

OPEN

# Transfer Learning in ECG Classification from Human to Horse Using a Novel Parallel Neural Network Architecture

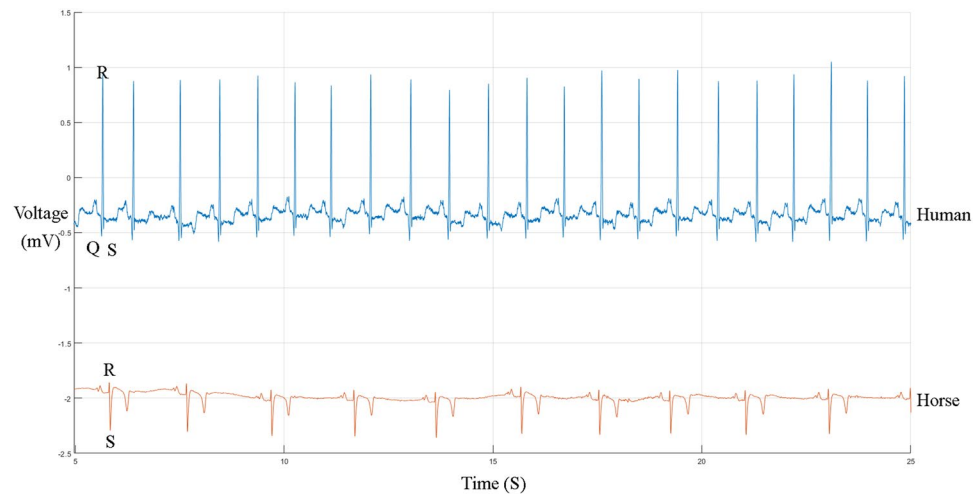
Glenn Van Steenkiste<sup>1\*</sup>, Gunther van Loon<sup>1</sup> & Guillaume Crevecoeur<sup>2,3</sup>

Automatic or semi-automatic analysis of the equine electrocardiogram (eECG) is currently not possible because human or small animal ECG analysis software is unreliable due to a different ECG morphology in horses resulting from a different cardiac innervation. Both filtering, beat detection to classification for eECGs are currently poorly or not described in the literature. There are also no public databases available for eECGs as is the case for human ECGs. In this paper we propose the use of wavelet transforms for both filtering and QRS detection in eECGs. In addition, we propose a novel robust deep neural network using a parallel convolutional neural network architecture for ECG beat classification. The network was trained and tested using both the MIT-BIH arrhythmia and an own made eECG dataset with 26.440 beats on 4 classes: normal, premature ventricular contraction, premature atrial contraction and noise. The network was optimized using a genetic algorithm and an accuracy of 97.7% and 92.6% was achieved for the MIT-BIH and eECG database respectively. Afterwards, transfer learning from the MIT-BIH dataset to the eECG database was applied after which the average accuracy, recall, positive predictive value and F1 score of the network increased with an accuracy of 97.1%.

A cardiac arrhythmia is an abnormal impulse generation or an abnormal conduction of the impulse from the sinoatrial node in the heart<sup>1</sup>. An electrocardiogram (ECG) allows to detect these abnormalities non-invasively. Horses often present arrhythmias at rest and/or during exercise and arrhythmias can range from being clinically irrelevant to potentially life threatening<sup>2-5</sup>. At the same time, sudden death during exercise occurs at an up to 10 times higher ratio in horses compared to human athletes<sup>6,7</sup>. Looking at the causes of sudden death, a cardiovascular problem is found in 8.8% of the cases, but in 68% of the cases no pathology is found on autopsy and a fatal arrhythmia is proposed as the most likely cause of sudden death<sup>6,8</sup>. Arrhythmias in horses may not be associated with obvious clinical signs which alert the owner, and simple auscultation is not always reliable, which requires the recording of an electrocardiogram to make a diagnosis<sup>9</sup>. Also in human medicine the use of an ECG has been proven to be the way to go for pre-participation screening<sup>10</sup>. However ECG interpretation is time-consuming and requires expertise. Furthermore, intra- and interobserver agreement for recognition and classification of arrhythmias varies from good during rest to poor during exercise which currently limits the usage of ECGs in horses<sup>11</sup>. Computer aided signal analysis can eliminate both intra- as interobserver variability and can be done quicker and more cost effective when compared to human interpretation.

Automated ECG interpretation is challenging since an ECG signal can vary between and within patients under different physical circumstances<sup>12</sup>. In order to perform diagnosis based on an ECG, the algorithm must be able to characterize and recognize ECG morphology and rhythm. For human medicine numerous algorithms have been proposed both for filtering, QRS detection and classification of ECGs, but for horses only one filtering algorithm, two QRS detection algorithms and no classification algorithms were described by the knowledge of the authors<sup>3,13,14</sup>. Hoofed animals, such as the horse, have a different nervous conduction system of the heart which results in a different ECG morphology as can be seen in Fig. 1<sup>15</sup>. One of the most striking differences between human and equine ECGs (eECGs) is that instead of having a large R peak, a large S peak is prominent in eECGs.

<sup>1</sup>Department of large animal internal medicine, Ghent University, Ghent, 9000, Belgium. <sup>2</sup>Department of Electromechanical, System and Metal Engineering, Ghent University, Ghent, 9000, Belgium. <sup>3</sup>Core Lab EEDT-DC, Flanders Make, Belgium. \*email: [Glenn.VanSteenkiste@ugent.be](mailto:Glenn.VanSteenkiste@ugent.be)

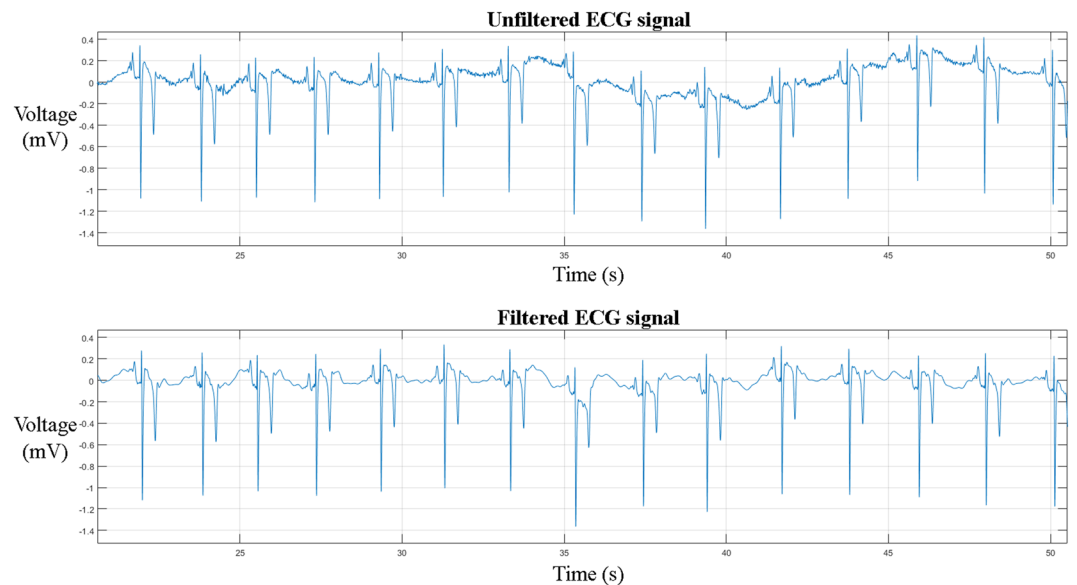


**Figure 1.** Humane (upper) and equine (lower) ECG at rest. Both ECGs are shown at the same paper speed and resolution. Note the difference in both morphology as basal heartrate.

