Fetal Gender Identification using Machine and Deep Learning Algorithms on Phonocardiogram Signals

Reza Khanmohammadi 1, Mitra Sadat Mirshafiee 2, Mohammad M. Ghassemi 3*, Tuka Alhanai 4*

Department of Computer Engineering, University of Guilan, Iran 1
Department of Industrial Engineering, Alzahra University, Iran 2
Department of Computer Science, Michigan State University, USA 3
Department of Electrical & Computer Engineering, New York University Abu Dhabi, UAE 4
rezanecessary@gmail.com, mitra.mirshafiee@gmail.com, ghassem3@msu.edu, tuka.alhanai@nyu.edu

Abstract

Phonocardiogram (PCG) signal analysis is a critical, widely-studied technology to noninvasively analyze the heart’s mechanical activity. Through evaluating heart sounds, this technology has been chiefly leveraged as a preliminary solution to automatically diagnose Cardiovascular diseases among adults; however, prenatal tasks such as fetal gender identification have been relatively less studied using fetal Phonocardiography (FPCG). In this work, we apply common PCG signal processing techniques on the gender-tagged Shiraz University Fetal Heart Sounds Database and study the applicability of previously proposed features in classifying fetal gender using both Machine Learning and Deep Learning models. Even though PCG data acquisition’s cost-effectiveness and feasibility make it a convenient method of Fetal Heart Rate (FHR) monitoring, the contaminated nature of PCG signals with the noise of various types makes it a challenging modality. To address this problem, we experimented with both static and adaptive noise reduction techniques such as Low-pass filtering, Denoising Autoencoders, and Source Separators. We apply a wide range of previously proposed classifiers to our dataset and propose a novel ensemble method of Fetal Gender Identification (FGI). Our method substantially outperformed the baseline and reached up to 91% accuracy in classifying fetal gender of unseen subjects.

Contributions

In this work, we take a step towards automatic Fetal Gender Identification (FGI) with an understudied approach of using Fetal Phonocardiogram (FPCG) signals. More specifically, we experiment with multiple ML and DL pipelines containing a variety of preprocessing, feature extraction, and classification techniques to map fetal cardio patterns to the gender of the fetus. However, PCG signals suffer from various noise and require certain denoising techniques to be initially applied for signals to become gender-detectable. We train a Denoising Autoencoder and a neural source separation technique to measure the effects of two fundamentally different audio cleansing techniques. Afterwards, we evaluate top PCG classification methodologies in CVD detection on our gender-tagged FPCG dataset to establish a firm FGI baseline. Finally, we propose a novel ensemble method that outperforms the rest of the classifiers in FGI. We use medically accepted metrics such as accuracy, sensitivity, and specificity (Trevethan 2017) to evaluate the efficiency of each method in the hold-out and leave one out settings.

The remaining of this paper is organized as follows: first, we describe previous studies in the Related Work section. Then we fully describe our approach in the Methodology section and compare the result with previous work in Results & Discussion before concluding the paper in the last section.

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
Related Work

Advantages of Phonocardiography Signals

Automatic classification of heartbeat signals has attracted considerable interest due to the practicality and effectiveness of proposed approaches in detecting Cardiovascular Diseases (CVDs) (Ismail, Siddiqi, and Akram 2018). Generally, the principal difference between the majority of such research is the monitoring method with which the heartbeats have been collected. Among noninvasive methods, Electrocardiography (ECG) has been relatively more studied than expensive methods such as Magnetocardiography (MCG) or less accurate methods such as Photoplethysmography (PPG) (Avanzato and Beritelli 2020). Because of the timing and sequence of the heart’s depolarization and repolarization, ECG signals can be collected by putting electrodes on different parts of the body surface to detect the electrical activity of different portions of the heart. Such signals consist of three major waves, which are recalled as P wave, QRS complex, and T wave, which respectively represent the impact of atrial depolarization, ventricular depolarization, and ventricular repolarization in the final ECG signal. Still, extracting the FHR using this method has shown to be rather time-consuming and difficult to apply on pregnant women (Samieinasab and Sameni 2018). In contrast, Phono-cardiography signals, which record the mechanical activity of the heart during cardiac cycles as murmurs and sounds, are known to be an entirely passive, cost-effective alternative. When obtained from adults, these waveforms contain two fundamentally normal sounds (namely, S1 and S2) and two other potentially abnormal sounds (S3 and S4). During development, the fetal heart differs from the adult heart, where the Patent Doramen Ovale (PDO) allows the blood flow to bypass the lungs (Remien and Majmundar 2021), which ends up FPCG signals having their unique characteristics. However, such a dissimilarity is not the only challenge that is posed in FPCG analysis.

