é classificado incorretamente como saudável.

## 6. Resultados e Discussão

Os seguintes resultados obtiveram-se do procedimento descrito no capítulo 5. Primeiramente, a base convolucional da rede pré-treinada é congelada. As redes são então acopladas a classificadores cujos parâmetros de cada camada são alterados individualmente em cada treino, apesar de manterem a arquitetura descrita em 5.5. O classificador que obtiver o melhor resultado é utilizado numa segunda fase que envolve descongelar gradualmente as últimas camadas da base convolucional da rede pré-treinada. Os resultados obtidos utilizando os MFCC e os espectrogramas de Mel são apresentados nas tabelas 6.1 e 6.2, respetivamente.

Tabela 6.1 – Resultados das redes treinadas na base de dados de imagens dos MFCC. Em negrito apresentam-se os melhores resultados para cada indicador. A linha com fundo cinza indica o conjunto dos melhores resultados.

Rede	Parâmetros	PerdaLog	Precisão	Sensibilidade	Especificidade
<b>DENSE169</b>	SGD(momentum=0.9),lre-4, L2=0.001	0.548416	0.721576	0.667802	<b>0.857464</b>
	SGD(Nesterov),lre-4	0.511224	0.763963	0.741357	0.786570
	Nadam,lre-5	0.536297	0.762633	0.754322	0.770945
<b>INCEPTION</b>	Nadam,lre-5	0.544584	0.756649	0.730054	0.783245
	SGD(momentum=0.9,Nesterov),lre-4	0.571776	0.760140	0.774352	0.745928
<b>VGG16</b>	Nadam,lre-5,L2=0.001	0.481802	0.761968	0.751995	0.771941
	Nadam,lre-6	0.593590	0.760971	0.765957	0.755985
	Nadam,lre-4	0.487037	0.734209	0.782497	0.685920
	Nadam,lre-5	0.537755	0.772274	0.782081	0.762467
	Nadam,lre-6, 15frozen	0.478886	0.772606	0.765764	0.779711
	Nadam,lre-5, 17frozen	0.788246	0.771277	0.697506	0.845204
<b>VGG19</b>	Nadam,lre-5	0.520114	0.774767	0.783993	0.765542
	Adamax,lre-5	0.551258	0.770445	0.776014	0.764877
	SGD(momentum=0.9,Nesterov),lre-4	0.546714	0.770944	0.764295	0.777593
	Nadam,lre-6, 18frozen	<b>0.392716</b>	<b>0.778424</b>	0.763404	0.793450

	Nadam,lre-6, 19frozen	0.431317	0.773105	<b>0.786048</b>	0.759388
--	--------------------------	----------	----------	-----------------	----------

\*Informação para leitura das tabelas.

Parâmetros: [Otimizador, Learning Rate, Regularizador, Camadas Congeladas da Base Convolutacional]

Abreviaturas: SGD: stochastic gradient descent, lre-x: learning rate =  $1^{-x}$ .

Se não indicado o contrário, o regularizador usado é L2=0.01.

Tabela 6.2 – Resultados das redes treinadas na base de dados de imagens dos Espectrogramas de Mel. Em negrito apresentam-se os melhores resultados para cada indicador. A linha com fundo cinza indica o conjunto dos melhores resultados.

Rede	Parâmetros	PerdaLog	Precisão	Sensibilidade	Especificidade
<b>DENSE169</b>	Nadam,lre-6	0.686490	0.781579	0.761843	0.801315
	Nadam,lre-5	0.511066	0.787993	0.776727	0.799260
	Nadam,lre-6, 5,60frozen	0.437348	0.775493	0.693543	<b>0.857335</b>
<b>INCEPTION</b>	Nadam,lre-5	0.553264	0.781908	0.768257	0.795559
	Nadam,lre-6	0.829615	0.779112	0.762418	0.795806
	Adamax,lre-5	0.908945	0.780099	0.778207	0.781991
<b>VGG16</b>	Nadam,lre-5	0.618122	0.794079	0.779277	0.808882
	Adamax,lre-4	0.641987	0.802254	<b>0.806097</b>	0.798412
	Nadam,lre-6, 17frozen	0.688437	0.795724	0.788226	0.802580
	Adamax,lre-6, 17frozen	<b>0.311134</b>	<b>0.808530</b>	0.797651	0.819367
<b>VGG19</b>	Nadam,lre-5	0.621181	0.787171	0.774670	0.799671
	Nadam,lre-7, 18frozen	1.228346	0.792599	0.787218	0.787976
	Nadam,lre-6, 18frozen	0.570918	0.794408	0.790082	0.798353
	Nadam,lre-7, 19frozen	1.415251	0.787993	0.784734	0.791242

Os resultados dos testes realizados nas bases de dados de imagens dos MFCC e Espectrogramas de Mel revelam que as redes VGG apresentam uma maior percentagem de classificações corretas em relação às redes *Inception* e *DenseNet*, tanto em relação aos MFCC como aos espectrogramas, por diferenças entre os 2% e 3%.

Todas as redes mostraram uma melhor performance na base de dados de espectrogramas de Mel. Isto poderá dever-se ao espectrograma ser uma representação de dimensão superior da onda sonora que preserva parcialmente a sua forma, e como as bases convolucionais são retiradas de redes cujo problema consistia em classificar imagens, as suas camadas podem traduzir essa

procura de cantos, limites e formas na imagem. Em comparação, os MFCC são retirados dos filtros aplicados no espectrograma, já consiste numa representação de menor dimensão da onda.