Because there is currently a lack of horse-adapted software, equine ECGs are currently only manually analysed by a trained equine clinician or cardiologist<sup>4,11</sup>.

The aim of this study is to develop algorithms for an end-to-end system for beat-to-beat eECG analysis for horses based upon conventional ECG signal processing techniques along with state of the art deep learning techniques for feature extraction and classification. Since deep learning approaches require massive amounts of data to be trained, which is not available for eECGs, transfer learning will be applied in order to improve results for eECG classification. In most automated ECG interpretation studies, the authors concentrate on conventional machine learning approaches: pre-processing, feature extraction, feature reduction and feature classification<sup>16</sup>. The advantage of using deep learning above the conventional techniques is that the essential steps, namely feature extraction, feature selection and classification can be developed without explicit definition. Improvements in deep learning network architectures and more powerful computing hardware has recently increased the usage of deep learning networks for cardiac arrhythmia diagnosis in ECGs. Two of the most recent papers use short segments of ECG, 1 second and 2 seconds respectively, and use a residual convolutional neural network (CNN)<sup>17,18</sup>. Murugesan *et al.* also introduce the combined use of a long short memory (LSTM) block with a CNN showing improved results. The improved results can be explained because the relationship with surrounding heartbeats can yield important information which is used by the LSTM block. Recurrent neural networks have traditionally been used for time series processing, but they are resource intensive and recent results have shown that convolutional neural network approaches can outperform recurrent neural networks for tasks such as audio synthesis and machine translation<sup>19</sup>. Therefore we introduce a parallel network in this paper, with a separate pathway for feature extraction from both the individual ECG morphology and the temporal relationship. Despite the papers above achieving a high accuracy, sensitivity and positive predictive value (PPV), they lack the ability of providing the cardiologist an exact number of abnormal beats which is an important prerequisite in eECG analysis. Horses can have a heartrate up to 240 bpm (4 beats per second) and thus a segment of 1 or 2 seconds can include both normal and abnormal heartbeats simultaneously<sup>20</sup>. Another study by Isin and Ozdalili<sup>21</sup> converts R-T segments to  $256 \times 256 \times 3$  images and uses the convolutional layers of a pre-trained alexNet<sup>22</sup> as neural network architecture to extract the features, that on their turn are fed into a hidden layer after a principal component analysis is applied to reduce the number of features. Zihlman *et al.*<sup>23</sup> and Rubin *et al.*<sup>24</sup> both used a similar approach by converting the ECG to a spectrogram and feeding this into a CNN. Furthermore, Zihlman *et al.* also introduces a long short memory block between the convolutional layers and the hidden layers as network architecture. Lou *et al.*<sup>25</sup> also converts the ECG to a spectrogram using a modified frequency slice wavelet transform before feeding the ECG to a stacked denoising auto-encoder for feature extraction, the result is then given to deep neural network for classification. Kachuee *et al.*<sup>26</sup> and Kiranyaz *et al.*<sup>27</sup> directly apply a CNN network using single unprocessed ECG beats as input to the network. The papers of Kiranyaz *et al.*, Kachuee *et al.*, Lou *et al.*, Rubin *et al.* and Isin and Ozdalili have the advantage of providing beat per beat classification.

There are multiple annotated human ECG datasets available, with the Physionet MIT-BIH arrhythmia dataset as one of the most commonly used databases in studies<sup>28,29</sup>. Since for eECGs such datasets are lacking we use transfer learning from a network trained upon the MIT-BIH dataset to an own generated dataset for horses with 20.000 beats which were recorded in a clinical setting. Transfer learning for ECGs has already successfully been used in different studies for human ECG classification, but has not yet been applied between different species<sup>17,21,27</sup>.



**Figure 2.** Equine ECG trace before (upper) and after (lower) filtering. A median filter has been applied in order to remove the low frequency baseline wander and a discrete wavelet filter was used to remove the high frequency noise.

Algorithm	False Positive (beats)	False Negative (beats)	Positive Predictive Value (%)	Recall (%)
Proposed	94	204	99.1	98.0
Pan-Tompkins	2156	737	81.5	91.5

**Table 1.** Comparison of proposed QRS detection algorithm with Pan-Tompkins<sup>32</sup> on 10356 equine QRS complexes.

## Results

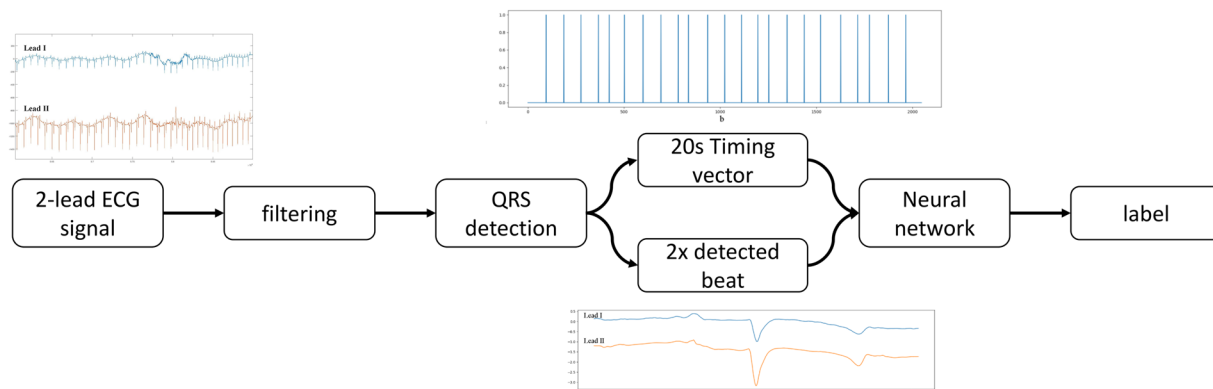
**Datasets.** Both the MIT-BIH and eECG dataset were split at random into a 60% training, 20% validation and 20% test dataset. The validation dataset was used as fitness value for the genetic algorithm and the validation and training dataset were used for training the final CNN. The test dataset was used for validation of the final trained CNN. Because both datasets are unbalanced in number of samples for each class, a resampling was applied for the training dataset. The MIT-BIH database has 15 different beat type classes, but since there is only limited knowledge and data about the different equine arrhythmias they were reallocated in the following 4 classes depending the underlying primary mechanism: Normal (N), atrial premature contraction (APC), ventricular premature contraction (VPC) or artefact (A).