Noise Processing Techniques

A certain impediment to adopting PCG signals in computational tasks is the necessity to handle various low and high-frequency noise through preprocessing (Yang et al. 2020). As (Tseng et al. 2021) suggests, noise generally originates either from the recording devices (instrumental noise) or the surroundings (ambient noise). The first can be a result of the microphone's internal electronics' functionality; however, its adverse effects are relatively minor in an analogy to the latter where a broad range of sources can lower signal quality. Background noise, muscle contractions, and respiratory sounds (Li et al. 2020b) are noises that are challenging for researchers to process and derive meaningful information from PCG signals (Leal et al. 2018; Pham et al. 2018). Since signal quality heavily influences the derivation of features (Li et al. 2020b), researchers have investigated denoising techniques to come up with proper filters that differentiate heartbeat signals from the unwanted (Tseng et al. 2021). Common PCG denoising filters are either static such as Butterworth low-pass filtering (Gomes and Pereira 2012), or adaptive such as the least mean square algorithm (Dewangan and Potdar 2014). Empirical Mode Decomposition (EMD), Intrinsic Mode Functions (IMFs), and Nonnegative Matrix Factorization (NMF) are among other techniques that have been successfully employed to individually/jointly clean a given signal (Pham et al. 2018). Other solutions leverage source separators (Guo et al. 2021), or autoencoder networks (Chuang et al. 2019), to distinguish heartbeat sounds - the latter technique is one we will further investigate in the Methodology section of this paper.

Featurizing Phonocardiography Signals

The success of PCG signal classification heavily depends on advancements in machine learning (Li et al. 2020b). Classifiers such as Support Vector Machines (SVM) (Cheng et al. 2019) or k-Nearest Neighbours (KNN) (Juniati et al. 2018) have previously succeeded in detecting abnormalities such as murmurs or extrasystole. However, these advancements heavily rely on handcrafted discriminant features that get fed as inputs to such models (Yang et al. 2020). In our case, PCG signals are characteristically two-folded: (1) By adopting a signal processing perspective, studies have investigated features such as mean, variance, or power in each of the time (Deng and Han 2016), frequency (Safara et al. 2013), and time-frequency domains (Thiyagaraja et al. 2018). (2) Whereas another point to process these recordings is to capture their sound features. Mel Spectrogram (Chowdhury, Poudel, and Hu 2020), and Mel-frequency Cepstrum (MFCC) (Yaseen, Son, and Kwon 2018) are among the most widely leveraged feature representations techniques defining sound characteristics using a specifiable number of coefficients.

Depending on the problem of interest, conditions regarding different PCG datasets, and applied preprocessing techniques, studies have focused on each individually or combined the two to train classifiers (Shi et al. 2019). Based on previous results and achievements, there is greater confidence that CVDs can almost accurately be detected using state-of-the-art ML-based methodologies.

Deep Learning

In the meantime, DL algorithms have followed the same trend and are no exceptions to this sense of confidence (Chen et al. 2021). Standing on Neural Networks' superiority where the extraction and selection of features are eradicated to be individually stepped through, plenty of studies have experimented with Multilayer Perceptron (MLP) (Raza et al. 2019), Convolutional Neural Networks (CNN) (Wu et al. 2019), and Recurrent Neural Networks (RNN) (Latif et al. 2020). Unlike ML algorithms, no handcrafted features are fed to these networks and instead, low-level sets of features are extracted to represent the signals. MFCC (Rubin et al. 2016), Spectrogram (Demir et al. 2019), or Constant-Q transform (Ijwari et al. 2021) are among the most frequently used sound feature representations that have substantially aided related studies. However, the favorability of their outcomes has heavily been dependent on the abundance of informative data (Khan et al. 2021). An interesting research line has recently tackled PCG signal classification using CNNs on the image represented sound features (Ren...
In this setting, low-level sound representations are converted into images, and CNNs are asked to extract image features to classify heartbeats. Widely employed in image processing tasks, pretrained image networks such as VGGNet (Simonyan and Zisserman 2015) and ResNet (He et al. 2015) have also been leveraged to improve classification accuracy. Research on CVD detection using Transfer Learning (Alafif et al. 2020) has allowed researchers to more effectively handle smaller datasets where scarce data cannot be fed to neural networks directly, thereby unfavourably improving the chances of overfitting (Boulares, Alafif, and Barnawi 2020).

