Non-Contact Atrial Fibrillation Detection From Face Videos by Learning Systolic Peaks

Zhaodong Sun , Juhani Junttila, Mikko Tulppo, Tapio Seppänen , and Xiaobai Li , *Member, IEEE*

Abstract—*Objective:* **We propose a non-contact approach for atrial fibrillation (AF) detection from face videos.** *Methods:* **Face videos, electrocardiography (ECG), and contact photoplethysmography (PPG) from 100 healthy subjects and 100 AF patients are recorded. Data recordings from healthy subjects are all labeled as healthy. Two cardiologists evaluated ECG recordings of patients and labeled each recording as AF, sinus rhythm (SR), or atrial flutter (AFL). We use the 3D convolutional neural network for remote PPG monitoring and propose a novel loss function (Wasserstein distance) to use the timing of systolic peaks from contact PPG as the label for our model training. Then a set of heart rate variability (HRV) features are calculated from the inter-beat intervals, and a support vector machine (SVM) classifier is trained with HRV features.** *Results:* **Our proposed method can accurately extract systolic peaks from face videos for AF detection. The proposed method is trained with subject-independent 10-fold cross-validation with 30 s video clips and tested on two tasks. 1) Classification of healthy versus AF: the accuracy, sensitivity, and specificity are 96.00%, 95.36%, and 96.12%. 2) Classification of SR versus AF: the accuracy, sensitivity, and specificity are 95.23%, 98.53%, and 91.12%. In addition, we also demonstrate the feasibility of non-contact AFL detection.** *Conclusion:* **We achieve good performance of non-contact AF detection by learning systolic peaks.** *Significance:* **non-contact AF detection can be used for self-screening of AF symptoms for suspectable populations at home or self-monitoring of AF recurrence after treatment for chronic patients.**

Index Terms—**Atrial fibrillation (AF), heart rate variability (HRV), photoplethysmography imaging, remote photoplethysmography, wasserstein distance.**

I. INTRODUCTION

ATRIAL fibrillation (AF) is a common heart arrhythmia, and about 2% of the global population is reported to have this disease [1]. AF might cause stroke and heart failure, but most AF episodes are asymptomatic in the early phase. Early detection of AF episodes is essential to avoid these severe diseases. The common method for AF diagnosis is to observe electrocardiography (ECG) signals measured from the electrodes attached to the chest. However, this requires specific medical equipment and complicated operations to measure ECG signals. Cardiologists should determine the diagnosis from ECG signals, which is not practical for daily checks or long-term monitoring. Although some works [2]–[5], [5]–[12] designed automatic AF detection algorithms from ECG and achieve a high accuracy, the acquisition of ECG signals is complicated and not convenient for daily use. Some works [13]–[21] propose to use contact photoplethysmography (PPG) from pulse oximeters, smartwatches or phone cameras on fingers for more convenient and accessible AF detection. However, all of the ECG and contact PPG methods require the sensors attached to the body/skin, which might cause irritation and hygiene issues. In addition, specific biomedical equipment such as ECG sensors or pulse oximeters also decrease the accessibility of these methods to more populations.

Compared with ECG and contact PPG, remote PPG[1] provides a non-contact way to monitor cardiac signals by using a camera. Remote PPG is measured from the face color change induced by the blood volume change in face videos [22], [25], and can reveal cardiac rhythms for AF detection. As cameras are ubiquitous nowadays, this might lead to potential solutions or products for convenient non-contact AF detection. There are several situations where AF detection from face videos is preferred. First, AF detection from face videos can be used for telemedicine during video conferences without specific medical instruments. Second, contact AF detection is not applicable to people with skin burns or other skin diseases on the measurement locations such as fingers or chest. It has been reported that 486,000 people in the United States were burned to receive medical treatment in the year 2011 [26]. Third, cameras are ubiquitous and accessible such as in smartphones or laptops, and much cheaper than the smartwatches with ECG/PPG sensors. Finally, as people spend a long time on using smart phones or laptops every day, the AF detection from face videos can be integrated to smart phones or laptops when people are using these devices.

Recent studies [24], [27]–[31] have made some contributions to non-contact AF detection by remote PPG. Most studies

Zhaodong Sun, Tapio Seppänen, and Xiaobai Li are with the Center for Machine Vision and Signal Analysis, University of Oulu, FI 90014 Oulu, Finland (e-mail: zhaodong.sun@oulu.fi; tapio.seppanen@oulu.fi; xiaobai.li@oulu.fi).

Juhani Junttila and Mikko Tulppo are with the Research Unit of Internal Medicine, Medical Research Center Oulu, FI 90014 Oulu, Finland (e-mail: juhani.junttila@oulu.fi; mikko.tulppo@oulu.fi).

---

[1]Remote PPG is used to be literally distinguished from contact PPG. There are other terms equivalent to remote PPG such as photoplethysmography imaging [22], video photoplethysmography [23], and videoplethysmography [24].

focus on heart rate variability (HRV) features derived from systolic peaks in remote PPG for AF detection and achieved promising results. The problem is that remote PPG signals are subtle compared to noises, and systolic peaks of AF patients have low amplitudes when the heart rate is high [21]. Therefore, it is challenging to develop a method that can measure systolic peaks accurately enough for AF detection. Therefore, we develop a method to accurately measure systolic peaks from face videos to facilitate non-contact AF detection and test the effectiveness of the method on our large-scale dataset.

Our main contributions are listed below:

- We develop a deep learning-based remote PPG algorithm by using the timing of the systolic peaks as the ground truth and demonstrate that learning systolic peaks from face videos can facilitate non-contact AF detection.
- We record the full version of Oulu Bio-face (OBF) dataset [27] for non-contact AF detection, including videos from 100 healthy subjects and 100 AF patients. Two cardiologists evaluated ECG recordings of patients and labeled each recording as AF, sinus rhythm (SR), or atrial flutter (AFL).
- We test two types of AF detection models for the self-screening of healthy subjects and patients, respectively, and achieve high accuracy. The classification between healthy subjects and patients with AF has an accuracy of 96.00%. The classification between patients with SR and patients with AF has an accuracy of 95.23%. Additional experiments also demonstrate the feasibility of non-contact AFL detection.

II. RELATED WORKS

A. Contact Methods for AF Detection

The AF detection from ECG can be based on 1) P-wave detection [2]–[5], 2) R-R interval variability [5]–[9], and 3) deep learning (DL)-based methods [10]–[12]. P waves are not prominent features and are often influenced by artifacts, while R peaks have higher amplitudes that can be easily detected. AF patients have a larger R-R interval variability, which means the cardiac rhythm of AF patients shows larger irregularity, so the features from R-R intervals can be a good indicator for AF detection. Some traditional methods [6]–[8] extracted features such as entropy [8] or HRV features [7] from R-R intervals and simply use a threshold to do AF classification. Some methods [9], [32], [33] applied machine learning methods such as support vector machine (SVM) [32], convolutional neural network [33] or neighborhood component analysis [9] to R-R intervals for AF detection and achieved higher performance than traditional methods. However, it is difficult for rhythm-based methods using R-R intervals to distinguish AF from other arrhythmias, while the morphology-based methods such as P-wave absence can avoid false-positive errors in AF detection [34]. Deep learning-based methods [10]–[12] can be directly applied to ECG signals and achieve good AF detection performance with large datasets, but they are sensitive to changes in ECG morphology and are unclear how they generalize to unseen data [34].

Although ECG is a gold standard for AF diagnosis, some studies found contact PPG [13]–[21] measured from fingertips and ballistocardiograph (BCG) [35]–[37] can also be used for AF detection. PPG and BCG signals do not have the P-wave in ECG, so the systolic peaks in PPG and BCG are the only clue for AF detection. Similar to the R-peaks in ECG, the systolic peaks are also an indicator of cardiac rhythm and the peak-peak intervals can also be used to derive the HRV features. For the daily use and better accessibility, some work proposed to use the accelerometer and gyroscope of a smartphone on the chest to get the BCG signal for AF detection [10], while others proposed to use the camera of a smartphone [6] on the fingertips or smart watches [18]–[21] to get the contact PPG signals for AF detection. The AF detection from contact PPG and BCG provides other alternatives to ECG signals and is more convenient for daily health monitor.

B. Remote PPG for non-Contact AF Detection

Recent studies [24], [29]–[31] showed that remote PPG from face videos can be used for non-contact AF detection. Couderc et al. [24] first used HRV features from remote PPG for AF detection on a small dataset of 11 patients, which demonstrates the feasibility of AF detection from face videos. Yan et al. [30] collected a dataset including 217 patients among which 75 patients showed AF. They used a pre-trained SVM with features from autocorrelation analysis and achieved 95.4% accuracy for classifying AF patients vs. non-AF patients. Shi et al. [29] proposed a HRV feature fusion approach by combining three remote PPG algorithms [27], [38], [39] for AF detection. The method was tested on the previous version of OBF dataset [27], including 30 AF patients and 100 healthy subjects, and achieved an accuracy of 92.56%. Corino et al. [31] used a simple model with three features to do AF detection from remote PPG on 68 patients and achieved 87% AF detection accuracy.