The MIT-BIH dataset had 237704 beats for training the CNN, with 186874N, 11799 APC, 37153 VPC and 1878 A and 59427 samples for testing with 46841N, 2892 APC, 9232 VPC, 462A. The different classes were resampled to 24476 samples for each class for training, thus resulting in 97904 training samples in total. The eECG dataset was split in a training dataset of 21152 samples, with 15991N, 1087 APC, 3699 VPC and 375A which were resampled to 2393 samples of each class and 5288 samples for testing with 3972N, 255 APC, 971 VPC and 90A.

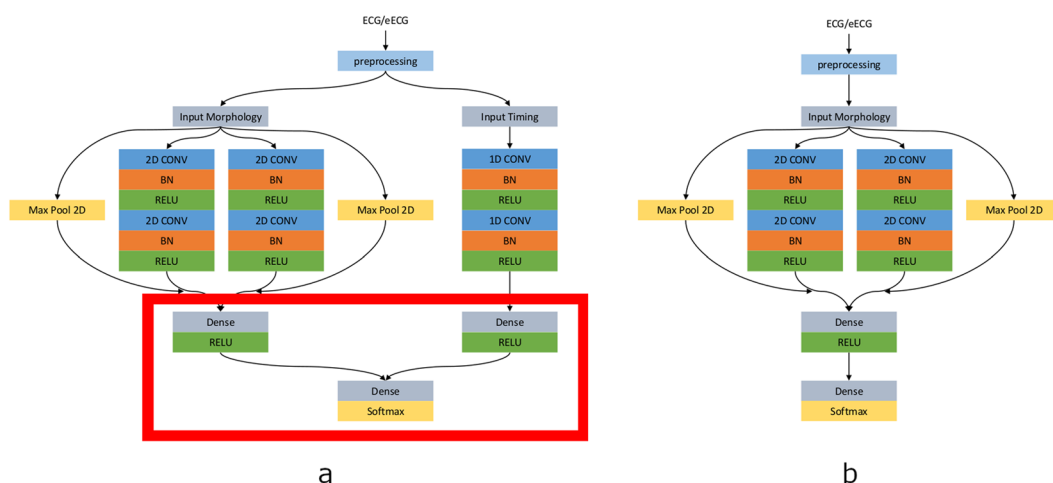
**Signal processing of the ECG signal.** The ECG signal was first pre-processed using a median filter to remove the baseline wander, followed by Discrete Wavelet transform to remove the remaining noise<sup>30,31</sup>. The result of the filtering is shown in Fig. 2. Next, 5 peaks in eECGs were detected based upon the modulus maxima of the Stationary Wavelet transform with a Symlet 4 wavelet and an adaptive threshold<sup>31</sup>. The detected beats were used for training and testing of the proposed CNN architectures.

The proposed QRS detection algorithm was compared to the popular Pan-Tompkins algorithm on 5 different 30 m eECG traces (10356 beats in total)<sup>32</sup>. The results of the comparison are shown in Table 1.

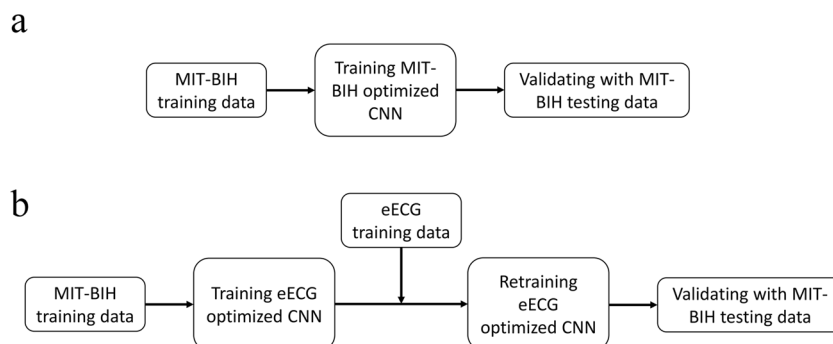
**The ECG classification algorithm.** The CNN were designed for a fixed network input of  $2 \times 500$  data points for the morphological input and 2000 data points for the timing input. The high-level architecture of the complete proposed ECG processing method is shown in Fig. 3 and the high-level architecture of the proposed CNN's is shown in Fig. 4. Weights were set using random weight initialization for the MIT-BIH dataset. For the eECG dataset, both randomly initialized weights and the weights from the pretrained CNN on the MIT-BIH dataset (transfer learning) were used for comparison. The used training method is shown in Fig. 5. A genetic algorithm (GA) was used to optimize the CNN parameters. The non-dominated sorting genetic algorithm II



**Figure 3.** High level architecture of the complete proposed ECG processing method. Two raw ECG leads are filtered and QRS detection is performed on lead II. After QRS detection 500 points around the detected beat, both from lead I and lead II, are extracted in addition with a vector representation of all detected QRS complexes 10s before and 10s after each beat. These inputs are presented to the network that performs the classification into 4 classes.



**Figure 4.** High level architecture of both proposed networks. Panel A shows the CNN with parallel timing pathway, panel B shows the CNN without timing pathway. The red rectangle highlights the layers that were retrained when only performing transfer learning on the top layers. CONV: convolutional block; BN: batch normalization; RELU: rectified linear activation unit; Max Pool 2D: Max pooling operation; Dense: fully connected layer; Softmax: Softmax activation block with the 4 different output classes.



**Figure 5.** Training method for obtaining the evaluation metrics for the MIT-BIH dataset (Panel a) and the eECG dataset (Panel b). The MIT-BIH dataset is directly trained on the MIT-BIH training set (80% of total dataset) with the MIT-BIH optimized CNN parameters for the proposed CNN architectures. For the eECG dataset, the same MIT-BIH training set was used for obtaining the initial weights with the eECG optimized CNN parameters. Afterwards the network is retrained with the eECG training data.