**Methodology**

To classify fetal gender using FPCG signals, we use the version 1.0.1 of the publicly available Shiraz University Fetal Heart Sounds Database (Samieinasab and Sameni 2015). This database contains PCG signals of 109 pregnant women and FPCG signals recorded by placing a JABEZ™ electronic stethoscope on their lower maternal abdomen. There were 7 cases of pregnancies where the mother was carrying a twin, which we excluded in this work, resulting in 102 samples which further are broken down to 7-second samples, among which a balanced number of male and female signals are considered to form our final dataset of 1000 FPCG samples. The original sampling rate of these recordings is 16,000 Hz and are, on average, 90 seconds long. A large number of previous studies have experimented with PCG samples of 5 to 10 seconds using major datasets (e.g. PhysioNet (Liu et al. 2016) or PASCAL (Gomes et al. 2013)). To balance between the informativeness of chunked samples and their final count, we chose seven seconds as the maximum length of each sound sample. To do so, we used the peak detection algorithm of the SciPy Python library (Virtanen et al. 2020) to find positive (heartbeats) and negative (silence) peaks from sound files and thus avoid cutting a signal when the heart beat cycle was not complete. To further describe our work, hereafter, we step through a pipeline to denoise, classify, and stack models for better efficiency.

**Denoising**

Removing noise is of great importance since our dataset's FPCG signals contain samples among which heartbeats may either have a high or low frequency. This condition also reigns in terms of noise, where the low-frequency (e.g. the friction between the stethoscope’s surface and human skin) and high-frequency (e.g. background noise) noise are diversely distributed. Hence, in our case, simply low-pass filtering (Raza et al. 2019) FPCG signals using static thresholds cannot distinguish heartbeats from noise. Instead, what we need is an adaptive filter that intelligently differentiates the two. Hence, we experiment with the following well-studied techniques to do so.

**Denoising Autoencoders (DAE):** An autoencoder consists of an encoder ($e$) and a decoder ($d$) that are connected through an encoded state. The encoder learns to map an input vector $x(i)$ into a compressed mid-level representation $z(i)$ such that the decoder is better modelled to reconstruct output vector $y(i)$ from $z(i)$.

\[
    z[i] = e(x[i]) \quad (1)
\]

\[
    y[i] = d(z[i]) \quad (2)
\]

Given two parallel sets of noisy and clean signals, we can train an autoencoder network with the objective to denoise signals. Here we use the PASCAL dataset, which contains plenty of clean PCG signals and common clinical artifacts. This way, through combining a random artifact with a clean PCG signal, we obtain a noisy $x(i)$ from which the autoencoder will learn to generate clean heartbeats as $y(i)$.

**Single-Channel Blind Source Separation (SCBSS):** we use a deep learning framework called DeWave (Hershey et al. 2013) to apply SCBSS on FPCG signals using the deep clustering algorithm. To separate audio sources, this method uses embedding vectors for each time-frequency bin. As indicated in equation 3, they have designed their objective function to make embedding vectors $v_i$ more similar/dissimilar when two bins do/do not belong to the same class.

\[
    C(\theta, Y) = |VV^T - YY^T|_F^2 = \sum_{i,j:y_i=y_j} \left( |v_i - v_j|^2 - 1 \right) + \sum_{i,j} \frac{1}{4} \left( |v_i - v_j|^2 - 2 \right)^2 \quad (3)
\]

Given $VV^T$ as a low-rank affinity matrix and $YY^T$ as an ideal affinity matrix, network parameters $\theta$ are trained to minimize the objective function. Whereas the first term compacts bins of the same class, the second term spreads out bins of different classes. To apply this technique, we need
to train the network on two distinct audio sources to model certain patterns of each. We again use the PASCAL dataset by feeding its artifact samples and pure heartbeats as two audio sources to the framework. At inference, given a single noisy heartbeat sound signal \( s_n(i) \), we obtain two audio files named \( s_h(i) \) (heartbeat signal) and \( s_p(i) \) (noise signal) where each belongs to one of the two audio sources.

\[
s_h[i], s_n[i] = w(s_i[i])
\]

Since the DeWave method \( w \) was initially proposed and evaluated on speech data, its effects have not been fully investigated in auscultation. However, in terms of our FPCG signals, they perform interestingly well in detecting background, instrumental, and continuous noise, whereas they fail to capture sudden noisy fluctuations. Next, to better detect mishandled high-frequency noise, we use a spectral gating noise reduction technique \( g \) (Sainburg, Thielk, and Gentner 2020). In this case, we used the stationary variant of this technique since a statically estimated threshold captures our noise of interest much better. Finally, we merge \( s_n(i) \) with \( g(s_n(i)) \) that of the stationary filter to obtain a final noise waveform \( m(i) \).