Em relação aos otimizadores, os Adam atingiram mais rapidamente convergência num valor e maior precisão do que os SGD. É, no entanto, argumentado em [88] que apesar dos resultados positivos, as redes Adam tenham depois dificuldade em generalizar fora do treino.

As redes VGG produziram a sensibilidade mais elevada em ambas as bases de dados de imagens, um fator desejável em sistemas aplicados na área da saúde. A rede *DenseNet*, por sua vez, obteve os resultados mais elevados de especificidade.

O descongelar das últimas camadas da base convolucional VGG obteve melhores resultados do que a base completamente congelada (aumento do valor de precisão até 0.7%), mas apenas quando descongeladas as última e penúltima camadas. Com maior número de camadas finais descongeladas o resultado de precisão tende a diminuir. No caso das restantes redes o descongelar das últimas camadas gerou sempre piores resultados. O processo de treino da rede que obteve os melhores resultados pode ser observado nas Figs.6.1 e 6.2.

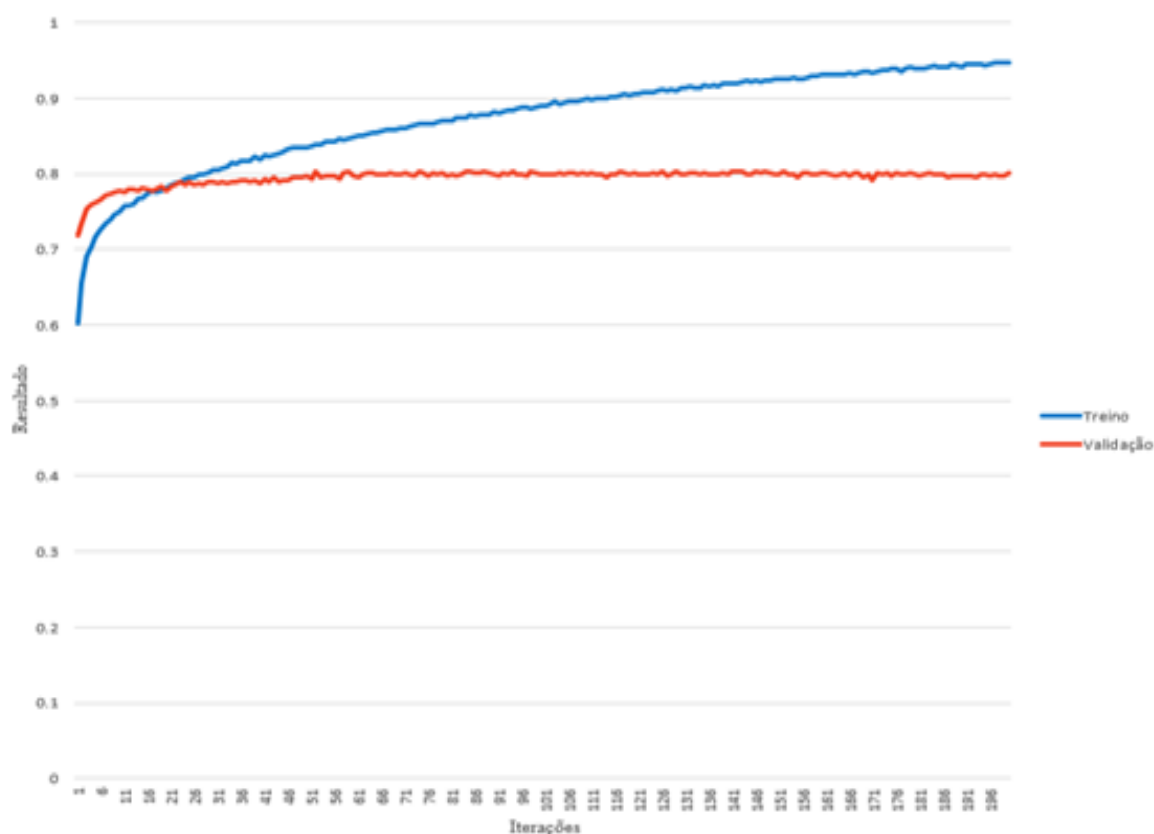


Figura 6.1 – Treino da rede com a base convolucional congelada e otimizador Adamax. A azul os resultados do grupo de treino e a vermelho os do grupo de validação.