	CNN 1	CNN 2	CNN timing
<b>MIT-BIH</b>			
Layer 1			
Subsampling	4	2	128
Convolution width	60	20	300
Filters	16	40	4
Layer 2			
Subsampling	16	16	8
Convolution Width	30	20	100
Filters	16	32	8
Neurons in fully connected layers	512		512
Neurons in final fully connected layer	32		
<b>eECG</b>			
Layer 1			
Subsampling	8	2	32
Convolution width	10	30	50
Filters	16	8	64
Layer 2			
Subsampling	4	16	16
Convolution Width	50	20	30
Filters	64	64	4
Neurons in fully connected layers	512		512
Neurons in final fully connected layer	32		

**Table 2.** Results from the genetic algorithm after optimizing both on the MIT-BIH dataset and the eECG dataset.

(NSGA-II) was used with a population size of 20 and optimized during 10 generations with the accuracy as fitness value and training was ended after 50 epochs or when the training accuracy stopped improving for 3 epochs<sup>33</sup>. The following parameters were optimized: number of filters, width of convolution and subsampling for each individual convolutional layer, number of neurons for each fully connected layer, L2 regularization and dropout used for training.

The details of the best network architecture calculated by the GA are given in Table 2. This network achieved an accuracy of 97.7% on the eECG validation dataset. The results of the different datasets and with or without transfer learning from the MIT-BIH dataset to the eECG dataset are given in Table 3 in addition with the results of training without resampling the dataset. The transfer learning was done using 2 different methods: the first method updated the weights of all the layers of the network during the training, while the second method only updated the weights of the fully connected layers. With the second transfer learning method the convolutional layers are used as pre-trained feature extractors while the fully connected layers are re-trained as classification layers. Results from the non-parallel network are shown in Table 4. It took around 13 h 46 m for running the GA, 20 m for training on the MIT-BIH dataset and 4 m 30 s for training on the eECG dataset for the entire network and 18 s for only retraining the top fully connected layers. In order to evaluate the performance of our proposed architecture on the eECG dataset, both the resampled MIT-BIH and eECG dataset were trained and evaluated with the deep CNN architecture proposed by Kachuee *et al.*<sup>26</sup>. Results of the deep CNN architecture shown in Table 5.

## Discussion and Conclusion

Optimization of neural network hyperparameters is a time intensive task that can consume multiple weeks of computing time in order to combine all possible (realistic) parameters. Here we used a GA for optimization of the CNN parameters, which only consumed 13 h 46 m of computing time in order to run once without human interaction. The GA also achieved a good network reduction with relatively small network dimensions as can be seen in Table 2.

Since there are no published results of QRS detection algorithms on eECGs, we validated our proposed QRS detection algorithm against the Pan-Tompkins algorithm<sup>32</sup>. This algorithm is the most common used algorithm in literature for QRS detection on human ECG traces. As shown in Table 1 the Pan-Tompkins algorithm has a lower positive predictive value and sensitivity in comparison with our proposed algorithm. The high number of false positive results, and thus low positive predictive value, of the Pan-Tompkins algorithm is mainly explained due to the detection of the high T-wave as a QRS complex. In certain circumstances the detection of the T-wave obstructed the detection of the actual QRS complex, especially during higher heart rates during exercise, which explains the higher number of false negative results. Most false positive detected beats of the proposed algorithm

Class	Recall (%)	Positive Predictive Value (%)	F1-score (%)	Accuracy (%)
<b>Scores with resampling</b>				
<i>MIT-BIH</i>				97.7
N	99.4	97.3	98.4	
APC	70.1	94.8	80.6	
VPC	93.6	94.8	80.6	
A	81.9	90.6	86.0	
<i>eECG without transfer learning</i>				92.6
N	99.2	81.7	95.2	
APC	43.3	95.7	59.5	
VPC	97.7	96.2	96.9	
A	81.7	91.5	96.2	
<i>eECG with transfer learning on the entire network</i>				97.1
N	99.5	97.1	98.3	
APC	73.2	95.7	81.6	
VPC	97.1	97.8	97.4	
A	87.8	87.8	87.8	
<i>eECG with transfer learning on the top layers</i>				86.4
N	99.0	83.7	90.7	
APC	41.6	85.5	55.4	
VPC	75.4	97.5	84.9	
A	85.4	90.9	88.1	
<b>Scores without resampling</b>				
<i>MIT-BIH</i>				98.2
N	98.2	99.2	99.1	
APC	90.2	88.3	89.2	
VPC	97.8	98.2	97.6	
A	91.8	89.6	87.0	
<i>eECG with transfer learning</i>				98.2
N	99.0	99.1	99.0	
APC	90.9	92.6	91.7	
VPC	97.8	97.8	97.8	
A	92.2	81.5	85.1	

**Table 3.** Scores for the parallel network for different datasets, testing methods with or without resampling the datasets and different transfer learning techniques.

Class	Recall (%)	Positive Predictive Value (%)	F1-score (%)	Accuracy (%)
<b>Scores with resampling</b>				
<i>MIT-BIH</i>				96.7
N	99.7	96.8	98.2	
APC	69.0	94.4	79.7	
VPC	95.3	96.8	96.1	
A	82.0	89.4	85.5	
<i>eECG with transfer learning on the entire network</i>				91.9
N	99.5	90.5	94.7	
APC	43.5	96.7	59.1	
VPC	96.9	96.5	96.7	
A	81.0	90.0	85.3	

**Table 4.** Scores for the CNN architecture without parallel timing input.

were high frequency and high amplitude artefacts on the ECG trace which also induced a short blanking period due to the adaptive threshold of the algorithm. This blanking period accounts for most of the false negative beats of the proposed algorithm. The higher number of false positives would have implied a higher number of A, namely more T waves that would have been detected as QRS complexes, in the eECG dataset and thus a better

Class	Recall (%)	Positive Predictive Value (%)	F1-score (%)	Accuracy (%)
<b>Scores with resampling</b>				
<i>MIT-BIH</i>				95.9
N	99.3	96.3	97.8	
APC	68.5	91.4	78.3	
VPC	92.3	96.0	94.1	
A	77.3	86.3	81.5	
<i>eECG without transfer learning</i>				92.0
N	99.1	91.3	95.0	
APC	46.5	92.2	61.8	
VPC	96.0	96.1	96.0	
A	78.5	88.3	84.5	
<i>eECG with transfer learning on the entire network</i>				93.8
N	98.6	93.8	96.2	
APC	56.6	85.5	68.1	
VPC	94.1	96.9	95.4	
A	75.8	87.6	81.7	

**Table 5.** Scores for the deep CNN architecture proposed by Kachuee *et al.*<sup>26</sup>.

balance of the different classes since there are currently only 465A in the eECG dataset or 2.2% of the total number of beats. However in combination with the lower sensitivity this would have resulted in a lower number of total true positives and therefore have an impact on the 1D timing vector which would be less representative for the relative timing of each QRS around the detected beat. If the 1D timing vector would include more artefacts and/or lesser QRS complexes this could possibly affect the usefulness of the 1D timing vector for the overall classification algorithm.