\[
m[i] = s_n[i] + g(s_n[i])
\]

This way, given a noisy FPCG signal as \( s_n(i) \), we obtain \( m(i) \), which is the merged waveform of \( s_n(i) \) high and low-frequency noise. To filter a waveform inside another, we take \( F \) where

\[
F = \left[ \sum_{i=1}^{N}(x_i - \bar{x})^2 \right] / (N - 1)
\]

1. **Variance:** Given \( \bar{x} \) as the mean of the signal and \( N \) as the number of time steps, this feature demonstrates how spread-out a signal is.

\[
variance = \sigma^2 = \frac{\sum(x - \bar{x})^2}{N - 1}
\]

2. **Skewness:** As \( \text{Tang et al. 2018} \) suggests, this feature assumes that noise has a different probability distribution than heartbeats.

\[
Skewness = \frac{3(\bar{x} - \tilde{x})}{\sigma}
\]

3. **Kurtosis:** Determines how "tailed" the probability distribution is.

\[
kurtosis = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{(N - 1)s^4}
\]

4. **Spectral Entropy:** Entropy is a measurement of how peaky and disorganized a signal is (Toh, Togneri, and Nordholm 2005). When extracted from PCG signals, this feature also assumes heartbeat signals having low-frequencies whereas high-frequency noise is uniformly distributed. Where \( P(\omega_t) \) stands for the spectrum of a signal’s, the Probability Density Function (equation ???) is calculated with which the Power Spectral entropy can be now calculated (equation [11]).

\[
p_i = \frac{1}{\sum} |P(\omega_i)|^2
\]

\[
PSE = -\sum_{i=1}^{n} p_i \ln p_i
\]

5. **Energy:** The Energy feature (equation [12]) captures sudden changes in mechanical energy of the signal (Kudriavtsev, Polyshchuk, and Roy 2007).

\[
E = \sum_{n=-\infty}^{\infty} |x[n]|^2
\]

6. **Root Mean Square (RMS):** RMS is used to measure continuous energy, which is obtained through a square root operation of the mean of the audio signal squared.

\[
RMS = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{N}}
\]

7. **Zero-crossing rate (ZCR):** As a popular acoustic feature in speech processing, ZCR informs us of the number of times the signal crosses the horizontal axis.

