BIOWISH: Biometric Recognition Using Wearable Inertial Sensors Detecting Heart Activity

Emanuele Maiorana*, Senior Member, IEEE, Chiara Romano*, Student Member, IEEE, Emiliano Schena*, Senior Member, IEEE, and Carlo Massaroni*, Senior Member, IEEE

Abstract—Wearable devices have been recently proposed to perform biometric recognition, leveraging on the uniqueness of the collectable physiological traits to generate discriminative identifiers. Most of the studies conducted on this topic have exploited heart-related signals, sensing the cardiac activity either through electrical measurements using electrocardiography, or with optical recordings employing photoplethysmography. In this article we instead propose a system performing BIOMetric recognition using Wearable Inertial Sensors detecting Heart activity (BIOWISH). In more detail, we investigate the feasibility of exploiting mechanical measurements obtained through seismocardiography and gyrocardiography to verify the identity of a subject. Several feature extractors and classifiers, including deep learning techniques relying on siamese training, are employed to derive distinctive characteristics from the considered signals, so as to differentiate between legitimate users and impostors. A multi-session database, comprising acquisitions taken from subjects performing different activities, is employed to perform experimental tests. The obtained results testify that identifiers derived from measurements of chest vibrations, collected by wearable inertial sensors, could be employed to guarantee high recognition performance, even when considering short-time recordings. Explainability methods have been also employed to derive some insights about the aspects relevant to perform predictions for both people and activity recognition tasks.

Index Terms—Biometrics, deep learning, explainability, seismocardiography, gyrocardiography.

I. INTRODUCTION

THANKS to their ability to carry out non-invasive measurements of physiological traits, wearable devices can be nowadays used for a wide set of possible applications, justifying their reputation as one of the “next big things” in technological innovation after smartphones [1]. While they have long been used for fitness purposes, with data related to the heart rate or the calorie consumption employed to track the achievement of specific goals [2], a growing interest is being devoted to the usage of wearable devices for medical purposes, exploiting their portability to monitor critical health parameters over long time periods, while preserving as much as possible the comfort and freedom of movement of the involved subjects [3]. Furthermore, wearable devices have been recently proposed to perform automatic biometric recognition [4], with the aim of providing security when sharing data in Internet of Things (IoT) frameworks, a major requirement when implementing sensitive value-added services such as smart payment [5].

Actually, physical, behavioral, and cognitive traits can be collected through wearable sensors, and processed to derive discriminative features that can be exploited to differentiate a legitimate subject wearing the employed devices from potential impostors. A notable peculiarity of recognition systems relying on wearable devices, with respect to traditional approaches, consists in the feasibility of collecting data autonomously, without the need for the involved users to interact with dedicated infrastructures, thus allowing a more convenient and user-friendly acquisition procedure. This is especially true for cognitive biometric traits [6], that is, physiological signals depending on the autonomic nervous system (ASN), that can be collected even when the involved subjects do not perform any voluntary action. Biometric systems relying on such traits can also guarantee improved security, due to the fact that the recorded characteristics typically cannot be captured remotely, inherently provide liveness detection, and can be also examined to infer about the mental and emotional states of an individual, providing the means to detect compulsion attacks [7]. Moreover, since the interested traits can be acquired at any time and place, wearable biometric systems can perform continuous recognition, i.e., a user’s identity can be verified throughout a session, preventing hijacking and avoiding attackers’ unauthorized access after an initial successful recognition [8].

Several cognitive characteristics collectable through wearable devices have been proposed in literature to perform biometric recognition [9], including traits related to cardiac [10], brain [11], respiratory [12], ocular [13], and epidermal [14] activities. Among such sources of information exploitable for biometric recognition purposes, the heart plays a pivotal role. In fact, most of the research conducted on wearable biometric systems has been focused on the analysis of cardiac signals [9]. This is due to the fact that heart activity is analyzed for several medical applications, and integrating biometric recognition within them could reduce the risks of identity fraud in healthcare, especially for remote monitoring [15]. Moreover, several commercial products able to detect heart activity are nowadays