Currently there are no published equine ECG classifications methods and/or published datasets, therefore it is difficult to benchmark the performance of the proposed method for equine ECGs. The intra- and interobserver agreement has been suggested to be similar as for human ECGs, but the accuracy of the individual equine physiologist has not yet been studied<sup>11</sup>. When comparing with the numerous described algorithms for human ECG classification, our proposed method shows good classification results with an accuracy of 97.7% for the MIT-BIH database, with an even a higher accuracy of 98.2% when not applying resampling to the training dataset. However it should be noted that direct comparison between publications is difficult due to the different metrics, classes and datasets that are used for evaluating the classification performance<sup>14</sup>. In addition, accuracy alone is not the most optimal measurement of performance for ECG classification due to the imbalanced nature of ECG data which is illustrated here with a 99:1 ratio for the N:A classes in the MIT-BIH training dataset and a 43:1 ratio for the N:A classes in the eECG dataset. Since neural networks are even more vulnerable for overfitting when the dataset is imbalanced we used a resampling technique for the training data in order to achieve better classification results for the individual classes as has been shown before<sup>23</sup>. As can be seen in Table 3, the resampling had a negative effect on both the accuracy and individual scores, which is in contradiction to previous publications. The average recall, PPV and F1-score decreased from 94.7%, 93.8% and 93.2% to 86.3%, 94.5% and 90.5% for the MIT-BIH database and changed from 95.0%, 92.7% and 93.4% to 89.4%, 94.3% and 91.8% for the eECG dataset. A higher accuracy without resampling can be expected due to overfitting of the neural network on the majority class, the decrease of the individual scores however is unexpected. The effect might be explained due our resampling technique that noticeably decreases the total number of training samples, from 237704 to 97904 for the MIT-BIH dataset and from 21152 to 9572 for the eECG dataset. In addition to the decrease of the total number of samples the network can be trained on, and thus loss of information for the network, samples of the smallest class are copied multiple times and overfitting for this class can be expected. We tested both theories by undersampling to even lower numbers per class, which even further lowered the overall accuracy, and oversampling to higher numbers per class which slightly increased the overall accuracy and individual scores, but never exceeded the scores without resampling. This indicates that no extra information was learned by the network by copying the smallest classes, but this increased the required learning time with 50%. Another method to decrease overfitting and exploit the entire dataset is ensembling<sup>23</sup>. Multiple trained networks of the same type are used which are trained and validated, with early stopping, on different but overlapping parts of the dataset, the individual predictions are combined by majority voting. We applied this technique on our datasets with the parallel network for our production model with 5 trained networks and achieved an increased accuracy of 99.1% for the MIT-BIH dataset and 98.7% for the eECG dataset with improved (1–3%) performance metrics for all classes. Further improvements could be achieved by creating larger ensembles.

The results in Table 4 indicate that our proposed architecture with parallel timing input achieves a higher classification performance on all metrics, this effect remains for both the MIT-BIH and eECG dataset. These results reflect the clinical importance of the relative timing of each beat in the ECG trace that is also used by the cardiologist for the interpretation of the ECG. Other publications also acknowledge the importance of the temporal relationship by creating recurrent neural networks, for example by using LSTM blocks in combination with a CNN

in 2 publications which had similar improvements in results as in our study by including the temporal relationship<sup>17,23</sup>. Larger CNN architectures with more layers or more parallel pathways, similarly to the inception block used in ECGNet, have been studied, but no further improvement in classification accuracy could be achieved<sup>17</sup>.

Because there are currently no other equine ECG classification methods published, we compared our method against the deep CNN network architecture of Kachuee *et al.*<sup>26</sup>. In order to validate our implementation of the deep CNN architecture, the deep CNN was also trained and validated on the MIT-BIH dataset as described in the paper. Since our implementation only use 4 classes instead of the 5 classes described in the original paper, our accomplished accuracy was slightly higher (95.9% vs. 93.4%) in comparison with the achieved accuracy in the paper as can be seen in Table 5. However the accuracy on all datasets remained lower in comparison with our proposed parallel CNN architecture.

When trained on the eECG dataset, the proposed network achieved a low accuracy of 92.6% when using randomly initialised weights. The results remained similar for both training with and without a resampled dataset, especially the recall and F1-score for the APC remained very low. Because the small and imbalanced dataset, the network is very vulnerable to overfitting despite the use of L2 regularisation and dropout layers. By using transfer learning from the similar but larger MIT-BIH dataset a significant improvement of accuracy and average recall, PPV and F1-score from 92.6%, 80.4%, 93.7% and 84.5% to 97.1%, 89.4%, 94.3% and 91.8% could be achieved due to the improved feature extraction of the CNN. However when the CNN layers were used a feature extractor and only the top classification layers were retrained for classification of the eECG dataset, the achieved average accuracy was even lower in comparison without transfer learning. This indicates that the features extracted from the MIT-BIH dataset, however being a good initial weight, cannot be directly applied to the eECG dataset without updating the weights of the acquired features. Transfer learning has been applied before for ECGs, both between datasets and for patient specific holter ECG interpretation, but by the knowledge of the authors never between species<sup>17,25,34</sup>. For the future, an improved, larger and better balanced, dataset should be used with the transfer learning for achieving even higher improvement of classification accuracy of eECGs with the current parallel network.

We propose two novel CNN architectures for ECG specific beat-to-beat classification using the beat morphology on 2 leads and a graphical representation of the timing of each beat. An equine specific filtering and QRS detection algorithm is proposed and validated against the popular Pan-Tompkins QRS detection algorithm. The classification model performance was validated on a human and an own-made equine dataset with and without resampling of the dataset for equally balanced classes. Transfer learning was used from the human to the equine dataset for improved classification performance for all classes of the equine dataset. Our results on both datasets were compared against the deep CNN network architecture of Kachuee *et al.*<sup>26</sup>. A genetic algorithm was used for successful optimization of the CNN parameters. By the best of the authors' knowledge this is the first equine specific ECG classification algorithm and this algorithm may help to improve diagnosis of possible life-threatening arrhythmias in horses. Applying the architecture to larger (equine) ECG datasets for improved classification performance and extending the number of classes for classification are interesting pathways to be explored in the future.

## Methods

Procedures on animals were approved by the Ethical Committee, Faculty of Veterinary Medicine, Ghent University (EC2016/35). All experiments were performed in accordance with the relevant guidelines and regulations.

**Dataset description.** The MIT-BIH arrhythmia database is one of the most used arrhythmia databases for research in ECG signal processing<sup>28,35</sup>. It contains 48 half-hour annotated, two-channel (lead II and modified V1, V2, V3, V4 or V5 leads) ambulatory ECG recordings, obtained from 47 patients sampled at 360 Hz per channel.