There are one or more of the following limitations in the previous works. 1) Their AF detection accuracy is limited, as the systolic peak information was not exploited for training remote PPG algorithms. We propose to utilize systolic peaks to train our remote PPG algorithm, and the reasons are in two aspects. First, studies [15], [16], [24], [27]–[29] showed that HRV features computed using systolic peaks are sufficient for AF detection, so the waveform is not needed. Second, the systolic peaks are the most prominent and reliable feature in remote PPG, while other waveform features are subtle and easily contaminated by noises. 2) Most previous studies were tested on limited AF cases, and a larger scale of AF data is needed to make a reliable evaluation of the methods. 3) The low diversity of previous datasets also limits their application scope. Previous studies mainly focused on classifying healthy subjects versus before-treatment AF patients. It is also essential to compare before-treatment (AF) and after-treatment (sinus rhythm, SR) of the same patients, which was not concerned in any previous study yet. 4) Previous studies (except [27]) did not compare the non-contact AF detection performance with results from ECG or contact PPG. The AF detection results from ECG or contact
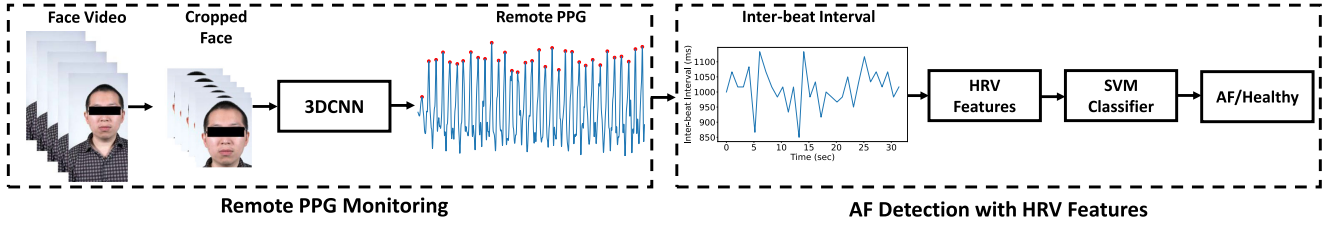
Fig. 1. The general framework of non-contact AF detection. The red dots in the remote PPG are the systolic peaks. 3DCNN = three dimensional convolutional neural network, PPG = photoplethysmography, AF = atrial fibrillation, HRV = heart rate variability, SVM = support vector machine.

PPG can be a performance upper bound from which we can see the gap between non-contact and contact AF detection.

III. METHOD

In this section, we will introduce the 3-dimensional convolutional neural network (3DCNN) model for remote PPG monitoring and how to select a loss function to train our model with the timing of systolic peaks to facilitate AF detection. We will also explain using the HRV features derived from systolic peaks for AF detection.

A. Remote PPG Monitoring

1) Preprocessing: We first need to crop the face region from the face video. The face region is obtained from the landmarks generated from OpenFace [40]. We first get the minimum and maximum horizontal and vertical coordinates of the landmarks to locate the central face point. The bounding box size is 1.2 times the range of vertical coordinates of landmarks and is fixed for each video. After getting the central point and the size of the bounding box, we can crop the face region from each frame as shown in Fig. 1. The cropped face is resized to 128×128.

2) 3DCNN for Remote PPG Monitoring: The cropped face is fed into the 3DCNN model [28] for remote PPG monitoring. 3DCNN uses the 3D kernels to do the convolution on the video along the width, height, and time axis. We will use the similar 3DCNN architecture as [28] with one modification that will be illustrated in the next part. The model is shown in Fig. 2. The input for 3DCNN is the cropped face video clip $x \in \mathbb{R}^{C \times H \times W \times T}$ where C is the number of color channels, H is the video height, W is the video width, and T is the video time length. The output remote PPG is $p_r = G_\theta(x) \in \mathbb{R}^T$. The input video is a 4D video signal, while the network eliminates the height, width, and color dimensions and converts the 4D signal into a 1D remote PPG signal. To train this network, we need to minimize a loss function. The loss function measures the distance between the ground truth, which is the contact PPG $p_c(t)$, and the network output, which is the remote PPG $p_r(t)$. A previous method [28] used negative Pearson correlation as the loss function with the contact PPG as the ground truth. This loss function encourages the remote PPG signals and contact PPG signals to have similar morphology. However, this loss function also encourages the model to learn redundant information from the contact PPG, such as diastolic peaks or artifacts.

3) Binary Systolic Peaks for 3DCNN Training: Using contact PPG as the ground truth to train the network is not the best.
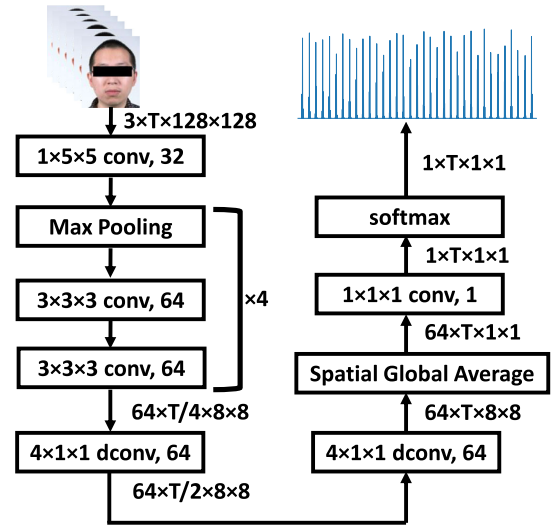

Fig. 2. Three-dimensional convolutional neural network (3DCNN) architecture. The input is the image sequence of the cropped face. The output is the systolic peak signal. "$3 \times 3 \times 3$ Conv, 64" means the 3D convolution operation with filter size 3x3x3 and the output channel is 64. "dconv" is 3D transposed convolution, which works as upsampling. There is also a batchnorm layer and a ReLU activation following each convolutional block. There are 4 max pooling layers. All of them downsample the spatial dimensions, and two of them only downsample the temporal dimension.

The systolic peak timing in the contact PPG is a better ground truth. There are two reasons. First, [41] demonstrated that the waveform morphology of contact PPG signals is not entirely consistent with remote PPG due to their different measurement locations. This means that contact PPG is not the actual ground truth for each face video due to the inconsistency of contact and remote PPG. However, [41] also showed that the timing of systolic peaks is the consistent information between contact and remote PPG. Second, contact PPG from patients usually contain systolic peaks with much lower amplitude than that from healthy subjects, as shown in Fig. 3. Väliaho *et al.* [21] also reported this phenomenon and explained that a higher heart rate in AF causes smaller stroke blood volume and lower systolic peak amplitudes. If we use the contact PPG as the ground truth for model training, the model might learn some redundant information and ignore some systolic peaks with low amplitude. Therefore, we will convert our contact PPG into binary systolic peaks as shown in Fig. 3. The binary systolic peaks only have zeros and ones, which means it only keeps the timing of the systolic peaks and removes other redundant information. This kind of ground truth

(a) Contact PPG from a healthy subject

(b) Binary systolic peaks from a healthy subject

(c) Contact PPG from a patient with AF

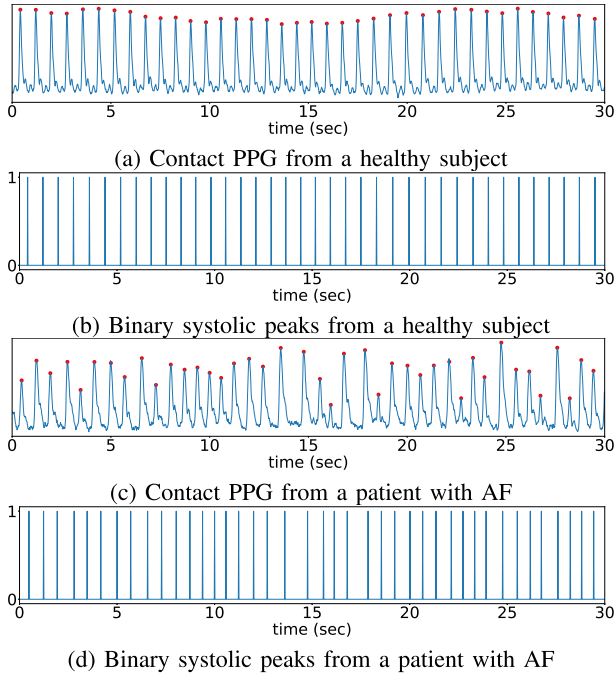(d) Binary systolic peaks from a patient with AF

Fig. 3. Contact PPG (with systolic peaks marked with red dots) and corresponding binary systolic peaks. Binary systolic peaks (b and d) are extracted from the contact PPG signals (a and c). The PPG signal from the AF patient contains complex patterns such as irregular heartbeat rhythms and different systolic peak amplitudes. PPG = photoplethysmography, AF = atrial fibrillation.

can encourage the model only to learn the timing of systolic peaks, which is also the only information used for AF detection.