Authorized licensed use limited to: University of Michigan. Downloaded on July 23,2023 at 10:16:00 UTC from IEEE Xplore. Restrictions apply.
available, and wearing such devices typically does not involve significant inconvenience for the users, thus resulting in high acceptability. Furthermore, the permanence of discriminative characteristics within cardiac signals, a fundamental property for their usage in practical biometric recognition systems, has been addressed in several studies relying on medical-grade equipment [16]. It has also to be mentioned that heart activity can be sensed through many different approaches. The technique most commonly employed to assess cardiac functioning is electrocardiography (ECG), performed placing electrodes on a subject’s skin to detect small electrical changes resulting from the heart depolarization and repolarization [17]. Photoplethysmography (PPG) instead measures variations in blood volume pulse (BVP) and oxygen saturation (SpO2) illuminating the skin at close distance with near-infrared (NIR) light [18]. Digital stethoscopes can be also employed to acquire acoustic information on the heart activity through phonocardiography (PCG) [19]. Eventually, also mechanical sensing can be exploited to acquire information on a subject’s cardiovascular activity. In more detail, seismocardiography (SCG) [20] and gyrocardiography (GCG) [21] respectively exploit accelerometers and gyroscopes to collect signals depending on the low-frequency vibrations caused by heart pumping, when placing sensors on a subject’s chest. The vibrations generated by the cardiac activity can also be detected in body parts far from the heart, such as the head, using ballistocardiography (BCG) [22]. Illustrative examples of the heart-related signals collected through different sensing techniques are given in Fig. 1, along with indications about the most relevant events performed during a cardiac cycle.

While ECG and PPG signals collected through wearable devices have been exploited for biometric recognition purposes in several researches, only a handful of studies have evaluated the feasibility of extracting discriminative characteristics from SCG or GCG data. Such a possibility would actually be concretely interesting for developing innovative real-world biometric systems. In fact, it is nowadays possible to collect accurate mechanical measurements with commercial micro-electromechanical sensors (MEMS) [24]. Such devices are typically inexpensive, unobtrusive, resistant to electromagnetic and acoustic interferences, safe and viable for prolonged use, since they commonly have an extremely low power consumption [23]. Actually, SCG and GCG techniques are rapidly spreading in medical scenarios as alternatives to ECG for monitoring the pre- and post-procedural conditions of patients [25], allowing to estimate hemodynamic parameters that can be hardly recovered using other sensing techniques [26], [27]. Several other applications, including posture analysis and context retrieval, have been also designed using chest-worn inertial sensors [28].

In this context, this paper presents a biometric recognition system referred to as BIOWISH, since it relies on Wearable Inertial Sensors to collect mechanical measurements related to the Heart activity. In more detail, we here evaluate:

- the feasibility of recognizing an individual, at an average time distance of 21 days from enrolment, using information about the subject’s heart activity collected with commercial wearable inertial measurement units (IMUs);
- the effectiveness of deep learning strategies relying on novel convolutional neural networks (CNNs) and siamese training to generate discriminative representations of SCG and GCG waveforms for a healthy subject. Relevant events include the mitral valve closure (MC) and opening (MO), the aortic valve closure (AC) and opening (AO), the S1 sound at the closure of the atriioventricular (mitral and tricuspid) valves, the S2 sound at the closure of the semilunar (aortic and pulmonary) valves, the systolic nadir (A) and end diastole (B). The H wave head-ward deflection begins close to the peak of the R wave and has maximum peak near the end of ejection; the I wave foot-ward deflection follows the H wave and occurs early in systole; the J wave largest head-ward wave follows the I wave and occurs late in systole; the K wave foot-ward wave follows the J wave and occurs before the end of systole (adapted from [23]).

The paper is organized as follows: Section II provides an overview on the state of the art of wearable biometric recognition systems based on heart activity. The techniques employed in the proposed BIOWISH system to perform biometric recognition using heart-based inertial measurements are presented in Section III. The collected database and the tests carried out on it are outlined in Section IV. Conclusions drawn from the performed work are reported in Section V.
II. Wearable Biometric Recognition Based on Heart Activity

As mentioned in the previous section, the physiological signals associated with the cardiac muscle functioning have long been studied with the aim of extracting person-specific information, and exploiting it for biometric recognition purposes, considering in most cases ECG signals [29]. Actually, the vast majority of such studies relied on medical-grade equipment with multiple electrodes, able to collect data with a high signal-to-noise ratio (SNR) [17]. Yet, due to the costs of the used devices, and to the commonly uncomfortable acquisition conditions, these solutions have been typically restricted only to laboratory environments, with limited real-world applications of heart-based biometric recognition so far proposed.