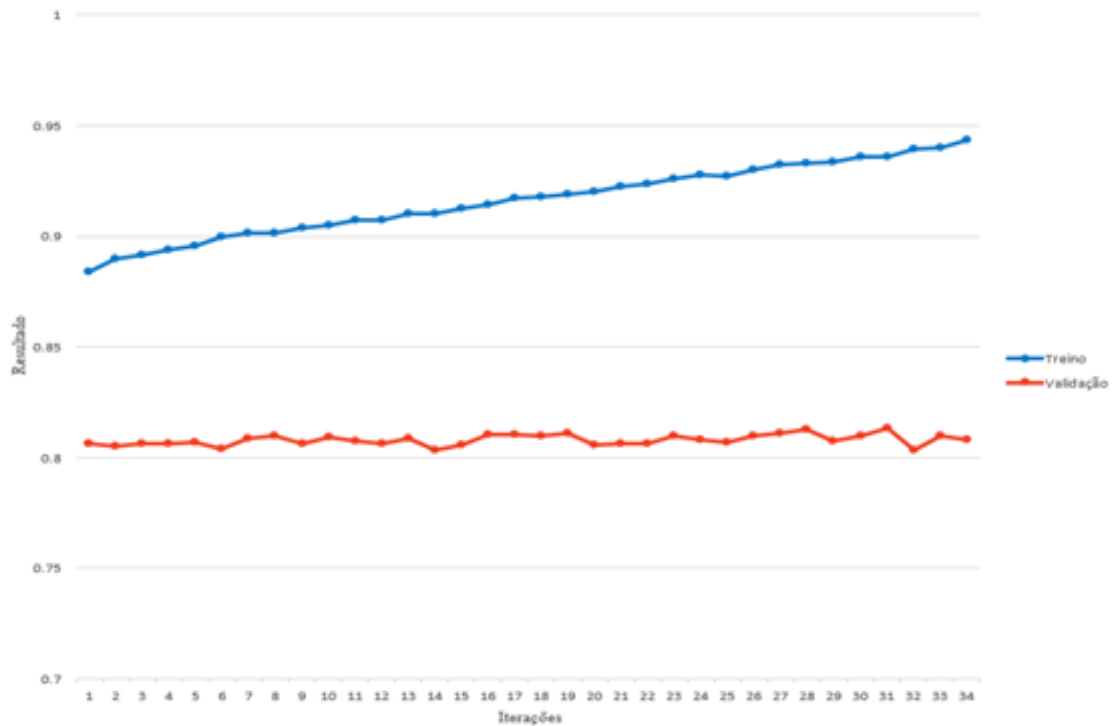


Figura 6.2 – Treino da rede com a última camada da base convolucional descongelada. Pesos do classificador mantidos da rede anterior. A azul os resultados do grupo de treino e a vermelho os do grupo de validação.

A Fig.6.1 representa o treino com a base convolucional congelada. Podemos observar que o modelo já apresenta um valor de precisão de cerca de 60% no grupo de treino e cerca de 72% no conjunto de validação logo na primeira iteração. Isto era esperado porque foram importados os pesos da rede VGG16 o que permitiu ao modelo não os inicializar aleatoriamente. Existe, no entanto, a partir da 60ª iteração sinais de *overfitting*, a precisão no conjunto de treino continua a subir enquanto no grupo de validação estagnou nos 80%, aproximadamente.

A Fig.6.2 representa o descongelamento e treino da última camada da base convolucional. Apesar de haver alguma melhoria no resultado de precisão, é também verificado *overfitting* pois o progresso no conjunto de treino é constante e o mesmo não se verifica no de validação. O facto de o grupo de validação apresentar oscilações no resultado de precisão pode indicar a necessidade de parâmetros de treino mais sensíveis.

Nos testes seguintes, são apenas testadas redes cuja base convolucional é retirada da rede VGG16 por apresentar semelhança e um tamanho menor que a VGG19.

Foram criadas novas bases de dados, uma de matrizes contendo os primeiros quatro coeficientes dos MFCC acoplados aos coeficientes delta e delta-delta respetivos, dos primeiros 12 GTCC (*Gammatone Cepstral Coefficients*) e dos primeiros 4 PNCC (*Power-Normalized Cepstral Coefficients*). Os resultados podem ser encontrados nas tabelas 6.3, 6.4 e 6.5, respetivamente.

Tabela 6.3 – Resultados das redes treinadas na base de dados de imagens dos MFCC acoplados aos coeficientes delta e delta-delta. Em negrito apresentam-se os melhores resultados para cada indicador. A linha com fundo cinza indica o conjunto dos melhores resultados.

Rede	Parâmetros	PerdaLog	Precisão	Sensibilidade	Especificidade
<b>VGG16</b>	Nadam,lre-4	0.709407	<b>0.696715</b>	0.712039	0.681390
	Nadam,lre-5	0.788595	0.692909	0.700522	0.685297
	SGD(momentum=0.9,Nesterov),lre-4	<b>0.579947</b>	0.691506	<b>0.720251</b>	0.662759
	Adamax,lre-5	0.789373	0.694611	0.697366	<b>0.691857</b>

Tabela 6.4 – Resultados das redes treinadas na base de dados de imagens dos GTCC. Em negrito apresentam-se os melhores resultados para cada indicador. A linha com fundo cinza indica o conjunto dos melhores resultados.

Rede	Parâmetros	PerdaLog	Precisão	Sensibilidade	Especificidade
<b>VGG16</b>	Adamax,lre-5	0.733484	<b>0.720923</b>	<b>0.709178</b>	0.732669
	Adamax,lre-4	<b>0.616829</b>	0.720613	0.708868	0.732358
	Nadam,lre-5	0.816267	0.715542	0.696451	<b>0.734633</b>

Tabela 6.5 – Resultados das redes treinadas na base de dados de imagens dos PNCC. Em negrito apresentam-se os melhores resultados para cada indicador. A linha com fundo cinza indica o conjunto dos melhores resultados.

Rede	Parâmetros	PerdaLog	Precisão	Sensibilidade	Especificidade
<b>VGG16</b>	Adamax,lre-5	0.719794	<b>0.709498</b>	0.711862	<b>0.707134</b>
	Adamax,lre-4	<b>0.664985</b>	0.689042	0.675216	0.702868
	Nadam,lre-5	0.772806	0.697471	<b>0.714382</b>	0.680560

Todos os testes apresentaram um resultado de classificação bastante inferior aos obtidos com os espectrogramas e MFCC. O acoplamento dos coeficientes delta e delta-delta e diminuição do número de MFCC piorou o resultado de precisão em relação aos 13 MFCC.