The next question is how to use the binary systolic peaks for our model training. We can regard our model output $p_r = G_\theta(x)$ and our binary systolic peak label s as two probability distributions so that we can use some probability distribution distances as our loss function. The reason is that probability distribution distances can compare a binary pulse signal with a continuous smooth signal. Our systolic peak signal is just a series of binary signals, and our model output is the continuous smooth signal, so probability distribution distances might be a good choice. Previous used negative Pearson correlation loss function can handle continuous values in contact PPG but cannot work with binary values in binary systolic peak signals.

To satisfy the requirement of probability distributions, we should ensure that the sum of a signal is one and the signal is non-negative. We can normalize the binary systolic peak as $\bar{s} = s/\Sigma_i s_i$. Since the softmax layer can make sure the model output is non-negative and the sum is one, we add a softmax layer to the output of the network $G_\theta(x)$ and the new network is $H_\theta(x) = \mathrm{softmax}(G_\theta(x))$. Therefore, we have our new model output $\bar{p}_r = H_\theta(x)$. Our new loss function $l(\bar{s}, \bar{p}_r)$ is defined to measure the distance between normalized systolic peak signal $\bar{s}$ and model output $\bar{p}_r$. We use Wasserstein distance as our loss function l for learning the systolic peaks. The reason why Wasserstein distance is used as the loss function and the full analysis about the loss function selection is described in the following.

B. Loss Function Selection for Learning Systolic Peaks

1) Candidate Loss Functions: For the proposed 3DCNN model, both the model output signal $\bar{p}_r$ and the normalized binary systolic peaks $\bar{s}$ can be regarded as probability distributions. Therefore, we can use some probability distribution distances to measure their similarity. These probability distribution distances can be used as loss functions for the model training. During training, the loss function is minimized so that our model output signal is well aligned with the systolic peak signals. In previous work [42], probability distribution distances were used for comparing two time series in the power spectral density (PSD) domain. However, the accurate representation of systolic peaks can only be in the time domain. Therefore, we directly use the probability distribution distances in the time domain. There are some options for probability distribution distances, such as squared Euclidean distance, Kullback-Leibler (KL) divergence, Jensen-Shannon (JS) divergence, and Wasserstein distance.

Squared Euclidean Distance (SED): Squared Euclidean distance is a straightforward way to compare two time series. It measures the Euclidean distance between two probability distributions. It is also widely used as a loss function for deep learning models. It is defined as

$$l_{SED}(\bar{p}_r, \bar{s}) = \parallel \bar{p}_r - \bar{s} \parallel_2^2 \tag{1}$$

Kullback-Leibler (KL) Divergence: The KL divergence is another probability distribution distance widely used as the loss function in classification tasks. In the classification tasks, it measures the distance between the true classification distribution and the predicted classification distribution. KL divergence is defined as

$$l_{KL}(\bar{p}_r, \bar{s}) = \mathrm{KL}(\bar{s} \parallel \bar{p}_r) = \sum_{t=1}^{T} \bar{s}(t) \log\left(\frac{\bar{s}(t)}{\bar{p}_r(t)}\right) \tag{2}$$

Jensen-Shannon (JS) Divergence: JS divergence is based on KL divergence. Compared with KL divergence, JS divergence is symmetric and bounded between 0 and 1. It is defined as

$$l_{JS}(\bar{p}_r, \bar{s}) = \frac{1}{2}\mathrm{KL}\left(\bar{p}_r \parallel \frac{\bar{p}_r + \bar{s}}{2}\right) + \frac{1}{2}\mathrm{KL}\left(\bar{s} \parallel \frac{\bar{p}_r + \bar{s}}{2}\right) \tag{3}$$

Wasserstein Distance: Wasserstein distance is based on optimal transport. It finds the minimum cost to move the mass of one probability distribution to turn the probability distribution into another. It is defined as

$$l_{WS}(\bar{p}_r, \bar{s}) = \min_{\pi \in \Pi(\bar{p}_r, \bar{s})} \mathbb{E}_{(x,y)\sim\pi}[\parallel x - y \parallel] \tag{4}$$

where $\Pi(\bar{p}_r, \bar{s})$ is a set containing all joint probability distributions with marginal distribution $\bar{p}_r$ and $\bar{s}$. Wasserstein distance finds the joint probability distribution (transport plan) π so that the movement cost is minimum. x and y are two random variables following the joint distribution π. However, the formula above has a min operation and cannot be directly used as a loss function. Since our signals $\bar{p}_r$ and $\bar{s}$ are one-dimensional, the Wasserstein distance has a closed-form in the one-dimensional
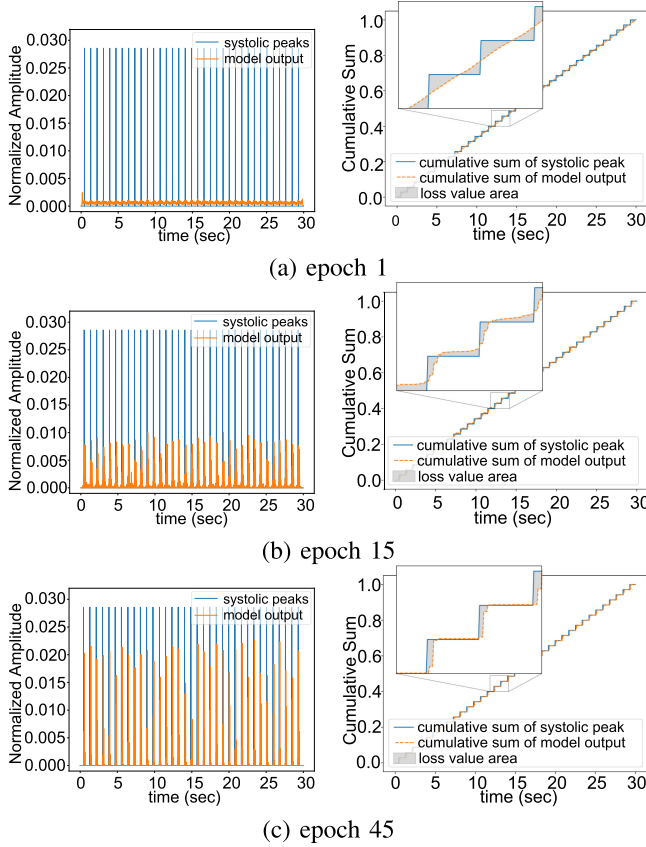
(a) epoch 1

(b) epoch 15

(c) epoch 45

Fig. 4. Illustration of how Wasserstein distance works. The first, second, and third rows are at epochs 1, 15, and 45 of model training. The first column shows the ground truth systolic peaks and our model output. The second column shows the cumulative sum of the two signals in the first column and also shows the zoom-in figures. The gray area is the Wasserstein distance value. During model training, we intend to minimize this gray area so that the model output is similar to the systolic peak signal. The decreasing of the area makes the peak signal from the model sharper and more accurately aligned with the true systolic peaks.

case, as shown below.

$$l_{WS}(\bar{p}_r, \bar{s}) = \int_0^1 |F_{\bar{p}_r}^-(u) - F_{\bar{s}}^-(u)|du = \sum_{t=1}^T |F_{\bar{p}_r}(t) - F_{\bar{s}}(t)|$$
(5)

where $F_{\bar{p}_r}$ is the cumulative sum of $\bar{p}_r$, which is also the cumulative distribution function. $F_{\bar{p}_r}^-$ is the inverse of function $F_{\bar{p}_r}$. The Wasserstein distance used for the model training is the rightmost part of (5). Fig. 4 shows the interpretation of the loss function. To calculate the value of the Wasserstein distance between two signals, we first get the cumulative sum of each signal. From (5), the Wasserstein distance is the absolute difference between the two cumulative sum curves, as the gray area shown in the third column of Fig. 4.