Because no datasets are available for eECGs we collected and annotated a dataset of 26.440 beats from 15 horses. Seven horses were used in a previous pacing study, in which ectopic depolarisations were introduced in the atrium and ventricle with electrical stimulation of the heart, resulting in an exact annotation of each beat<sup>36</sup>. The remaining 8 eECGs were selected from clinical cases based upon a high frequency of ectopic beats in order to include more clinical realistic timings of premature beats compared to the paced beats. Two veterinarians with experience in cardiology annotated these recordings using a Python based ECG annotation tool designed for this work. All eECGs were recorded at 500 Hz using the modified base-apex electrode configuration<sup>37</sup>. Because the eECGs were recorded at 500 Hz, the MIT-BIH dataset was upsampled in order to match eECGs. The ECG was subsampled to 100 Hz for the timing vector in order to reduce the memory and computational cost, the resolution remained 500 Hz for the morphology input. Lead I and II were used for the algorithm from the eECG dataset.

The median number of samples of each class was calculated for each dataset and the samples of each training class were randomly under- or oversampled with replacement in order to acquire the median number of samples for each class. This was done using the imbalanced-learn toolkit v 4.3<sup>38</sup>.

**eECG specific pre-processing.** Baseline wandering is the low frequency noise typically induced by movement and breathing of the horse and can induce large differences between the amplitudes of the beats. A median filter can be used to remove this artefact without chancing or disturbing the characteristics of the waveform. The original eECG signal was processed with a median filter of 200 ms width in order to remove P waves and QRS complexes, followed by a median filter of 600 ms width to remove T waves<sup>30</sup>. The signal resulting from the second filtering contains the baseline of the ECG signal and is subsequently subtracted from the original ECG signal to obtain an ECG signal without baseline wandering.

After removing the baseline wandering, the resulting ECG signal still contains other residual noise in the higher frequency range. In this study we used Discrete Wavelet transform using a Coiflet 2 wavelet to filter the



signal<sup>31</sup>. The signal was first decomposed into several subbands by applying Wavelet Transform with a Daubechies mother wavelet of order 8, since this has been proven to be the most appropriate wavelet basis function for denoising ECGs<sup>39</sup>. Next, all values below a certain threshold were set to 0, after which the signal was reconstructed. Using these pre-processing methods high frequency components of the ECG signal decrease as lower details are removed from the original signal. An example of an ECG trace filtered by a subsequent median and Discrete Wavelet filter is shown in Fig. 2.

**Detection of S peak in the eECG.** The detection of the S peak is based upon the modulus maxima of the Stationary Wavelet transform using a Symlet 4 wavelet<sup>31</sup>. A Symlet 4 wavelet was used to decompose the signal since this best represents the shape of a physiologic equine QRS complex. For detection of the S peak, only the quadrated values of coefficient detail scale 4 ( $cd4^2$ ) are used for applying a search for a maxima modulus line exceeding an adaptive threshold  $t_{QRS}$ <sup>40</sup>.  $t_{QRS}$  is calculated proportional to a rolling mean over 200 relative maxima of the selected detail coefficients. So for each 200 local maxima we take:

$$t_{QRS} = \frac{1.25}{200} \sum_{k=0}^{199} cd4_{n-k}^2 \quad (1)$$

If multiple values exceed  $t_{QRS}$  within 200 ms, the largest value is selected as the QRS complex. A value of 200 ms is selected because this corresponds to the shortest duration of the effective refractory period of the action potential in horses with a 99% confidence interval<sup>41</sup>. For this study, detection occurred on lead II and a sample was taken from 0.5 s before and after each detected S peak on lead I and II for input to the CNN.

Since there are no published results from eECG specific QRS detection algorithms, the proposed method was compared against the Pan-Tompkins algorithm<sup>32</sup>. This is one of the most common used QRS detection algorithms. Both the Pan-Tompkins algorithm and the proposed algorithm were implemented in Matlab 2018b (MathWorks). Both algorithms were run on 30 m eECGs of 5 different horses. The results were manually checked for correctness.

**Parallel convolutional neural network architecture for eECG and ECG classification.** Convolutional layers of different size have been shown to learn distinct feature representations and thus combining these features will provide better feature representation compared to one single filter<sup>17,42</sup>. Based on this, the input layers for the filtered signal is connected to 2 parallel paths with each identical architectures but different parameters. We propose a CNN architecture with two parallel input layers: one layer ingests the filtered signal of the detected beats of lead I and II in a  $2 \times 512$  matrix while a second layer ingests a 1D vector ( $1 \times 2048$  samples) representing the relative timing of each detected beat in a 20 s segment, using 1 for a detected beat and 0 otherwise. Sample inputs are shown in Fig. 3. By doing so, the network is forced to learn both features that represent the morphology, the filtered signal input, and the temporal relationship of the detected beat, the 1D vector, compared to the surrounding beats. Both of these features are essential for correct ECG interpretation. Most approaches for individual beat classification only process the individual beat, without information of the surrounding ECG trace. While some ECG classification strategies involve the processing of longer traces with multiple beats, thus extracting possible temporal relationship features, they do not offer the benefit of beat-to-beat classification. Each parallel path consists of 2 2D convolutional layers, each followed by a batch normalization layer, a rectified linear unit activation (ReLU) block and a dropout layer<sup>43–45</sup>. The CNN layer performs convolution with a convolution width kernel size of  $2 * \text{convolution width}$  with  $1 * \text{subsampling strides}$ , so no interaction exists between both leads, the convolution width and subsampling stride are shown in Table 2 for each network. Zero-padding was used after convolution in order to keep the temporal order of the input signal. The batch normalization layers normalize the learned features of the previous CNN layer resulting in reduced overfitting and improved learning speed. The ReLU layer permits to add a non-linearity to the network which favours a deeper representation. The dropout layer improves the generalization capabilities of the network and avoids co-adaptation by randomly setting a fraction ( $=1 - p$ ) of the neurons to zero. Based upon the residual like architecture of Rajpurkar *et al.*<sup>18</sup>, a residual connection is used within each path in a similar manner to those found in the Residual Network architecture<sup>46</sup>. This shortcut connection subsamples the input using a Max Pooling operation with the same subsample factor as the combined convolutional layers of that path<sup>47</sup>. The 2 parallel paths are flattened, concatenated and connected to a fully connected layer with a ReLU layer.