A classificação com as bases de dados dos GTCC e PNCC conta com os valores mais baixos, distando entre 8% e 10%, respetivamente, do valor máximo obtido com os espectrogramas de Mel.

Foram resultados inesperados em função de artigos presentes na literatura [81][82]. Os sistemas em ambos os artigos eram testados em problemas de classificação da fala, o que oferece um intervalo de frequências diferente daquele do PCG e não eram classificados com redes convolucionais. Isto pode ter, por si, influenciado os resultados. Além disso, as imagens utilizadas na criação das bases de dados que geraram as nossas redes são uma codificação RGB dos valores desses coeficientes guardadas em formato .png. Uma utilização mais direta dos valores desses coeficientes evitaria a sua codificação e possível perda de informação. Um formato alternativo no pré-processamento poderá obter melhores resultados.

De modo a verificar a capacidade da rede criada se adaptar a novo tipo de dados, foi criada uma nova base de dados de imagens de espectrogramas de Mel, desta vez recorrendo aos PCG disponibilizados pelo grupo PASCAL [89]. Os PCG são divididos em duas bases de dados, a base A, contendo amostras de som recolhidas através do *iStethoscope Pro iPhone*, e portanto do público em geral, e a base B, cujas amostras são recolhidas em ambiente clínico através do estetoscópio digital *DigiScope*. Uma descrição pormenorizada da base de dados PASCAL pode ser encontrada em [89]. Os resultados obtidos podem ser encontrados na tabela 6.6.

Tabela 6.6 – Resultados de precisão da rede gerada em ambas as bases de dados do grupo PASCAL. A azul indicam-se os resultados corretamente previstos e a vermelho os incorretamente.

		Normal	Patológico
<b>Dataset A</b>	Normal	79.28%	20.72%
	Extrassístole	64.76%	35.24%
	Sopro	80.68%	19.32%
<b>Dataset B</b>	Normal	73.1%	26.9%
	Normal c/ ruído	75.35%	24.65%
	Extrassístole	74.86%	25.14%
	Sopro	75.98%	24.02%
	Sopro c/ ruído	74.12%	25.88%

Uma vez que o nosso classificador tratava de um problema de classificação binária (sinal normal ou patológico) a base de dados PASCAL foi adaptada ao problema. Os grupos de extrassístole e sopro são atribuídos ao grupo patológico sem se procurar distinção entre estas duas classes.

A rede obteve um valor de precisão aceitável face aos resultados do treino quando contemplado o grupo de amostras normais da base A (79.3%). As amostras normais da base B obtiveram valores entre 4% e 6% inferiores em termos de precisão, mas também contavam com um número maior de amostras. É de notar que a rede revela alguma robustez a ruído, uma vez que tanto o grupo normal como o normal ruidoso da base de dados B apresentam valores de classificação semelhantes (distam 2.25%). O mesmo ocorre quando contemplado o grupo de sopros e de sopros com ruído da base B (distam 1.16%).

No entanto, quando considerado o grupo patológico, a rede falhou a sua classificação em maior parte dos casos, tanto nas amostras da base A como da base B, apesar de apresentar diferenças entre a percentagem de classificações corretas de sopros e de extrassístoles, sendo a rede capaz de identificar com maior precisão sopros na base B (por cerca de 5%), e extrassístoles na base A (por cerca de 10%).

O facto de tanto as amostras da base A, adquiridas através do *iStethoscope Pro iPhone* pelo público em geral, como da base B, recolhidas em ambiente clínico através do *DigiScope*, apresentarem resultados relativamente semelhantes, indica que a rede poderá ser sensível a novos métodos de aquisição de sinal. Isto porque a rede apresenta robustez ao ruído e porque a precisão de classificação da base de dados B não foi mais elevada que a da base de dados A, apesar da aquisição dos seus sinais ser em ambiente clínico. Isto indica que novos métodos de aquisição poderão ter de ser treinados na rede antes de esta ser capaz de classificar os sinais destes métodos provenientes.



## 7. Conclusão

A abordagem adotada nesta dissertação considerou algumas variações, explorando diferentes tipos de parâmetros acústicos e diferentes estratégias de treino, tendo ficado o desenvolvimento limitado pelas capacidades de computação dos recursos disponíveis. De qualquer forma, os resultados obtidos demonstram a viabilidade da utilização da metodologia *transfer learning* quando aplicada à classificação de PCGs. Os cenários considerados permitiram obter bons resultados, em linha com outros resultados de abordagens diferentes. O método utilizado não exige segmentação síncrona prévia do sinal e por isso mostra-se particularmente capaz de lidar com dados em bruto.

A metodologia *transfer learning* revelou ser eficiente na redução do tempo de treino da rede e no aumento da percentagem de acerto nas primeiras instâncias do treino. É verificado pelos dados obtidos que as redes geradas são bastante robustas a ruído e que há a possibilidade de serem aplicadas ao problema de classificação dos PCG.

No entanto, após usufruir do conhecimento da base convolucional importada, a rede teve dificuldade em generalizar as características para o problema de classificação do PCG. Ou seja, não conseguiu criar camadas finais com pesos que traduzissem a especificidade do problema.