2) Analysis of Peak Misalignment: We first analyze the loss function response when two comparing signals are misaligned. We first generate a smooth peak signal p_s from a truncated Gaussian distribution curve $\mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 0.1$ to imitate the 3DCNN model output and normalize it to make sure the sum is one. We also generate a binary peak signal p_b to imitate the ground truth systolic peak. These two signals are shown in
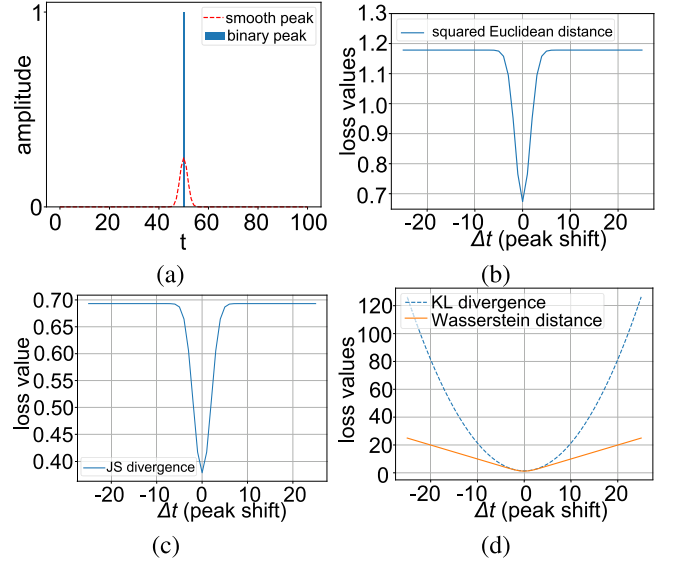

Fig. 5. (a) The smooth peak and binary peak. (b) Squared Euclidean Distance with respect to the peak shift Δt. (c) JS divergence with respect to the peak shift Δt. (d) KL divergence and Wasserstein distance with respect to the peak shift Δt.

Fig. 5(a). We shift the binary peak signal $p_b(t)$ to $p_{b,\Delta t}(t) = p_b(t - \Delta t)$ and get the loss function values $l(p_s, p_{b,\Delta t})$. We can plot the loss function values $l(p_s, p_{b,\Delta t})$ with respect to peak shift Δt to see the loss functions response to the peak misalignment. The results are shown in Fig. 5(b–d). For squared Euclidean distance and JS divergence in Fig. 5(b–c), the loss value becomes saturated and constant when the absolute peak shift is too large, which means the loss function cannot assign a larger penalty to a larger peak alignment. This disadvantage might prevent the model from learning the accurate systolic peaks at these saturated locations. On the other side, KL divergence and Wasserstein distance in Fig. 5(d) assign a larger penalty when the peak misalignment is larger and will not be saturated when the peak shift is large, which means KL divergence and Wasserstein distance could be the promising options.

3) Analysis of Peak Sharpness: We also analyze the loss function response to peak sharpness. We hope the 3DCNN model will output signals with sharper peaks rather than flat ones, as sharper peaks are less impacted by noises and easier to detect. We use the binary peak and a series of smooth peak signals with different variances σ^2 to analyze the response of the loss function to the peak sharpness. We can change the variance σ^2 of the truncated Gaussian distribution curve $\mathcal{N}(0, \sigma^2)$ to control the sharpness of the smooth peak p_{s,σ^2}. Smaller variance σ^2 means sharper peaks. This phenomenon is illustrated in Fig. 6(a). The smooth peak with $\sigma^2 = 10$ is almost flat, while the smooth peak with $\sigma^2 = 0.01$ is very sharp and similar to the binary peak. By changing the variance σ^2 of the smooth peak, we can plot the loss values $l(p_{s,\sigma^2}, p_b)$ with respect to the variance σ^2 of the smooth peak in Fig. 6(b). From the plot, Wasserstein distance has larger penalty values than other loss functions, especially at large variance positions (small sharpness), which indicates that Wasserstein distance can assign a much larger
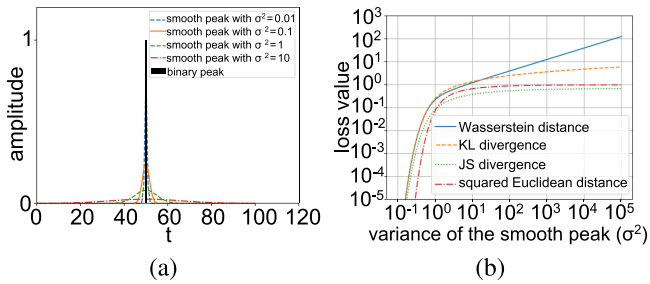
Fig. 6. (a) The binary peak and the smooth peak with difference variances (sharpness). (b) The loss values of the four candidate loss function with respect to the variance σ^2 of the smooth peak.

penalty value when the output peaks are too flat. Therefore, Wasserstein distance is the best to motivate the model to produce sharp peaks.

C. AF Detection With HRV Features

The AF detection part is shown in the right part of Fig. 1. We select 20 features as listed below according to previous studies [15], [16], [24], [27], [29] to train a support vector machine (SVM) classifier with radial basis function (RBF) kernel.

- Time domain: mean inter-beat interval (IBI), standard deviation of IBI (SDNN), standard deviation of successive difference of IBI (SDSD), percentage of samples with more than 50 ms difference from the consecutive beat (pNN50), percentage of samples with more than 20 ms difference from the consecutive beat (pNN20), the number of samples with more than 50 ms difference from the consecutive beat (NN50), the number of samples with more than 20 ms difference from the consecutive beat (NN20), the root mean square of successive differences of IBI (RMSSD), median of IBI, range of IBI, the coefficient of variation of successive differences (CVSD), the coefficient of variation (CVNNI), maximum heart rate, minimum heart rate, and the standard deviation of heart rate.
- Spectral domain: the power in the low frequency (LF) (0.04Hz-0.15 Hz), the power in the high frequency (HF) (0.15Hz-0.4 Hz), and the ratio of LF and HF.
- Geometrical domain: Poincaré plot standard deviations (SD1, SD2).

IV. EXPERIMENTAL SETUP

In this section, we first present our dataset. We will provide the experimental protocol and evaluation metrics for two experiments. One experiment is remote PPG monitoring, and another is non-contact AF/AFL detection.

A. Dataset

1) Participants: We record the full version of the Oulu Bioface (OBF) dataset with 100 healthy subjects and 100 AF patients. The healthy subjects were recruited from the University of Oulu, and AF patients were from Oulu University Hospital. The study was performed according to the Declaration of Helsinki,

TABLE I
STATISTICAL INFORMATION OF THE PARTICIPANTS

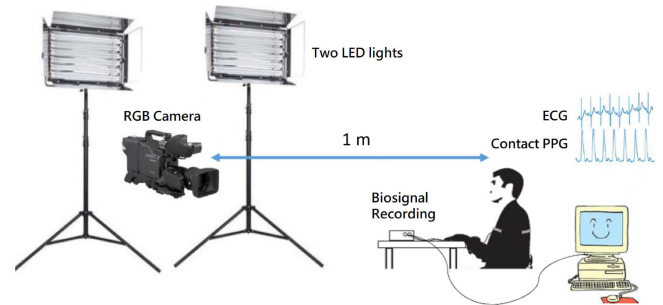|	OBF-H (n=100)	OBF-P (n=100)
Age (year)	31.6 ± 8.8, [18, 68]	64.2 ± 9.2, [43, 88]
Gender	61% Male 39% Female	83% Male 17% Female
Ethnic	Caucasian: 32%, Asian: 37%, Others: 31%.	Caucasian: 100%
Weight (kg)	71 ± 16	96 ± 21
Ratio of wearing eyeglasses	39%	45%

Fig. 7. Data Recording Setup. RGB = red green blue, ECG = electrocardiography, PPG = photoplethysmography, LED = light-emitting diode.

TABLE II
RECORDING EQUIPMENT AND SETTINGS

Devices	Speicifications	Settings	Outputs
RGB camera	Blackmagic URFA mini	ISO 400, FPS 60, Resolution 1920 x 1080	RGB video
ECG sensor	NX-EXG2B	Sampling rate 256 Hz	ECG signal
Contact PPG sensor	NX-BVP1C	Sampling rate 128Hz	Contact PPG signal
LED lights	Aputure, LightStorm LS 1c	Brightness: 3500 lux, Temperature: 5500 k	N/A

RGB = red green blue, ECG = electrocardiography, PPG = photoplethysmography, LED = light-emitting diode.

and the local committee of research ethics of the Northern Ostrobothnia Hospital District approved the protocol (reference number: 116/2016). All participants gave their written consent before recording. The statistical information of participants is summarized in Table I. The partitions of healthy subjects and patients in the OBF dataset are OBF-H and OBF-P, respectively.