A separate parallel pathway processes the 1D vector with the relative timing information of the detected beat. The path consists of two 1D convolutional layers each followed by a dropout layer, no improvements in overall accuracy could be achieved by adding batch normalization blocks and residual connections as was done for the morphology paths. The last dropout is also connected to a fully connected layer with a ReLU layer. The outputs of the fully connected layers of the morphology and timing pathway are concatenated and connected to a final fully connected layer with soft-max activation in order to produce a distribution over the 4 output classes for each detected beat. The soft-max layer is used to perform closed-set identifications. The output of this layer is an integer label correlating to a predefined class. By adding a dedicated fully connected layer for each parallel pathway, the extracted features can be processed by a 'specialized' layer for each type of feature, timing or morphology, before feeding it into a general fully connected classification layer. Both a CNN with and without parallel pathway for timing were built and tested. The high-level architectures of both networks are shown in Fig. 4.

Since there are no other published equine ECG classification methods to compare against, the deep CNN architecture of Kachuee *et al.* was implemented and trained on our eECG dataset<sup>26</sup>. This architecture was chosen because of both the good accuracy on the MIT-BIH dataset and the ability to perform beat per beat classification. All deep CNN network parameters were set as described in the paper.

**Training method.** The softmax cross entropy was used as loss function for training. The Adam optimizer with the default parameters described in the paper and with a learning rate of 0.0001 was used for updating the weights<sup>48</sup>. Dropout and L2 regularization was set to 0.2 and 0.001 respectively. Batch size was fixed to 500 and number of epochs to 20. The networks were build using Keras with Tensor Flow backend<sup>49</sup>.

The deep CNN was trained as described in the paper with the Adam optimizer, with the learning rate, beta-1, and beta-2 of 0.001, 0.9 and 0.999, respectively<sup>26</sup>. Learning rate is decayed exponentially with a decay factor of 0.75 every 10000 iterations. The MIT-BIH optimised CNN parameters were used for evaluating the performance on the MIT-BIH dataset, the eECG optimized CNN parameters were used for evaluating the performance on the eECG dataset. The training method is visualised in Fig. 5.

Transfer learning offers a useful solution for models were limited training data, a lack of expertise for training and moderate computer resources for training are available. In transfer learning a pre-trained network is used for a different task, with fine-tuning of some or all layers with the data from the different task. With a CNN, the convolutional layers can be reused as a feature extractor while the final layers (typically deep layers) can be retrained for a different classification task. Since the eECG dataset (26,440) is significantly smaller compared to the MIT-BIH ECG dataset (297,131 beats) transfer learning was applied from the MIT-BIH trained model to the eECG trained model. After training the complete model on the MIT-BIH ECG dataset, transfer learning to the eECG dataset was applied in two different ways:

- All weights of the pre-trained network are updated. After training the model on the MIT-BIH dataset all weights are reused as initial weights for training on the eECG dataset, so both the feature extractor (convolutional layers) and classification part (fully connected layers) are adapted to the eECG dataset.
- Using the pre-trained MIT-BIH network as a feature extractor. In this case all layers were frozen for training after training on the MIT-BIH dataset (i.e. the weights were not updated during training) except the fully connected layers. By doing so the pre-trained CNN acts as a feature extractor for the ECG and only the classification layers are retrained.

**Evaluation metrics.** Standard metrics for classification tasks were used for evaluation: accuracy (ACC), Recall; positive predictive value (PPV) and F1-score, which are defined next:

$$\begin{aligned}
 ACC &= \frac{(TP + TN)}{(TP + TN + FP + FN)} \\
 Recall &= \frac{TP}{(TP + FN)} \\
 PPV &= \frac{TP}{(TP + FP)} \\
 F1 - score &= \frac{2 * R * PPV}{(R + PPV)}
 \end{aligned}
 \tag{2}$$

TP: true positive, TN: true negative, FP: false positive, FN: false negative, R: recall.

Each CNN was trained 3 times with identical initial parameters, the metrics were averaged over the results of these 3 trainings. The following tests were conducted:

- Each CNN (deep CNN, proposed CNN with and without parallel timing pathway) was tested with the resampled MIT-BIH and eECG dataset.
- The resampled eECG dataset was tested with both the network without transfer learning and with both transfer learning methods.
- In order to test the effect of the resampling, both the MIT-BIH and eECG dataset without resampling were used to train the parallel and deep CNN network with transfer learning from MIT-BIH to eECG.

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 5 July 2019; Accepted: 20 December 2019;

Published online: 13 January 2020

## References

1. van Loon, G. & Pattenon, M. *Electrophysiology and arrhythmogenesis. Cardiology of the Horse* (Elsevier Ltd, 2010).
2. Ryan, N., Marr, C. M. & Mcgladdery, A. J. Survey of cardiac arrhythmias during submaximal and maximal exercise in Thoroughbred racehorses. *Equine Vet. J.* **37**, 265–268 (2010).
3. Physick-Sheard, P. W. & McGurrin, M. K. J. Ventricular arrhythmias during race recovery in standardbred racehorses and associations with autonomic activity. *J. Vet. Intern. Med.* **24**, 1158–1166 (2010).
4. Buhl, R., Petersen, E. E., Lindholm, M., Bak, L. & Nostell, K. Cardiac arrhythmias in standardbreds during and after racing-possible association between heart size, valvular regurgitations, and arrhythmias. *J. Equine Vet. Sci.* **33**, 590–596 (2013).
5. Barbesgaard, L., Buhl, R. & Meldgaard, C. Prevalence of exercise-associated arrhythmias in normal performing dressage horses. *Equine Vet. J.* **42**, 202–207 (2010).