A incapacidade da rede gerar melhores resultados terá vários motivos. As redes neuronais convolucionais exigem uma quantidade de dados enorme de modo a tornarem-se eficientes na tarefa de classificação, o que também é um motivo para a importação das bases convolucionais. As redes pré-treinadas utilizadas contavam com milhões de exemplos para cada uma das suas classes. Além da quantidade de dados não ser ainda suficiente, também a qualidade dos dados fica aquém do necessário. É necessário rotular as diferentes patologias em cada ficheiro de som, e também identificá-las temporalmente nos casos em que necessário, por exemplo, nas extrassístoles que não ocorrem em todos os batimentos cardíacos de um indivíduo e que portanto, ao rotular um ficheiro de som inteiro como extrassístole, irá resultar em batimentos normais serem rotulados como patológicos. Morales et. al. demonstrou que a transformação do PCG causa melhoria na classificação face à onda em formato .wav. O facto das *features* recolhidas nesta dissertação serem *heat maps* guardadas em formato .png poderá ser um ponto a melhorar.

Futuras melhorias à tecnologia passarão por criar um algoritmo que permita à rede não classificar apenas batimentos mas considerar todos os batimentos classificados ao longo de um PCG (útil na classificação de anomalias que não se verifiquem em todos os batimentos). Explorar o impacto dos hiperparâmetros do classificador em maior detalhe e aprimorar os valores dos mesmos. A união com outros sistemas de classificação e criação de um algoritmo de votação poderá originar melhores resultados do que os sistemas isolados.



## Bibliografia

1. Pagidipati Neha Jadeja, e Gaziano Thomas A. «**Estimating Deaths From Cardiovascular Disease: A Review of Global Methodologies of Mortality Measurement**». *Circulation*, vol. 127, n. 6, Fevereiro de 2013, pp. 749–56. ahajournals.org (Atypon), doi:10.1161/CIRCULATIONAHA.112.128413.
2. McGill HC, McMahan CA, Gidding SS (março de 2008). «**Preventing heart disease in the 21st century: implications of the Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study**». *Circulation*. 117 (9): 1216–27.
- 3 S. Mangione and L. Nieman, “**Cardiac auscultatory skills of internal medicine and family practice trainees: A comparison of diagnostic proficiency**,” *JAMA*, vol. 278, no. 9, pp. 717–722, 1997.
4. J. Constant, “**Essentials of Bedside Cardiology**”, second edition. Humana Press Inc, 2003.
5. Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, Castells F, Roig JM, Silva I, Johnson AE, Syed Z, Schmidt SE, Papadaniil CD, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab MR, Sameni R, Mark RG, Clifford GD. “**An open access database for the evaluation of heart sound algorithms. Physiological Measurement**” 2016; 37(9).
6. Uguz H 2012, “**A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases**”, *J. Med. Syst.* 36 61–72.
7. Ari S, Hembram K, Saha G, 2010, “**Detection of cardiac abnormality from pcg signal using lms based least square svm classifier Expert Syst**”. *Appl.* 37 8019–26.
8. Saracoglu R 2012 “**Hidden markov model-based classification of heart valve disease with pca for dimension reduction**” *Eng. Appl. Artif. Intell.* 25 1523–8.
9. Quiceno-Manrique A F, Godino-Llorente J I, Blanco-Velasco M and Castellanos-Dominguez G, 2010, “**Selection of dynamic features based on time-frequency representations for heart murmur detection from phonocardiographic signals**”, *Ann. Biomed. Eng.* 38 118–37.
10. A. Leatham, “**Auscultation and phonocardiography: a personal view of the past 40 years.**” *Br. Heart J.* 57, 397–403 (1987).
11. Joyanta Kumar Roy, Tanmay Sinha Roy and Subhas Chandra Mukhopadhyay, “**Heart Sound: Detection and Analytical Approach Towards Diseases**”, 2018.
12. Le T, Bhushan V. “**First Aid for the USMLE**” Step 1 2015. McGraw-Hill Education; 2014.
13. Gersh BJ. “**Auscultation of heart sounds.**” In: Post TW, ed. *UpToDate*. Waltham, MA: UpToDate. <https://www.uptodate.com/contents/auscultation-of-heart-sounds>. Last updated December 18, 2015. Accessed February 24, 2017.
14. Mangla A. “**Heart Sounds. In: Lange RA. Heart Sounds.**” New York, NY: WebMD. <https://emedicine.medscape.com/article/1894036>. Updated April 24, 2014. Accessed December 18, 2017.