2) Data Acquisition: The recording setup is shown in Fig. 7. A participant was seated in front of an RGB camera at a one-meter distance. Two LED lights on both sides of the RGB camera face toward the participant at a 45-degree angle at a 1.5-meter distance. Meanwhile, ECG and contact PPG signals are measured by ECG and PPG sensors, respectively, and synchronized with the video. The recording equipment and settings are shown in Table II.

Modalities	Dataset Partitions	Participants	Sessions	Labels
RGB Videos, ECG, Contact PPG	OBF-H	Healthy Subjects	Before Exercise (100)	Healthy (200)
			After Exercise (100)	
	OBF-P	AF Patients	Before Cardioversion (90)	AF (73) SR (61)
			After Cardioversion (79)	AFL (11) Other (24)

RGB = red green blue, ECG = electrocardiography, PPG = photoplethysmography, AF = atrial fibrillation, SR = sinus rhythm, AFL = atrial flutter.

During the recording, the participant was seated facing the camera, and the ECG/contact PPG sensors were attached to them. For each healthy subject, there are two recording sessions, and each lasts for five minutes. One session was recorded at a resting status, and the other was recorded after 5 minutes of exercise (climbing the stairs) so that a wider range of heart rates is covered. The mean heart rates before and after exercise are 73.6 bpm and 83.2 bpm, respectively. The heart rate ranges before and after exercise are 50.09-107.2 bpm and 60.64-130.1 bpm, respectively. For each AF patient, there are also two recording sessions. One was recorded before cardioversion treatment (AF symptom presented), and the other was recorded after the treatment (back to sinus rhythm) so that each patient has both AF and SR data. The mean heart rates before and after treatment are 78.9 bpm and 66.5 bpm, respectively. The corresponding ECG and contact PPG were recorded simultaneously with each video. Finally, OBF-H has 200 videos. Due to data loss, some patients only have one session, and the total number of videos in OBF-P is 169.

3) Data Labelling: The dataset composition and labels are shown in Table III. In order to do the AF detection, all the videos in OBF-H are labeled as healthy. All patients' data in OBF-P were labeled by two independent cardiologists by observing ECG signals. Only persistent AF/SR/AFL (i.e., no rhythm transition during the recording) cases were labelled as 'AF/SR/AFL', and complex cases (e.g., paroxysmal, ventricular extrasystole, supraventricular extrasystoles) were labelled as 'Other' (24 cases). Finally, there are 73 videos labeled as AF, 61 videos labeled as SR, 11 videos labeled as AFL, and 24 videos labeled as other that are not used for classification experiments.

B. Experiment I: Remote PPG Monitoring

1) Experimental Protocol: We train our models with three dataset partitions: OBF-P, OBF-H, or the whole OBF dataset. We also test our models on these three dataset partitions. The model is trained for 45 epochs with a learning rate of 0.0001. Each training video clip has 512 frames, and the batch size for training is 4. The binary systolic peaks are obtained by NeuroKit2 [43] from our contact PPG and will be used as the

ground truth to train the 3DCNN model. Wasserstein distance is used as the loss function. For testing, the videos are divided into non-overlapping 30-seconds clips and used as inputs. We use subject independent 10-folds cross-validation in all experiments except two cases, i.e., training on OBF-H and testing on OBF-P, and training on OBF-P and testing on OBF-H, which use cross-set protocol. We also compare our method (3DCNN-PEAK) with four remote PPG methods: 3DCNN-BVP [28] is a deep learning-based method using contact PPG waveforms for training. POS [38], CHROM [44], and PBV [45] are classical remote PPG algorithms without training.

2) Evaluation Metrics: We evaluate accuracy on two levels, i.e., the average heart rate level and IBI level. Accurate measure on the IBI level is more challenging and essential for AF detection since it requires accurate systolic peaks.

We use mean absolute error (MAE), root mean squared error (RMSE), and Pearson correlation (R) to evaluate the error of the average heart rate. MAE for heart rate is defined as $\mathrm{MAE}_{HR} = \sum_{n=1}^{N} |\mathrm{HR}_{video} - \mathrm{HR}_{true}|/N$, where N is the number of samples, HR_{video} is the heart rate measured from face videos, and HR_{true} is the true heart rate obtained from the contact PPG. RMSE is defined as $\mathrm{RMSE}_{HR} = \sqrt{\sum_{n=1}^{N}(\mathrm{HR}_{video} - \mathrm{HR}_{true})^2/N}$. Small MAE and RMSE values indicate accurate heart rate estimation. Pearson correlation is the linear correlation between the heart rates measured from videos and the true heart rates. When the Pearson correlation is close to 1, the heart rate estimation is accurate.

We use three metrics of mean absolute error (MAE), standard deviation (STD), and accuracy as used in [46] to evaluate the error of IBI. We can first define the absolute error of IBI as $\mathrm{AE} = \sum_{t=1}^{K} |\mathrm{IBI}_{video}(k) - \mathrm{IBI}_{true}(k)|/K$ where K is the length of the IBI, IBI_{video} is the IBI curve from face video, and IBI_{true} is the true IBI curve from contact PPG. Since the original IBI curve is an irregularly spaced time series, we should resample the original IBI curve to get an evenly spaced time series to calculate the absolute error between two IBI curves. The MAE and STD of IBI are the mean and standard deviation of AE for all samples. The accuracy of IBI is defined as $\mathrm{AC}_{IBI} = 1 - \frac{1}{N}\sum_{n=1}^{N} \mathrm{AE}_n/(T/(B_n - 1))$ where T is the time length of a IBI curve and B_n is the number of systolic peaks in the nth sample.

C. Experiment II: AF/AFL Detection

1) Experimental Protocol: Remote PPG is obtained from the model trained on both OBF-P and OBF-H. HRV features (as described in Section III-C) are calculated from the systolic peaks in remote PPG and used for AF detection experiments. We perform two kinds of AF detection. 1) We use videos from healthy subjects and videos with AF labels from patients to classify healthy vs. AF. 2) We use the patient videos with SR labels and patient videos with AF labels to classify SR vs. AF. We perform the classification with different clip lengths of 10 s, 20 s, 30 s, 60 s, and 120 s to see how the clip length influences the AF detection. 3) We also perform the classification of SR vs. AFL with clip length 30 s to investigate whether AFL can be detected from face videos.

In the classification experiments of healthy vs. AF and SR vs. AF, we use subject-independent 10-fold cross-validation.[2] In the classification experiment of SR vs. AFL, the numbers of two labels are unbalanced, and the number of samples is limited, so we use another evaluation protocol in [29]. We randomly select 6 SR patients and 6 AFL patients to form the training set while randomly selecting another 5 SR patients and the left 5 AFL patients as the testing set. Thus, subjects in the training and testing sets are balanced and not overlapped. The experiments are independently performed ten times, and the average performance is reported.

We also report results achieved by using the ECG signals and contact PPG signals with the same HRV features in Section III-C, which are the ECG/contact PPG reference methods. Theoretically, the ECG and contact PPG results should be the upper bound that the model can achieve and are the reference results compared with remote PPG results. We also report results of ECG-based AF detection methods [9], [11] for healthy vs. AF and SR vs. AF. For remote PPG, we compare our results with four previous methods [28], [29], [38], [45]. For the baseline methods [28], [38], [45] and ECG/contact PPG reference methods, we use the same SVM classifier and HRV features from Section III-C, while only change the input IBI series. For [9], [11], [29], we use the features and models in their papers.

2) Evaluation Metrics: AF detection results are summarized into true positive (TP), true negative (TN), false positive (FP), and false negative (FN). We use accuracy, sensitivity, and specificity as the classification metrics. Accuracy is defined as $(TP + TN)/(TP + TN + FP + FN)$. Sensitivity is defined as $TP/(TP + FN)$. Specificity is defined as $TN/(TN + FP)$. We also add metrics of F1 score and area under curve (AUC) for more convincing comparison.

V. RESULTS AND DISCUSSION

In this section, we will present the results of our two experiments. One experiment is remote PPG monitoring, and another is non-contact AF/AFL detection. Results about computational speed are also reported. Finally, extensive discussion about the results is provided.

A. Results for Experiment I: Remote PPG Monitoring

We compare these four candidate loss functions mentioned in Section III-B1. Table IV shows that Wasserstein distance has the best performance for remote PPG monitoring. Fig. 4 also shows how Wasserstein distance boosts the predicted systolic peaks. Therefore, our experimental results validate our analysis conclusion in Section III-B that Wasserstein distance is the best option.