\[
ZCR_t = \frac{1}{2} \sum_{k=t-k}^{t+t-K} |\text{sgn}(x(k)) - \text{sgn}(x(k+1))|
\]
| Classifier | Input | Denoising method | Classifier | Hold-out | LOOCV |
|------------|-------|------------------|------------|----------|-------|
|            |       | SCBSS            | KNN        | Acc      | PR    |
|            |       |                  | SVM        | 0.59     | 0.38  |
| C<sub>R</sub>(x) | R(x)  |                  | XGB        | 0.56     | 0.47  |
|            |       |                  | LDA        | 0.66     | **0.55** | **0.8** | **0.57** |
|            |       |                  | LR         | 0.65     | 0.75   |
|            |       |                  | DAE        | KNN      | 0.56   |
|            |       |                  | SVM        | 0.6      | 0.47   |
|            |       |                  | XGB        | 0.55     | 0.51   |
|            |       |                  | LDA        | 0.62     | 0.55   |
|            |       |                  | LR         | 0.50     | 0.45   |
| C<sub>M</sub>(x) | M(x)  | SCBSS            | KNN        | 0.52     | 0.39   |
|            |       |                  | SVM        | 0.53     | 0.41   |
|            |       |                  | XGB        | **0.66** | **0.54** | **0.75** | **0.61** |
|            |       |                  | LDA        | 0.58     | 0.51   |
|            |       |                  | LR         | 0.54     | 0.46   |
|            |       |                  | DAE        | KNN      | 0.59   |
|            |       |                  | SVM        | 0.58     | 0.40   |
|            |       |                  | XGB        | 0.65     | 0.58   |
|            |       |                  | LDA        | 0.61     | 0.46   |
| C<sub>M,F</sub>(x) | MF(x) | SCBSS            | KNN        | 0.66     | 0.66   |
|            |       |                  | SVM        | **0.71** | **0.58** | **0.86** | **0.62** |
|            |       |                  | XGB        | 0.67     | 0.58   |
|            |       |                  | LDA        | 0.59     | 0.51   |
|            |       |                  | LR         | 0.57     | 0.48   |
|            |       |                  | DAE        | KNN      | 0.59   |
|            |       |                  | SVM        | 0.58     | 0.40   |
|            |       |                  | XGB        | 0.65     | 0.58   |
|            |       |                  | LDA        | 0.61     | 0.46   |
| C<sub>Q</sub>(x) | Q(x)  | SCBSS            | KNN        | 0.67     | 0.52   |
|            |       |                  | SVM        | 0.60     | 0.34   |
|            |       |                  | XGB        | 0.63     | 0.52   |
|            |       |                  | LDA        | 0.54     | 0.37   |
|            |       |                  | LR         | 0.59     | 0.40   |
|            |       |                  | DAE        | KNN      | 0.61   |
|            |       |                  | SVM        | 0.61     | 0.48   |
|            |       |                  | XGB        | 0.59     | 0.44   |
|            |       |                  | LDA        | 0.52     | 0.37   |
| C<sub>S</sub>(x) | T<sub>1</sub>−T<sub>7</sub>(x), F<sub>1</sub>−F<sub>35</sub>(x), T<sub>F</sub>1−28(x) | SCBSS | KNN | 0.54 |
|            |       |                  | SVM        | 0.68     | 0.57   |
|            |       |                  | XGB        | **0.91** | **0.86** | **0.96** | **0.86** |
|            |       |                  | LDA        | 0.56     | 0.55   |
|            |       |                  | LR         | 0.56     | 0.41   |
|            |       |                  | DAE        | KNN      | 0.56   |
|            |       |                  | SVM        | 0.87     | 0.79   |
|            |       |                  | XGB        | 0.86     | 0.78   |
|            |       |                  | LDA        | 0.71     | 0.59   |
|            |       |                  | LR         | 0.63     | 0.45   |
| Ensemble  |       | SCBSS            | KNN | 0.81 |
|            |       |                  | SVM        | 0.91     | 0.93   |
|            |       |                  | XGB        | **0.93** | **0.89** | **0.96** | **0.90** |
|            |       |                  | LDA        | 0.78     | 0.74   |
|            |       |                  | LR         | 0.79     | 0.82   |
|            |       |                  | DAE        | KNN | 0.75 |
|            |       |                  | SVM | 0.83 |
|            |       |                  | XGB | 0.79 |
|            |       |                  | LDA | 0.72 |
|            |       |                  | LR | 0.80 |

Table 1: A comparison between the efficiencies of individual models and how each impacts our final Ensemble method. Best results are **bolded**. Results reported are Accuracy (Acc), Precision (PR), in addition to clinically-used Sensitivity (SN) and Specificity (SP) scores, with validation performed on a Hold-out set and leave-one-out cross-validation (LOOCV).
9. Chroma \((R(x))\): Majorly used to measure the tonal content of audio pieces \(\text{(Shah et al. 2019)}\).
10. Mel Spectrogram \((M(x))\): As a variant of spectrograms, this technique generates perceptually relevant amplitude/frequency representations.
11. Mel-frequency cepstrum \((MF(x))\): MFCCs have widely been used as audio feature representations of the short-term power spectrum of a sound.
12. Constant-Q Transform \((Q(x))\): CQT simply transforms data from the time domain to the time-frequency domain with large Q-factors \(\text{(Schörkhuber 2010)}\).

Our goal is to improve our understanding of fetal heartbeats by transforming an FPCG signal from the time domain to the frequency and time-frequency domains. This way, a signal’s characteristic gets unveiled in two more domains with which we can better model FGI. Below is a description of how feature extraction is applied in each domain:

- **Time domain (12 features):** Features 1 to 7 \(T_{1-7}(x)\), Chroma mean \(R(i)\), Mel Spectrogram mean \(M(i)\), MFCC mean \(MF(i)\), and CQT mean \(Q(i)\) are extracted.
- **Frequency domain (35 features):** To handle information density, we limit our frequency of interest range down to 0-1500 Hz and break it down to 5 equal intervals of 300 Hz from which we extract features 1 to 7 as \(F_{1-35}(x)\).
- **Time-frequency domain (28 features):** We used the coeif wavelet to apply Discrete Wavelet Transform (DWT) and extracted features 1 to 7 from 3 levels of coefficients to \(TF_{1-28}(x)\) features.