15. S. Ari, P. Kumar, and G. Saha, “**A robust heart sound segmentation algorithm for commonly occurring heart valve diseases,**” *J. Med. Eng. Technol.*, vol. 32, no. 6, pp. 456–465, 2008.
16. Meyer ThE. “**Auscultation of cardiac murmurs in adults.**” In: Post TW, ed. *UpToDate*. Waltham, MA: UpToDate. <https://www.uptodate.com/contents/auscultation-of-cardiac-murmurs-in-adults>. Last updated December 22, 2016. Accessed April 9, 2017.
17. Sukryool Kang, Robin Doroshov, James McConnaughey, Raj Shekhar. “**Automated identification of innocent Still's murmur in children.**” *IEEE Trans Biomed Eng.* 2017; 64(6): pp. 1326–1334. doi: 10.1109/tbme.2016.2603787.
18. Herold G. “**Internal Medicine**”. Cologne, Germany: Herold G; 2014.
19. Amboss.com. (2014). “**Cardiovascular examination.**” [online] Available at: <https://www.amboss.com/us/library#xid=rM0fJg&anker=Za94cbf56c65ba879048421c95766464c> [Accessed 29 Aug. 2019].
20. F. Plesinger, I. Viscor, J. Halamek, J. Jurco, and P. Jurak, “**Heart sounds analysis using probability assessment,**” *Physiol. Meas.*, vol. 38, pp. 1685–1700, 2017.
21. S. Yuenyong, A. Nishihara, W. Kongprawechnon, and K. Tungpimolrut, “**A framework for automatic heart sound analysis without segmentation,**” *Biomed. Eng. Online*, vol. 10, no. 13, pp. 1–23, 2011.
22. A. Gharehbaghi, T. Dutoir, A. Sepehri, P. H. I, and P. Ask, “**An Automatic Tool for Pediatric Heart Sounds Segmentation,**” *Comput. Cardiol.*, pp. 37–40, 2011.
23. Dwivedi A, Imtiaz S, Rodriguez-Villegas E. “**Algorithms for Automatic Analysis and Classification of Heart Sounds – A Systematic Review.**” London, United Kingdom; 2018.
24. Rebala G., Ravi A., Churiwala S. “**(Artificial) Neural Networks. In: An Introduction to Machine Learning.**” Springer, Cham, 2019.
25. Kamath U., Liu J., Whitaker J. “**Basics of Deep Learning. In: Deep Learning for NLP and Speech Recognition.**” Springer, Cham, 2019.
26. Matthew D. Zeiler and Rob Fergus. “**Visualizing and understanding convolutional networks.**” In *European Conference on Computer Vision*, pages 818–833, Springer, 2014. DOI: 10.1007/978-3-319-10590-1\_53. 44, 94, 95, 97, 170.
27. Khan S., Rahmani H., Shah S. Bennamoun M. “**A Guide to Convolutional Neural Networks for Computer Vision.**” *Synthesis Lectures on Computer Vision*, 2018. DOI: 10.2200/S00822ED1V01Y201712COV015.
28. Samer Hijazi, Rishi Kumar, and Chris Rowen, “**Using Convolutional Neural Networks for Image Recognition.**” IP Group, Cadence.
29. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. “**Deep Learning**”. MIT Press, 2016. <http://www.deeplearningbook.org> 54.
30. G. D. Clifford et al., “**Recent advances in heart sound analysis,**” *Physiol. Meas.*, vol. 38, no. 8, pp. 10–25, 2017.
31. Gari D Clifford, Chengyu Liu, Benjamin Moody, David Springer, Ikaro Silva, Qiao Li, Roger G. Mark, “**Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016**” - IEEE Conference Publication

- [Internet]. Ieeexplore.ieee.org. 2019 [cited 4 September 2019]. Available from: <https://ieeexplore.ieee.org/abstract/document/7868816>.
32. D.Kumar, P.Carvalho, M.Antunes, R.P.Paiva, J.Henriques, “**Noise detection during heart sound recording using periodicity signatures**”, *Physiol.Meas.*32 (May(5)) (2011) 599–618.
  33. J.P.Ramos, P.Carvalho, R.P.Paiva, J.Henriques, “**Modulation filtering for noise detection in heart sound signals**”, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August, Boston, IEEE, 2011, pp.6013–6016.
  34. Wenfei Cao , Kaidong Wang, Guodong Han , Jing Yao, Andrzej Cichocki, “**A Robust PCA Approach With Noise Structure Learning and Spatial–Spectral Low-Rank Modeling for Hyperspectral Image Restoration**”, 1939-1404 © 2018 IEEE.
  35. Mohammed Nabih-Ali, EL-Sayed A. El-Dahshan, Ashraf S. Yahia, “**Heart Diseases diagnosis using intelligent algorithm based on PCG signal analysis**”. (2017) Egyptian E-Learning University (EELU), Physics Department, Faculty of Sciences, Ain Shams University, Abbasia, Cairo, Egypt.
  36. Shahid Ismail, Imran Siddiqi, Usman Akram, “**Localization and classification of heart beats in phonocardiography signals —a comprehensive review.**” *EURASIP Journal on Advances in Signal Processing*, volume 2018, Article number: 26 (2018).
  37. Tien-En Chen, Shih-I Yang, Li-Ting Ho, Kun-Hsi Tsai, Yu-Hsuan Chen, Yun-Fan Chang, Ying-Hui Lai, Syu-Siang Wang, Yu Tsao, Chau-Chung Wu, “**S1 and S2 Heart Sound Recognition using Deep Neural Networks**”. *IEEE Journals & Magazine* [Internet]. Ieeexplore.ieee.org. 2019 [cited 7 September 2019]. Available from: <https://ieeexplore.ieee.org/abstract/document/7460939>.
  38. P. Sharma, S. A. Imtiaz, and E. Rodriguez-Villegas, “**An Algorithm for Heart Rate Extraction from Acoustic Recordings at the Neck,**” *IEEE Trans. Biomed. Eng.*, vol. PP, no. c, p. 1, 2018.
  39. S. Sun, H. Wang, Z. Jiang, Y. Fang, and T. Tao, “**Segmentation-based heart sound feature extraction combined with classifier models for a VSD diagnosis system,**” *Expert Syst. Appl.*, vol. 41, pp. 1769–1780, 2014.
  40. S. Sun, Z. Jiang, H. Wang, and Y. Fang, “**Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified Hilbert transform,**” *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 219–230, 2014.
  41. S. Sun, “**An innovative intelligent system based on automatic diagnostic feature extraction for diagnosing heart diseases,**” *Knowledge-Based Syst.*, vol. 75, pp. 224–238, 2015.
  42. A. Gharehbaghi, T. Dutoir, A. Sepehri, P. H. I, and P. Ask, “**An Automatic Tool for Pediatric Heart Sounds Segmentation,**” *Comput. Cardiol.*, pp. 37–40, 2011.
  43. S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, “**Segmentation of heart sound recordings by a duration-dependent hidden Markov model,**” *Physiol. Meas.*, vol. 31, no. 4, pp. 513–529, 2010.
  44. H. Tang, T. Li, T. Qiu, and Y. Park, “**Segmentation of heart sounds based on dynamic clustering,**” *Biomed. Signal Process. Control*, vol. 7, no. 5, pp. 509–516, 2012.