Table V shows the test results on OBF-P. No matter what training set is used, our proposed method (3DCNN-PEAK)

TABLE IV
PERFORMANCE COMPARISON OF THE CANDIDATE LOSS FUNCTIONS

	Heart Rate			Inter-beat Interval		
	MAE (bpm)	RMSE (bpm)	R	MAE (ms)	STD (ms)	Accuracy (%)
KL	8.77	16.45	0.51	189.44	435.86	83.84
JS	14.80	23.59	0.30	390.62	705.08	75.83
SED	7.18	14.39	0.58	159.77	437.77	86.07
WS	**1.46**	**3.53**	**0.96**	**50.74**	**68.06**	**94.18**

KL = Kullback-Leibler Divergence, JS = Jensen-Shannon Divergence, SED = Squared Euclidean Distance, WS = Wasserstein Distance, MAE = mean absolute error, RMSE = root mean squared error, R = Pearson correlation, STD = standard deviation.

achieves the best performance on heart rate and IBI measurement. Using both OBF-P and OBF-H for training is preferable since this can further improve the model performance compared with using OBF-P or OBF-H alone. The cross dataset test shown in the OBF-H row indicates that our 3DCNN-PEAK method has better generalization capability than the 3DCNN-BVP method. Table VI shows the test results on OBF-H. Our 3DCNN-PEAK method still outperforms other baseline methods. Using both OBF-P and OBF-H for training also achieves the best performance on OBF-H. The reason is that a larger dataset (OBF-P + OBF-H) can include more face videos with different noise types and facial skin types from both healthy subjects and patients, which can train the model to be more robust to noise and better capture subtle remote PPG. It is obvious that the error of heart rate and IBI on OBF-H in Table VI is much lower than on OBF-P in Table V for all methods, which means the remote PPG monitoring for patients is more difficult than for healthy subjects. We provide the final results of training and testing on the whole OBF dataset in Table VII, and our proposed 3DCNN-PEAK achieves the best results where MAE values of heart rate and IBI are 1.4658 BPM and 50.7481 ms, respectively.

B. Results for Experiment II: AF/AFL Detection

Fig. 8 shows the accuracy, sensitivity, and specificity with clip lengths of 10 s, 20 s, 30 s, 60 s, and 120 s for both healthy vs. AF and SR vs. AF. In general, ECG has the best performance among all, and accuracy is high even with the shortest 10 s clip length. Our proposed 3DCNN-PEAK outperforms other remote PPG methods at most clip lengths. We can observe that longer clips can provide better performance. However, longer recording time also decreases the convenience and application scopes of our method. Therefore, we need to balance the recording time and classification performance. It can be observed that the classification performance increases significantly from 10 s to 30 s but does not increase too much after 30 s for our method and most other methods. In addition, the 30 s recording time is not very long. Therefore, we select 30 s length to show results from our method and other methods in the following part.

TABLE V
RESULTS OF HEART RATE AND INTER-BEAT INTERVAL ON OBF-P

Training Set (Validation Protocol)	Remote PPG Algorithm	Heart Rate			Inter-beat Interval		
		MAE (bpm)	RMSE (bpm)	R	MAE (ms)	STD (ms)	Accuracy (%)
OBF-P and OBF-H (10-folds)	3DCNN-PEAK	**2.5304**	**5.0275**	**0.9409**	**81.2244**	88.6318	**90.91**
	3DCNN-BVP [28]	3.7356	7.1297	0.8795	95.8779	122.6897	89.45
OBF-P (10-folds)	3DCNN-PEAK	3.3825	6.7187	0.8916	92.8611	99.0031	89.32
	3DCNN-BVP [28]	4.4277	8.3049	0.8350	104.1833	107.9021	88.21
OBF-H (cross dataset)	3DCNN-PEAK	3.6671	6.6110	0.8957	91.9969	**84.8351**	89.08
	3DCNN-BVP [28]	5.4465	9.9317	0.7562	109.1921	107.8691	86.68
-	POS [38]	6.2537	9.7026	0.7836	121.1604	93.5224	85.99
	CHROM [44]	7.2949	10.7810	0.7223	152.2947	90.2355	82.40
	PBV [45]	9.2961	13.0999	0.6294	159.7383	115.8414	82.28

MAE = mean absolute error, RMSE = root mean squared error, R = Pearson Correlation, STD = standard deviation.

TABLE VI
RESULTS OF HEART RATE AND INTER-BEAT INTERVAL ON OBF-H

Training Set (Validation Protocol)	Remote PPG Algorithm	Heart Rate			Inter-beat Interval		
		MAE (bpm)	RMSE (bpm)	R	MAE (ms)	STD (ms)	Accuracy (%)
OBF-P and OBF-H (10-folds)	3DCNN-PEAK	**0.4534**	**1.2868**	0.9954	**25.2568**	22.5873	**96.91**
	3DCNN-BVP [28]	0.7363	2.1517	0.9873	26.8545	25.6097	96.70
OBF-H (10-folds)	3DCNN-PEAK	0.5824	1.3059	**0.9955**	25.4702	**21.9592**	96.79
	3DCNN-BVP [28]	1.2912	3.4956	0.9676	33.1952	37.6494	95.78
OBF-P (cross dataset)	3DCNN-PEAK	1.2621	3.1478	0.9751	33.6131	29.8656	95.78
	3DCNN-BVP [28]	1.4451	4.3984	0.9447	39.4681	56.3431	95.16
-	POS [38]	2.1005	5.3898	0.9221	47.0708	50.3133	94.13
	CHROM [44]	4.1686	7.6021	0.8434	93.3947	51.5717	88.24
	PBV [45]	3.2652	6.2210	0.9106	63.8180	52.8818	92.24

MAE = mean absolute error, RMSE = root mean squared error, R = Pearson Correlation, STD = standard deviation.

TABLE VII
RESULTS OF HEART RATE AND INTER-BEAT INTERVAL ON THE WHOLE OBF DATASET

	Heart Rate			Inter-beat Interval		
	MAE (bpm)	RMSE (bpm)	R	MAE (ms)	STD (ms)	Accuracy
3DCNN-PEAK	**1.4658**	**3.5351**	**0.9686**	**50.7481**	**68.0636**	**94.18**
3DCNN-BVP [28]	2.1371	5.0769	0.9347	58.2923	91.6229	93.40
POS [38]	4.0234	7.6933	0.8568	81.3753	82.3035	90.36
CHROM [44]	5.6161	9.2114	0.7900	120.6662	77.8512	85.53
PBV [45]	6.0576	10.0119	0.7711	108.2304	100.0128	87.63

MAE = mean absolute error, RMSE = root mean squared error, R = Pearson Correlation, STD = standard deviation.

Table VIII shows the classification results for healthy vs. AF when the clip length is 30 s. The accuracy, sensitivity, and specificity for our method are 96.00%, 95.36%, and 96.12%, respectively. For SR vs. AF results in Table IX, the accuracy, sensitivity, and specificity for our method are 95.23%, 98.53%, and 91.12%, respectively. For SR vs. AFL in Table X, the accuracy, sensitivity, and specificity for our method are 88.43%, 90.04%, and 87.04%, respectively. Compared with other methods, our

method performs the best among remote PPG methods and is close to ECG and contact PPG methods.

C. Results for Computational Speed

Our experiments were conducted on an Intel Xeon E5-2650 2.30 GHz CPU and Nvidia Tesla V100 GPU. The cropped face is first extracted from the original face videos by OpenFace [40].
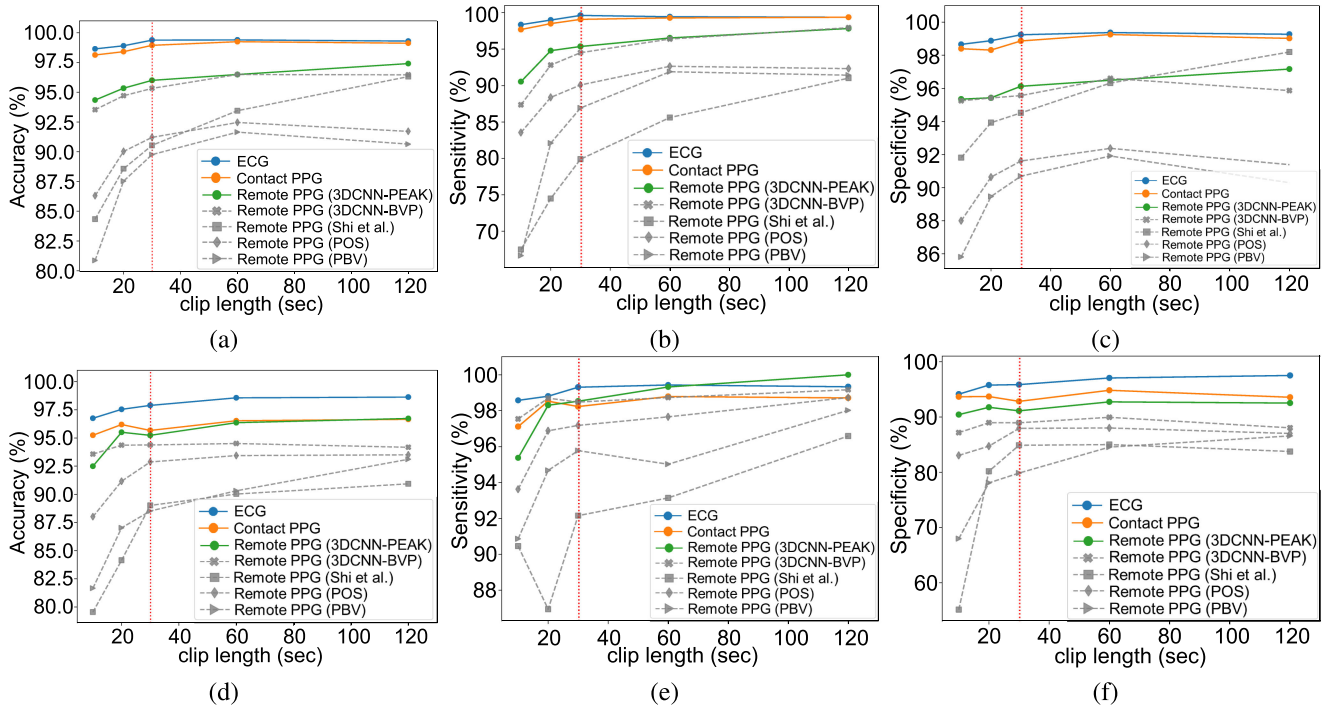
Fig. 8. (a–c) The accuracy, sensitivity, and specificity with respect to the clip length for healthy vs. AF. (d–f) The accuracy, sensitivity, and specificity with respect to the clip length for SR vs. AF. The red dot lines are the 30 s clip length that we choose for our classification experiments. ECG = electrocardiography, PPG = photoplethysmography.