### Fetal Gender Classification

We propose an ensemble method of FGI using different ML algorithms. In an ensemble setting, many individually trained learning algorithms are teamed up to perform at the same classification/regression task that they were initially trained on together. Generally, in an ensemble method, there are three types of approaches that researchers have majorly taken. They have either used the Boosting or the Bagging method to combine homogeneous inefficient models or the Stacking method to increase the efficiency of heterogeneous weak learners. Same as the latter, we stack different individually-trained FGI classifiers and design a Meta-learner to learn from their mistakes. However, unlike common stacking methods where plenty of different classifiers are trained on the exact same feature representations, we train individual models using different feature representations of FPCG signals. These models are as follows:

1. **Statistical Classifier:** We train an XGBoost Classifier \(\text{Meintjes, Lowe, and Legget 2018; Juniati et al. 2018)}\) on the training set \((X)\) to solely learn statistical features of denoised signals as \(C_S[(T_{1-7}(X), F_{1-35}(X), T_{F1-28}(X))]\).

2. **Chroma Classifier:** Among sound features, this classifier learns to classify tonal features as \(C_R(R(x))\) using Linear discriminant analysis (LDA).

3. **Mel Spectrogram Classifier:** We define an XGBoost Classifier as \(C_M(M(x))\) to classify denoised FPCG signals by their perceptual relevancy.

4. **MFCC Classifier:** an SVM is trained to classify power spectrums as \(C_{MF}(MF(x))\).

5. **CQT Classifier:** As our final individual learner, the \(C_Q(Q(x))\) classifier uses the K-Nearest Neighbors (KNN) algorithm to predict the gender using time-frequency information of a signal.

Finally, having trained these models, we stack them to build an ensemble using the XGBoost Classifier as the meta-learner. Passing a denoised FPCG signal to each of the individual models, we obtain predicted class probability distributions which will be concatenated as an input vector to the Meta-learner, after which the meta-learner would learn the correct final classification of fetal gender based on individually-trained models’ class distributions.

### Performance Metrics

To demonstrate our method’s applicability and establish a baseline for our problem, we implemented methodologies that have either taken a classical ML or DL approach. We chose a diverse set of methodologies to compare various effects of preprocessing, feature extraction techniques, and classification methods. We report results using the two conventionally used ML classification metrics of Accuracy (Acc) and Precision (PR), in addition to clinically-used Sensitivity (SN) and Specificity (SP) scores. These metrics are once reported for each approach when it has been validated using a Hold-out set and again using Leave-One-Out Cross-Validation (LOOCV). It is noteworthy that at inference, we did not solely leave one sample out, since this may cause certain training samples belong to the subject (fetus) to which the test sample also belongs. To avoid this, we left all samples that belonged to a certain subject out (i.e. Leave-One-Subject-Out Cross-Validation) to evaluate a model.

### Results & Discussion

As shown in Table 1 to form an ensemble, we compare the performance of different classifiers (i.e. KNN, SVM, XGBoost, LDA, and Logistic Regression (LR)) to choose the top-task-performing one as a final individual FGI model. For instance, discriminant MFCC features can be best identified using an SVM, whereas LR falls short. However, the impact of classifiers is shown to be less in an analogy to the denoising method. When trained on features extracted from the SCBSS-denoised signals, the classifiers are generally superior. The major difference between these neural denoising techniques is the structure of their network. Whereas the DAE is based on Autoencoders, the SCBSS relies on RNN and embeddings. Interestingly, where the first eliminates classification-worthy information through denoising, the latter filters such that models better understand them. By stacking the chosen classifiers, we create an ensemble that performs better than any one model alone.

Improvements made by our method are largely determined by two key factors. In contrast to most previously proposed ML methodologies (Table 2), we let more than just one algorithm be in charge of classification. Studies such as \(\text{Meintjes, Lowe, and Legget 2018; Juniati et al. 2018)}\), who have also handcrafted their features of interest, mostly failed
Table 2: Fetus Gender Identification baseline results compared to our best approach. Results reported are Accuracy (Acc), Precision (PR), in addition to clinically-used Sensitivity (SN) and Specificity (SP) scores, with validation performed on a Hold-out set and leave-one-out cross-validation (LOOCV).