45. Y. Zheng, X. Guo, and X. Ding, “**A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification,**” *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2710–2721, 2015.
46. V. N. Varghees and K. I. Ramachandran, “**A novel heart sound activity detection framework for automated heart sound analysis,**” *Biomed. Signal Process. Control*, vol. 13, no. 1, pp. 174–188, 2014.
47. H. Nazeran, “**Wavelet-based segmentation and feature extraction of heart sounds for intelligent PDA-based phonocardiography,**” *Methods Inf. Med.*, vol. 46, no. August, pp. 1–7, 2007.
48. D. Boutana, M. Benidir, and B. Barkat, “**Segmentation and identification of some pathological phonocardiogram signals using time-frequency analysis,**” *IET Signal Process.*, vol. 5, no. 6, p. 527, 2011.
49. H. Tang, T. Li, Y. Park, and T. Qiu, “**Separation of heart sound signal from noise in joint cycle frequency-time-frequency domains based on fuzzy detection,**” *IEEE Trans. Biomed. Eng.*, vol. 57, no. 10, pp. 2438–2447, 2010.
50. G. Redlarski, D. Gradolewski, and A. Palkowski, “**A system for heart sounds classification,**” *PLoS One*, vol. 9, no. 11, 2014.
51. Springer DB, Tarassenko L, Clifford GD. “**Logistic regression-HSMM-based heart sound segmentation.**”, *IEEE Trans Biomed Eng.* 2016;63:822–32.
52. A. Moukadem, A. Dieterlen, N. Hueber, and C. Brandt, “**A robust heart sounds segmentation module based on S-transform,**” *Biomed. Signal Process. Control*, vol. 8, no. 3, pp. 273–281, 2013.
53. Y. Soeta and Y. Bito, “**Detection of features of prosthetic cardiac valve sound by spectrogram analysis,**” *Appl. Acoust.*, vol. 89, pp. 28–33, 2015.
54. G. Amit, N. Gavriely, and N. Intrator, “**Cluster analysis and classification of heart sounds,**” *Biomed. Signal Process. Control*, vol. 4, no. 1, pp. 26–36, 2009.
55. S. Debbal and A. Tani, “**Heart sounds analysis and murmurs,**” *Int. J. Med. Eng. Inform.*, vol. 8, no. 1, pp. 49–62, 2016.
56. P. Lubaib and A. M. KV, “**The Heart Defect Analysis Based on PCG Signals Using Pattern Recognition Techniques,**” *Procedia Technol.*, vol. 24, pp. 1024–1031, 2016.
57. W. Zhang, J. Han, and S. Deng, “**Heart sound classification based on scaled spectrogram and tensor decomposition,**” *Expert Syst. Appl.*, vol. 84, pp. 220–231, 2017.
58. H. M. Fahad, M. U. Ghani Khan, T. Saba, A. Rehman, and S. Iqbal, “**Microscopic abnormality classification of cardiac murmurs using ANFIS and HMM,**” *Microsc. Res. Tech.*, no. July 2017, pp. 1–10, 2018.
59. R. C. King, E. Villeneuve, R. J. White, R. S. Sherratt, W. Holderbaum, and W. S. Harwin, “**Application of data fusion techniques and technologies for wearable health monitoring,**” *Med. Eng. Phys.*, vol. 42, pp. 1–12, 2017.
60. F. Safara, S. Doraisamy, A. Azman, A. Jantan, and S. Ranga, “**Wavelet packet entropy for heart murmurs classification,**” *Adv. Bioinformatics*, vol. 327269, p. 6, 2012.