TABLE VIII
RESULTS OF CLASSIFICATION BETWEEN HEALTHY SUBJECTS AND PATIENTS
WITH AF FOR 30 S CLIPS

Signal types	Methods	AC (%)	SE (%)	SP (%)	F1 (%)	AUC
ECG	Reference	99.38	99.63	99.24	99.06	0.9943
	Islam et al. [9]	99.27	99.24	99.28	98.87	0.9915
	Fan et al. [11]	97.52	99.15	96.44	97.12	0.9806
Contact PPG	Reference	98.94	99.08	98.86	98.22	0.9897
Remote PPG	3DCNN-PEAK	**96.00**	**95.36**	**96.12**	**92.67**	**0.9574**
	3DCNN-BVP [28]	95.32	94.51	95.56	91.64	0.9504
	Shi et al. [29]	90.53	79.88	94.51	81.96	0.8720
	POS [38]	91.21	90.08	91.61	85.19	0.9084
	PBV [45]	89.75	86.92	90.69	82.23	0.8881

ECG = electrocardiography, PPG = photoplethysmography, AF = atrial fibrillation, AC = accuracy, SE = Sensitivity, SP = Specificity, AUC = area under curve.

TABLE IX
RESULTS OF CLASSIFICATION BETWEEN PATIENTS WITH SR AND PATIENTS
WITH AF FOR 30 S CLIPS

Signal types	Methods	AC (%)	SE (%)	SP (%)	F1 (%)	AUC
ECG	Reference	97.89	99.31	95.85	98.13	0.9758
	Islam et al. [9]	98.32	99.49	96.87	98.52	0.9762
	Fan et al. [11]	96.53	97.95	94.93	96.11	0.9617
Contact PPG	Reference	95.67	98.23	92.83	96.22	0.9553
Remote PPG	3DCNN-PEAK	**95.23**	**98.53**	**91.12**	**95.87**	**0.9482**
	3DCNN-BVP [28]	94.38	98.48	88.95	95.17	0.9380
	Shi et al. [29]	89.01	92.15	84.87	89.87	0.8851
	POS [38]	92.87	97.18	87.93	93.66	0.9256
	PBV [45]	88.51	95.77	79.86	90.08	0.8782

ECG = electrocardiography, PPG = photoplethysmography. SR = sinus rhythm. AF = atrial fibrillation, AC = accuracy, SE = Sensitivity, SP = Specificity, AUC = area under curve.

TABLE X
RESULTS OF CLASSIFICATION BETWEEN PATIENTS WITH SR AND PATIENTS
WITH AFL FOR 30 S CLIPS

Signal types	Methods	AC (%)	SE (%)	SP (%)	F1 (%)	AUC
ECG	Reference	91.96	92.31	91.36	91.75	0.9183
Contact PPG	Reference	91.33	89.79	94.67	89.50	0.9223
Remote PPG	3DCNN-PEAK	**88.43**	**90.04**	**87.04**	**88.78**	**0.8854**
	3DCNN-BVP [28]	83.18	84.35	82.22	83.47	0.8329
	Shi et al. [29]	79.39	78.64	77.39	76.61	0.7801
	POS [38]	77.76	64.95	79.33	67.34	0.7214
	PBV [45]	73.68	64.36	83.46	66.98	0.7391

ECG = electrocardiography, PPG = photoplethysmography. SR = sinus rhythm. AFL = atrial flutter, AC = accuracy, SE = Sensitivity, SP = Specificity, AUC = area under curve.

The running speed of this step is 30 fps. We train our 3DCNN model with the whole OBF dataset (about 31 hours of face videos) for 45 epochs, which costs about 45 hours. The inference time of the model is about 2 ms (9×10^5 fps) for one 30 s video clip. The training of SVM classification takes 60 ms. For each 30 s clip, the inference time of SVM classification is about 1.2 ms (1.5×10^6 fps).

D. Discussion

From experiment I: remote PPG monitoring, there are three points we can conclude from Tables V, VI, and VII. 1) The evaluation of OBF-P, OBF-H, and the whole OBF dataset shows

that our proposed method can achieve the best performance of remote PPG monitoring on both healthy subject data and patient data. 2) Patient data is more challenging than healthy subject data, but the proposed 3DCNN-PEAK can achieve better results than other baselines on patient data. The results demonstrate that the proposed method with learning systolic peaks effectively improves remote PPG monitoring performance for patients. 3) Training with the whole OBF dataset can perform better than training with only OBF-P or OBF-H. Therefore, we use the remote PPG signals from the model trained on the whole OBF dataset for our classification tasks.

From experiment II: AF/AFL detection, two points can be observed from the results. First, the ECG and contact PPG achieve high accuracy, which validates that the HRV features are effective for AF/AFL detection. Second, our proposed 3DCNN-PEAK achieves slightly lower performance than the contact PPG, but it works best among all remote PPG methods, which indicates that learning systolic peaks can facilitate the AF/AFL detection. We also discuss the performance difference among healthy vs. AF, SR vs. AF, and SR vs. AFL. It can be seen that the results of SR vs. AF are slightly lower than that of healthy vs. AF, which means the SR vs. AF classification task is more challenging. This is expected since both AF and SR samples are from the patients, while the classification of AF vs. healthy uses data from two different groups of subjects whose data is more heterogeneous. In addition, results in Experiment I show that patient data is more challenging and has lower remote PPG monitoring performance, which may cause low classification performance between SR patients and AF patients. It is noted that the performance of all methods for SR vs. AFL is lower than healthy vs. AF or SR vs. AF. There are two reasons. 1) There is only a limited number of AFL patients in our dataset, and the numbers of SR (61 subjects) and AFL (11 subjects) are highly unbalanced, which might cause the performance decreasing. A larger AFL dataset could give a performance improvement and more reliable evaluation. 2) It was reported that adding PPG morphology features helped detect AFL compared to only using HRV features [15]. Since only HRV features are used in our experiment, future works could focus on extracting the remote PPG morphology from face videos to improve AFL detection.

We checked some cases when the detector fails. One reason is that some facial videos contain spontaneous facial motions that lead to noises in remote PPG. Another reason is that some PPG signals from patients contain systolic peaks with very low amplitudes due to small blood stroke volume [21]. The two factors may cause wrongly detected peaks and eventually lead to wrong classification.

Our study may lead to healthcare products, e.g., self-screening of AF symptoms of suspectable populations at home or self-monitoring for the chronic patients to check for recurrence of AF after treatment. Non-contact AF detection can also be used in public places for rapids screening of AF since this method does not need time to attach sensors to the subject, and the computational speed is high enough for real-time application. In addition, the non-contact way is more hygienic without contact sensors.

VI. CONCLUSION

In this paper, we propose a novel method for non-contact AF detection by utilizing the systolic peaks to train our deep learning model. A large-scale OBF dataset was collected for non-contact AF detection from face videos. We achieve high accuracy for non-contact AF detection and demonstrate the feasibility of non-contact AFL detection.