| Year | Author          | Input Features   | Method      | Hold-out | LOOCV |
|------|-----------------|------------------|-------------|----------|-------|
| 2015 | Zheng et al.    | EMD              | SVM         | 0.5      | 0.49  | 0.66  | 0.38 |
| 2016 | Nilanon et al.  | Spectrograms     | 2D-CNN      | 0.65     | 0.49  | 0.52  | 0.48 |
| 2017 | Li et al.       | FFT              | Logistic Regression | 0.58   | 0.50  | 0.52  | 0.49 |
| 2017 | Boulares et al. | Mel Spectrogram  | DNN + TL    | 0.71     | 0.67  | 0.87  | 0.68 |
| 2018 | Meintjes et al. | CWT              | KNN         | 0.60     | 0.55  | 0.57  | 0.57 |
| 2018 | Juniati et al.  | DWT              | KNN         | 0.69     | 0.55  | 0.66  | 0.54 |
| 2019 | Noman et al.    | Signal + MFCC    | CNN Ensemble | 0.58   | 0.50  | 0.57  | 0.49 |
| 2019 | Nogueira et al. | MFCC             | SVM         | 0.65     | 0.54  | 0.71  | 0.66 |
| 2019 | Shahid et al.   |                | AdaBoost    | 0.82     | 0.77  | 0.88  | 0.75 |
| 2020 | Khan et al.     | MFCC             | LSTM        | 0.59     | 0.62  | 0.59  | 0.52 |
| 2020 | Li et al.       | Statistical      | 1D-CNN      | 0.61     | 0.51  | 0.78  | 0.54 |
| 2020 | Alafif et al.   | MFCC             | 2D-CNN + TL | 0.68     | 0.61  | 0.69  | 0.6 |

| 2021 | This work      | Statistical + Sound | Ensemble | 0.93 | 0.89 | 0.96 | 0.90 | 0.91 | 0.86 | 0.96 | 0.86 |

Conclusion

The prenatal identification of fetal gender is a day-to-day clinical task that is mostly carried out using Ultrasounds. In this work, we tackle the problem of fetus gender identification using acoustical sounds of the heart as recorded by PCG signals, and extend upon previous studies in CVD detection to the task presented in this paper. We established a baseline by experimenting with different denoising methods and feature extraction techniques to derive meaningful information from these signals to aid the modeling of fetus gender using classical ML and DL algorithms. Our proposed ensemble method significantly improved the baseline by correctly classifying 91% of the FPCG signals among unseen subjects. Our results motivate further work in robust denoising techniques, and the development of end-to-end systems for modeling heart activity and FGI.

References

[Alafif et al. 2020] Alafif, T.; Boulares, M.; Barnawi, A.; Alafif, T.; Althobaiti, H.; and Alferaidi, A. 2020. Normal and abnormal heart rates recognition using transfer learning. In 2020 12th International Conference on Knowledge and Systems Engineering (KSE), 275–280.
[Avanzato and Beritelli 2020] Avanzato, R., and Beritelli, F.
cation using the continuous wavelet transform and convolutional neural networks. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 409–412.

[Nilan et al. 2016] Nilan, T.; Yao, J.; Hao, J.; Pushrotham, S.; and Liu, Y. 2016. Normal / abnormal heart sound recordings classification using convolutional neural network. In 2016 Computing in Cardiology Conference (CinC), 585–588.

[Ornoy, Weinstein-Fudim, and ergaz shaltiel 2020] Ornoy, A.; Weinstein-Fudim, L.; and ergaz shaltiel, Z. 2020. Methods for Prenatal Sex Determination and Their Importance in Understanding and Prevention of Gender-Related Birth Defects.

[Pham et al. 2018] Pham, D.-H.; Meignen, S.; Dia, N.; Fontecave-Jallon, J.; and Rivet, B. 2018. Phonocardiogram signal denoising based on nonnegative matrix factorization and adaptive contour representation computation. IEEE Signal Processing Letters 25(10):1475–1479.

[Raza et al. 2019] Raza, A.; Mehmood, A.; Ullah, S.; Ahmad, M.; Choi, G. S.; and On, B.-W. 2019. Heartbeat sound signal classification using deep learning. Sensors 19(21).

[Remien and Majmundar 2021] Remien, K., and Majmundar, S. H. 2021. Physiology, fetal circulation.

[Ren et al. 2018] Ren, Z.; Cummins, N.; Pandit, V.; Han, J.; Qian, K.; and Schuller, B. 2018. Learning image-based representations for heart sound classification. Proceedings of the 2018 International Conference on Digital Health.

[Rubin et al. 2016] Rubin, J.; Abreu, R.; Ganguli, A.; Nelaturi, S.; Matei, I.; and Sricharan, K. 2016. Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients. In 2016 Computing in Cardiology Conference (CinC), 813–816.

[Safara et al. 2013] Safara, F.; Doraisamy, S.; Azman, A.; Jantan, A.; and Abdullah Ramaiah, A. R. 2013. Multi-level basis selection of wavelet packet decomposition tree for heart sound classification. Computers in Biology and Medicine 43(10):1407–1414.

[Sainburg, Thielsk, and Gentnert 2020] Sainburg, T.; Thielsk, M.; and Gentnert, T. Q. 2020. Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. PLoS computational biology 16(10):e1008228.