61. M. Tschannen, T. Kramer, G. Marti, M. Heinzmann, and T. Wiatowski, “**Heart Sound Classification Using Deep Structured Features**,” *Comput. Cardiol.*, vol. 43, pp. 565–568, 2016.
62. T. Gokhale, “**Machine learning based identification of pathological heart sounds**,” *Comput. Cardiol.*, vol. 43, pp. 553–556, 2016.
63. L. Atallah, B. Lo, R. King, and G. Z. Yang, “**Sensor positioning for activity recognition using wearable accelerometers**,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 4, pp. 320–329, 2011.
64. Cristhian Potes, Saman Parvaneh, Asif Rahman, Bryan Conroy, “**Ensemble of Feature-based and Deep learning-based Classifiers for Detection of Abnormal Heart Sounds**”, Philips Research North America, Acute Care Solutions, Cambridge, MA, USA, (2017).
65. M. Zabihi, A. B. Rad, S. Kiranyaz, M. Gabbouj and A. K. Katsaggelos, "Heart sound anomaly and quality detection using ensemble of neural networks without segmentation," *2016 Computing in Cardiology Conference (CinC)*, Vancouver, BC, 2016, pp. 613-616.
66. Li Wan, Matthew Zeiler, Sixin Zhang, Yann LeCun, Rob Fergus, “**Regularization of Neural Networks using DropConnect**”, Dept. of Computer Science, Courant Institute of Mathematical Science, New York University (2013).
67. Edmund Kay, Anurag Agarwal, “**DropConnected neural networks trained on time-frequency and inter-beat features for classifying heart sounds**”, Engineering Department, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom (2017).
68. Elmar Messner, Matthias Zohrer, Franz Pernkopf, “**Heart Sound Segmentation - An Event Detection Approach using Deep Recurrent Neural Networks**”, Signal Processing and Speech Communication Laboratory, Graz University of Technology, Graz, Austria (2018).
69. XIEFENG CHENG, JIANZHONG HUANG, YUNYI LI, GUAN GUI, “**Design and Application of a Laconic Heart Sound Neural Network**”, current version September 13, 2019.
70. W. Zhang, J. Han, S. Deng, “**Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation**”, *Biomedical Signal Processing and Control*, 53 (2019) 101560.
71. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, “**Dropout: a simple way to prevent neural networks from overfitting**”, *J. Mach. Learn. Res.* 15 (1) (2014) 1929–1958.
72. Juan P. Dominguez-Morales, Angel F. Jimenez-Fernandez, Manuel J. Dominguez-Morales, Gabriel Jimenez-Moreno, “**Deep Neural Networks for the Recognition and Classification of Heart Murmurs Using Neuromorphic Auditory Sensors**”, *IEEE Transactions on Biomedical Circuits and Systems*, 2019.
73. Suyi Li, Feng Li, Shijie Tang, Wenji Xiong, "A Review of Computer-Aided Heart Sound Detection Techniques", *BioMed Research International*, vol. 2020, Article ID 5846191, 10 pages, 2020. <https://doi.org/10.1155/2020/5846191>
74. Rawat, W., & Wang, Z, “**Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review**.” (2017), *Neural Computation*, 29(9), 2352–2449.
75. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. “**PhysioBank, PhysioToolkit, and PhysioNet: Components**

of a New Research Resource for Complex Physiologic Signals”. *Circulation* 101(23):e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/content/101/23/e215.full>]; 2000 (June 13).

76. Brian McFee, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenbergk, Oriol Nieto, “**librosa: Audio and Music Signal Analysis in Python.**” Proc. Of the 14<sup>th</sup> Python in Science Conference. (SCIPY 2015)

77. Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffre Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskeyer, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu Xiaoqiang Zheng, “**Tensorflow: Large-scale machine learning on heterogeneous systems**”, 2015. Software available from [tensorflow.org](http://tensorflow.org)

78. François Chollet et. al., “**Keras.**”, GitHub repo: {<https://github.com/fchollet/keras>}

79. G. Heinzel, A. Rüdiger, R. Schilling “**Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flat-top windows.**”, Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Teilinstitut Hannover, February 15, 2002.

80. Kishore Prahallad, “**Speech Technology: A Practical Introduction**”, Carnegie Mellon University & International Institute of Information Technology Hyderabad.

81. Liu, Jia-Ming & You, Mingyu & Li, Guo-Zheng & Wang, Zheng & Xu, Xianghuai & Qiu, Zhongmin & Xie, Wenjia & An, Chao & Chen, Sili. (2013). “**Cough signal recognition with Gammatone Cepstral Coefficients.**” 2013 IEEE China Summit and International Conference on Signal and Information Processing, ChinaSIP 2013 - Proceedings. 160-164. 10.1109/ChinaSIP.2013.6625319.

82. C. Kim and R. M. Stern, “**Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition,**” in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 7, pp. 1315-1329, July 2016, doi: 10.1109/TASLP.2016.2545928.

83. G. Rajkumar, R. Jayabharathy, K. Narasimhan, N. Raju, M. Easwaran, V. Elamaram, Gustavo Ramirez-gonzalez, Marlon Burbano-fernandez, “**Spectral and SNR improvement analysis of normal and abnormal heart sound signals using different windows**”, *Future Generation Computer Systems* (2018), <https://doi.org/10.1016/j.future.2018.09.047>

84. **Optimizing Gradient Descent-** <http://sebastianruder.com/optimizing-gradient-descent/>

85. Karen Simonyan, Andrew Zisserman et al., “**VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION**”, Department of Engineering Science, University of Oxford (2015).

86. Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, “**DENSELY CONNECTED CONVOLUTIONAL NETWORKS**”, 2018.

87. Christian Szegedy, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, “**Going Deeper with Convolutions**”, 2014.

88. N Shirish Keskar, N. and Socher, R., **“Improving Generalization Performance by Switching from Adam to SGD”**, 2017.
89. Bentley, P. and Nordehn, G. and Coimbra, M. and Mannor, S., **"The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results"**, 2011.
90. Heger D., **“An Introduction to Artificial Neural Networks ( ANN )-Methods , Abstraction , and Usage.”** (2015).
91. Braga D., Madureira A., Coelho L., Ajith R., **"Automatic detection of Parkinson’s disease based on acoustic analysis of speech"**, 2019.
92. Marcelino P., **“Transfer learning from pre-trained models”**, 2018.