6. Lyle, C. H. *et al.* Risk factors for race-associated sudden death in Thoroughbred racehorses in the UK (2000–2007). *Equine Vet. J.* **44**, 459–465 (2012).
7. Corrado, D., Basso, C., Schiavon, M., Pelliccia, A. & Thiene, G. Pre-Participation Screening of Young Competitive Athletes for Prevention of Sudden Cardiac Death. *J. Am. Coll. Cardiol.* **52**, 1981–1989 (2008).
8. Lyle, C. H. *et al.* Sudden death in racing Thoroughbred horses: An international multicentre study of post mortem findings. *Equine Vet. J.* **43**, 324–331 (2011).
9. Martinez, M. W., Ayub, B. & Barrett, M. Is Auscultation the Achilles' Heel of Sports Screening? -Evidence on the Proficiency of Primary Care Physicians. In *Circulation* **126**, A9502 (2012).
10. Harmon, K. G., Zigman, M. & Drezner, J. A. The effectiveness of screening history, physical exam, and ECG to detect potentially lethal cardiac disorders in athletes: A systematic review/meta-analysis. *J. Electrocardiol.* **48**, 329–338 (2015).
11. Trachsel, D. S., Bitschnau, C., Waldern, N., Weishaupt, M. A. & Schwarzwald, C. C. Observer agreement for detection of cardiac arrhythmias on telemetric ECG recordings obtained at rest, during and after exercise in 10 Warmblood horses. *Equine Vet. J.* **42**, 208–215 (2010).
12. Banerjee, S. & Mitra, M. Application of cross wavelet transform for ECG pattern analysis and classification. *IEEE Trans. Instrum. Meas.* **63**, 326–333 (2014).
13. Lanata, A., Guidi, A., Baragli, P., Valenza, G. & Scilingo, E. P. A Novel Algorithm for Movement Artifact Removal in ECG Signals Acquired from Wearable Systems Applied to Horses. *PLoS One* **10**, e0140783, <https://doi.org/10.1371/journal.pone.0140783> (2015).
14. Kaplan Berkaya, S. *et al.* A survey on ECG analysis. *Biomed. Signal Process. Control* **43**, 216–235 (2018).
15. Meyling, H. A. & Ter Borg, H. The conducting system of the heart in hoofed animals. *Cornell Vet.* **47**, 419–455 (1957).
16. Acharya, U. R. *et al.* A deep convolutional neural network model to classify heartbeats. *Comput. Biol. Med.* **89**, 389–396 (2017).
17. Murugesan, B. *et al.* ECGNet: Deep Network for Arrhythmia Classification. *IEEE International Symposium on Medical Measurements and Applications (MeMeA)* 1–6, <https://doi.org/10.1109/MeMeA.2018.8438739> (2018).
18. Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C. & Ng, A. Y. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. *Computer vision and Pattern Recognition* Available at, <https://arxiv.org/pdf/1707.01836.pdf> (2017).
19. Bai, S., Kolter, J. Z. & Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. Available at, <https://arxiv.org/abs/1803.01271> (2018).
20. Vincent, T. L. *et al.* Retrospective study of predictive variables for maximal heart rate (HRmax) in horses undergoing strenuous treadmill exercise. *Equine Vet. J.* **38**, 146–152 (2006).
21. Isin, A. & Ozdalili, S. Cardiac arrhythmia detection using deep learning. *Procedia Comput. Sci.* **120**, 268–275 (2017).
22. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 1097–1105, <https://doi.org/10.1016/B978-008046518-0.00119-7> (2012).
23. Zihlmann, M., Perekrestenko, D. & Tschannen, M. Convolutional Recurrent Neural Networks for Electrocardiogram Classification. *Comput. Cardiol.* (2010). **44**, 1–4 (2017).
24. Rubin, J., Parvaneh, S., Rahman, A., Conroy, B. & Babaeizadeh, S. Densely Connected Convolutional Networks and Signal Quality Analysis to Detect Atrial Fibrillation Using Short Single-Lead ECG Recordings. *Comput. Cardiol.* (2010). **44**, 2–5 (2017).
25. Luo, K., Li, J., Wang, Z. & Cuschieri, A. Patient-Specific Deep Architectural Model for ECG Classification. *J. Healthc. Eng.* **2017**, 1–13 (2017).
26. Kachuee, M., Fazeli, S. & Sarrafzadeh, M. ECG heartbeat classification: A deep transferable representation. *2018 IEEE International Conference on Healthcare Informatics (ICHI)* 443–444, <https://doi.org/10.1109/ICHI.2018.00092> (2018).
27. Kiranyaz, S., Ince, T., Hamila, R. & Gabbouj, M. Convolutional Neural Networks for patient-specific ECG classification. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 2608–2611 (2015).
28. Merone, M., Soda, P., Sansone, M. & Sansone, C. ECG databases for biometric systems: A systematic review. *Expert Syst. Appl.* **67**, 189–202 (2017).
29. Goldberger, A. L. *et al.* PhysioBank, PhysioToolkit, and PhysioNet. *Circulation* **101**, E215–20 (2000).
30. De Chazal, P. *et al.* Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE Trans. Biomed. Eng.* **50**, 686–696 (2003).
31. Sasikala, P., Wahidabanu, R. & Robust, R. Peak and QRS detection in Electrocardiogram using wavelet transform. *Int. J. Adv. Comput. Sci. Appl.* **1**, 48–53 (2010).
32. Pan, J. & Tompkins, W. J. A Real-Time QRS Detection Algorithm. *IEEE Trans. Biomed. Eng.* 230–236 (1985).
33. Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182–197 (2002).
34. Kiranyaz, S., Ince, T. & Gabbouj, M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Trans. Biomed. Eng.* **63**, 664–675 (2016).
35. Moody, G. B. & Mark, R. G. The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.* **20**, 45–50 (2001).
36. Van Steenkiste, G., De Clercq, D., Vera, L. & van Loon, G. Specific 12-lead electrocardiographic characteristics that help to localize the anatomical origin of ventricular ectopy in horses: preliminary data. *Journal of Veterinary Internal Medicine* **33**, 1548, <https://doi.org/10.1111/jvim.15447> (2019).
37. Verheyen, T. *et al.* Electrocardiography in horses – part 1: how to make a good recording. *Vlaams Diergeneesk. Tijdschr.* **79**, 331–336 (2010).
38. Lemaitre, G., Nogueira, F. & Aridas, C. K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **18**, 1–5 (2017).
39. Singh, B. N. & Tiwari, A. K. Optimal selection of wavelet basis function applied to ECG signal denoising. *Digit. Signal Process.* **16**, 275–287 (2006).
40. Merah, M., Abdelmalik, T. A. & Larbi, B. H. R-peaks detection based on stationary wavelet transform. *Comput. Methods Programs Biomed.* **121**, 149–160 (2015).
41. De Clercq, D., Broux, B., Vera, L., Decloedt, A. & van Loon, G. Measurement variability of right atrial and ventricular monophasic action potential and refractory period measurements in the standing non-sedated horse. *BMC Vet. Res.* **14**, 101–118 (2018).
42. Szegedy, C. *et al.* Going Deeper with Convolutions. **18**, 186–191, <https://doi.org/10.1089/pop.2014.0089> (2014).
43. Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Available at, <http://arxiv.org/abs/1502.03167> (2015).
44. Nair, V. & Hinton, G. E. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on Machine Learning*, 807–814 (2010).
45. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
46. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *IEEE Conf.* **32**, 428–429 (2015).
47. Scherer, D., Müller, A. & Behnke, S. Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. In *20th International Conference on Artificial Neural Networks*, 92–101 (2010).
48. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. In *ICLR 2015*, 1–15 (2014).
49. Chollet, F. Keras, <https://keras.io> (2015).

## Acknowledgements

Glenn Van Steenkiste is a PhD fellow of the Research Foundation Flanders (FWO-Vlaanderen) with grant 1S56217N.

## Author contributions

G.C. and G.v.L. conceived the study. G.V.S. designed the algorithm with contributions from G.C. and G.v.L. G.V.S. annotated the equine dataset, performed the experiments and wrote the manuscript. All authors read and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to G.V.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020