[Samieinasab and Sameni 2015] Samieinasab, M., and Sameni, R. 2015. Fetal phonocardiogram extraction using single channel blind source separation. In 2015 23rd Iranian Conference on Electrical Engineering, 78–83.

[Schrökhuber 2010] Schrökhuber, C. 2010. Constant-q transform toolbox for music processing.

[Shah et al. 2019] Shah, A.; Kattel, M.; Nepal, A.; and Shrestha, D. 2019. Chroma feature extraction.

[Shi et al. 2019] Shi, K.; Schellenberger, S.; Michler, F.; Steigleder, T.; Malessa, A.; Lurz, F.; Oßgathe, C.; Weigel, R.; and Koelpin, A. 2019. Automatic signal quality index determination of radar-recorded heart sound signals using ensemble classification. IEEE Transactions on Biomedical Engineering 67:773–785.

[Simonyan and Zisserman 2015] Simonyan, K., and Zisserman, A. 2015. Very deep convolutional networks for large-scale image recognition.

[Springer et al. 2016] Springer, D.; Springer, D.; Brennan, T.; Nusi, N.; Abdelrahman, H.; Zühlke, L.; Mayosi, B.; Tarassenko, L.; Clifford, G.; and Uhlke, L. 2016. Automated signal quality assessment of mobile phone-recorded heart sound signals automated signal quality assessment of mobile phone-recorded heart sound signals. Journal of Medical Engineering & Technology 40.

[Swaminathan et al. 2020] Swaminathan, A.; Chaba, M.; Sharma, D. K.; and Chaba, Y. 2020. Gender classification using facial embeddings: A novel approach. Procedia Computer Science 167:2634–2642. International Conference on Computational Intelligence and Data Science.

[Tang et al. 2018] Tang, H.; Dai, Z.; Jiang, Y.; Li, T.; and Liu, C. 2018. Pcg classification using multidomain features and svm classifier. BioMed Research International 2018.

[Thiyagaraja et al. 2018] Thiyagaraja, S. R.; Dantu, R.; Shrestha, P. L.; Chitnis, A.; Thompson, M. A.; Anumandla, P. T.; Sarma, T.; and Dantu, S. 2018. A novel heart-mobile interface for detection and classification of heart sounds. Biomedical Signal Processing and Control 45:313–324.

[Tiwari et al. 2021] Tiwari, S.; Jain, A.; Sharma, A. K.; and Mohamad Almustafa, K. 2021. Phonocardiogram signal based multi-class cardiac diagnostic decision support system. IEEE Access 9:110710–110722.

[Toh, Togneri, and Nordholm 2005] Toh, A.; Togneri, R.; and Nordholm, S. 2005. Spectral entropy as speech features for speech recognition. Proceedings of PEECS.

[Trevethan 2017] Trevethan, R. 2017. Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. Frontiers in Public Health 5.

[Tseng et al. 2021] Tseng, K.; Wang, C.; feng Huang, Y.; Chen, G.-R.; Yung, K.; and Ip, W. 2021. Cross-domain transfer learning for pcg diagnosis algorithm. Biosensors 11.

[Vashisth and Meehan 2020] Vashisth, P., and Meehan, K. 2020. Gender classification using twitter text data. In 2020 31st Irish Signals and Systems Conference (ISSC), 1–6.

[Virtanen et al. 2020] Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 17:261–272.

[WHO 2021] WHO. 2021. Cardiovascular diseases (cvds).

[Wu et al. 2019] Wu, J. M.-T.; Tsai, M.-H.; Huang, Y. Z.; Islam, S. H.; Hassan, M. M.; Alelaiwi, A.; and Fortino, G. 2019. Applying an ensemble convolutional neural network
with savitzky–golay filter to construct a phonocardiogram prediction model. Applied Soft Computing 78:29–40.

[Yang et al. 2020a] Yang, L.; Li, S.; Zhang, Z.; and Yang, X. 2020a. Classification of phonocardiogram signals based on envelope optimization model and support vector machine. Journal of Mechanics in Medicine and Biology 20:1950062.

[Yang et al. 2020b] Yang, L.; Li, S.; Zhang, Z.; and Yang, X. 2020b. Classification of phonocardiogram signals based on envelope optimization model and support vector machine. Journal of Mechanics in Medicine and Biology 20(01):1950062.

[Yaseen, Son, and Kwon 2018] Yaseen; Son, G.-Y.; and Kwon, S. 2018. Classification of heart sound signal using multiple features. Applied Sciences 8(12).

[Yildiz 2019] Yildiz, T. 2019. A comparative study of author gender identification. TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES 27:1052–1064.