ACKNOWLEDGMENT

REFERENCES

[1] A. F. Members et al., "2012 focused update of the ESC guidelines for the management of atrial fibrillation: An update of the 2010 ESC guidelines for the management of atrial fibrillation developed with the special contribution of the European heart rhythm association," *Eur. Heart J.*, vol. 33, no. 21, pp. 2719–2747, 2012.

[2] M. Fukunami et al., "Detection of patients at risk for paroxysmal atrial fibrillation during sinus rhythm by P wave-triggered signal-averaged electrocardiogram," *Circulation*, vol. 83, no. 1, pp. 162–169, 1991.

[3] S. Mehta, N. Lingayat, and S. Sanghvi, "Detection and delineation of P and T waves in 12-lead electrocardiograms," *Expert Syst.*, vol. 26, no. 1, pp. 125–143, 2009.

[4] S. A. Guidera and J. S. Steinberg, "The signal-averaged P wave duration: A rapid and noninvasive marker of risk of atrial fibrillation," *J. Amer. College Cardiol.*, vol. 21, no. 7, pp. 1645–1651, 1993.

[5] R. Couceiro, P. Carvalho, J. Henriques, M. Antunes, M. Harris, and J. Habetha, "Detection of atrial fibrillation using model-based ECG analysis," in *Proc. 19th Int. Conf. Pattern Recognit.*, 2008, pp. 1–5.

[6] K. Tateno and L. Glass, "Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals," *Med. Biol. Eng. Comput.*, vol. 39, no. 6, pp. 664–671, 2001.

[7] S. Dash, K. Chon, S. Lu, and E. Raeder, "Automatic real time detection of atrial fibrillation," *Ann. Biomed. Eng.*, vol. 37, no. 9, pp. 1701–1709, 2009.

[8] M. S. Islam, N. Ammour, N. Alajlan, and H. Aboalsamh, "Rhythm-based heartbeat duration normalization for atrial fibrillation detection," *Comput. Biol. Med.*, vol. 72, pp. 160–169, 2016.

[9] M. S. Islam, M. M. B. Ismail, O. Bchir, M. Zakariah, and Y. A. Alotaibi, "Robust detection of atrial fibrillation using classification of a linearly-transformed window of RR intervals tachogram," *IEEE Access*, vol. 7, pp. 110012–110022, 2019.

[10] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati, "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, 2017, pp. 141–144.

[11] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1744–1753, Nov. 2018.

[12] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 12, pp. 2095–2104, Dec. 2018.

[13] M.-Z. Poh et al., "Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms," *Heart*, vol. 104, no. 23, pp. 1921–1928, 2018.

[14] S. Kwon et al., "Deep learning approaches to detect atrial fibrillation using photoplethysmographic signals: Algorithms development study," *JMIR mHealth uHealth*, vol. 7, no. 6, 2019, Art. no. e12770.

[15] L. M. Eerikäinen et al., "Detecting atrial fibrillation and atrial flutter in daily life using photoplethysmography data," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 6, pp. 1610–1618, Jun. 2020.

[16] J. Lee, B. A. Reyes, D. D. McManus, O. Maitas, and K. H. Chon, "Atrial fibrillation detection using an iphone 4 s," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 203–206, Jan. 2013.

[17] L. Krivoshei et al., "Smart detection of atrial fibrillation," *EP Europace*, vol. 19, no. 5, pp. 753–757, 2017.

[18] A. G. Bonomi et al., "Atrial fibrillation detection using a novel cardiac ambulatory monitor based on photo-plethysmography at the wrist," *J. Amer. Heart Assoc.*, vol. 7, no. 15, 2018, Art. no. e009351.

[19] M. Dörr et al., "The watch AF trial: Smartwatches for detection of atrial fibrillation," *JACC: Clin. Electrophysiol.*, vol. 5, no. 2, pp. 199–208, 2019.

[20] V. D. Corino, R. Laureanti, L. Ferranti, G. Scarpini, F. Lombardi, and L. T. Mainardi, "Detection of atrial fibrillation episodes using a wristband device," *Physiol. Meas.*, vol. 38, no. 5, 2017, Art. no. 787.

[21] E.-S. Väliaho et al., "Wrist band photoplethysmography in detection of individual pulses in atrial fibrillation and algorithm-based detection of atrial fibrillation," *EP Europace*, vol. 21, no. 7, pp. 1031–1038, 2019.

[22] T. Wu, V. Blazek, and H. J. Schmitt, "Photoplethysmography imaging: A new noninvasive and noncontact method for mapping of the dermal perfusion changes," in *Proc. SPIE Opt. Techn. Instrum. Meas. Blood Comp., Struct., Dyn.*, 2000, vol. 4163, pp. 62–70.

[23] L. Iozzia, L. Cerina, and L. Mainardi, "Relationships between heart-rate variability and pulse-rate variability obtained from video-PPG signal using ZCA," *Physiol. Meas.*, vol. 37, no. 11, 2016, Art. no. 1934.

[24] J.-P. Couderc et al., "Detection of atrial fibrillation using contactless facial video monitoring," *Heart Rhythm*, vol. 12, no. 1, pp. 195–201, 2015.

[25] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Exp.*, vol. 16, no. 26, pp. 21434–21445, 2008.

[26] American Burn Association, "Burn incidence fact sheet," Accessed: Oct. 5, 2021. [Online]. Available: https://ameriburn.org/who-we-are/media/burn-incidence-fact-sheet/

[27] X. Li et al., "The OBF database: A large face video database for remote physiological signal measurement and atrial fibrillation detection," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2018, pp. 242–249.

[28] Z. Yu, X. Li, and G. Zhao, "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," in *Proc. 30th Brit. Mach. Vis. Conf.*, Cardiff, U.K., 2019, Art. no. 277.

[29] J. Shi, I. Alikhani, X. Li, Z. Yu, T. Seppänen, and G. Zhao, "Atrial fibrillation detection from face videos by fusing subtle variations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2781–2795, Aug. 2020.

[30] B. P. Yan et al., "Contact-free screening of atrial fibrillation by a smartphone using facial pulsatile photoplethysmographic signals," *J. Amer. Heart Assoc.*, vol. 7, no. 8, 2018, Art. no. e008585.

[31] V. D. Corino, L. Iozzia, G. Scarpini, L. T. Mainardi, and F. Lombardi, "A simple model to detect atrial fibrillation via visual imaging," *Biomed. Eng./Biomedizinische Technik*, vol. 65, no. 6, pp. 721–728, 2020.

[32] S. Islam, N. Ammour, and N. Alajlan, "Atrial fibrillation detection with multiparametric RR interval feature and machine learning technique," in *Proc. Int. Conf. Inform., Health Technol.*, 2017, pp. 1–5.

[33] R. S. Andersen, A. Peimankar, and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," *Expert Syst. Appl.*, vol. 115, pp. 465–473, 2019.

[34] M. Butkuviene, A. Petrenas, A. Solosenko, A. Martin-Yebra, V. Marozas, and L. Sornmo, "Considerations on performance evaluation of atrial fibrillation detectors," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 11, pp. 3250–3260, Nov. 2021.

[35] C. Bruser, J. Diesel, M. D. Zink, S. Winter, P. Schauerte, and S. Leonhardt, "Automatic detection of atrial fibrillation in cardiac vibration signals," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 1, pp. 162–171, Jan. 2013.

[36] T. Hurnanen et al., "Automated detection of atrial fibrillation based on time-frequency analysis of seismocardiograms," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1233–1241, Sep. 2017.

[37] O. Lahdenoja et al., "Atrial fibrillation detection via accelerometer and gyroscope of a smartphone," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 108–118, Jan. 2018.

[38] W. Wang, A. C. den Brinker, S. Stuijk, and G. De Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017.

[39] L. Feng, L.-M. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 879–891, May 2015.

[40] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2018, pp. 59–66.

[41] D. McDuff, S. Gontarek, and R. W. Picard, "Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 12, pp. 2948–2954, Dec. 2014.

[42] E. Cazelles, A. Robert, and F. Tobar, "The wasserstein-fourier distance for stationary time series," *IEEE Trans. Signal Process.*, vol. 69, pp. 709–721, 2020.

[43] D. Makowski et al., "Neurokit2: A Python toolbox for neurophysiological signal processing," *Behav. Res. Methods*, vol. 53, pp. 1689–1696, 2021.

[44] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.

[45] G. De Haan and A. Van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiol. Meas.*, vol. 35, no. 9, 2014, Art. no. 1913.

[46] X. Liu, X. Yang, J. Jin, and A. Wong, "Detecting pulse wave from unstable facial videos recorded from consumer-level cameras: A disturbance-adaptive orthogonal matching pursuit," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 12, pp. 3352–3362, Dec. 2020.