Wearable devices have instead recently allowed to easily collect information related to a subject’s cardiac activity even outside controlled environments. In detail, biometric approaches relying on one- or two-lead ECG data have been proposed by exploiting sensors embedded in chestbands [30], [31], armbands [32], [33], and also electronic textiles, that is, fabrics with embedded electronics [10], [34]. Unfortunately, the quality of ECG signals collected through wearable devices is typically lower than that guaranteed by medical instrumentation [35], with the attainable recognition performance thus negatively affected.

Also PPG has been employed to design biometric systems relying on wearable devices in a considerable amount of papers [36]. Such optical measurements can be taken by either sensing transmitted light placing sensors on fingertips [37], or detecting reflected light with sensors placed on the wrist [38]. The collected PPG signals could be either directly processed to extract discriminative features to perform recognition [39], or employed to estimate coarse metrics concerning cardiovascular activity such as heart rate variability (HRV) [40] or metabolic equivalent of task (MET) [41], before exploiting the obtained measures to recognize a subject.

Quite large medical equipment has been also usually employed to collect PCG data [42], yet recent advances in wearable technology, consisting in the development of small sound sensors, could allow improving the acceptability of this acquisition modality, paving the way for practical recognition systems relying on PCG [43].

Mechanical measurements of the heart activity have been instead employed for biometric recognition purposes only in [44], [45], [46], and [47], where SCG data have been considered, and in [48], where also GCG signals have been exploited for the first time. The involved subjects have performed data collection in sitting conditions in [44], [45], lying supine in [46], while also standing and walking conditions have been considered in [48] and [47].

Head-mounted devices have been used to collect BCG signals in [22], [49]. However, BCG measurements are notably affected by any contact of the body with external objects, including the floor and the measuring devices, since they may interfere with, or even impede, the body displacement induced by recoil forces [50], with BCG therefore commonly providing less discriminative information than SCG or GCG.

A. State of the Art Analysis

In order to verify the effectiveness of the BIOWISH system here proposed, proper tests have been carried out exploiting an in-house multi-session database, specifically collected with the aim of checking the stability over time of the discriminative capabilities of the considered biometric traits. Actually, it is worth noting that the permanence of the employed identifiers is an aspect often neglected in biometric studies, especially when dealing with traits requiring a considerable amount of setup time to be acquired. This happens despite it is a widely known fact that recognition results computed by comparing data from distinct sessions are notably worse than those accomplished on single-session datasets [18]. It is also worth remarking that, for physiological signals employed as biometric characteristics, the extracted features may be additionally affected by personal habits, such as those related to the current diet, and even by the specific physical condition and emotional state of a subject. Furthermore, when using wearable devices to collect the interested biometric characteristics, also the specific placement of the employed on-body sensors can influence the performed recordings, since it is very unlikely to attach the employed devices each time at the very same position. It is therefore incorrect to estimate the recognition performance achievable in real-world applications using only single-session databases.

When considering heart-based biometric recognition systems relying on wearable devices, only few works have till now investigated the permanence of the exploited identifiers, performing tests on multi-session databases. In more detail, such works are listed in the summary reported in Table I. It can be seen that longitudinal studies have been conducted only on ECG, PPG, and SCG signals, among all the possible heart-based biometric traits. Furthermore, time spans significantly larger than a week between acquisition sessions have been considered only in a handful of studies [10], [33], [37], [38], [51], [52]. Commercial devices have been employed in most of the studies focused on a single modality, while prototype systems comprising multiple sensors have been used in multi-biometric systems relying on both ECG and PPG. Some works have evaluated the feasibility of recognizing a subject while performing multiple activities of daily living (ADL) such as sleeping [52], walking [51], or writing [38], with samples taken in the same conditions compared for recognition purposes. Recordings taken in the wild, i.e., with subjects involved in daily routines, have been instead considered only for ECG [10], [30], [34], with recognition there carried out regardless the specific performed activity. The duration of the signals employed to perform recognition is often greater than 10 s, sometimes requiring recordings lasting even more than a minute, especially when coarse metrics such as HRV or MET are derived from the collected data [40], [51].

It is also worth observing that, to test the mentioned approaches, most of the performed studies have considered identification scenarios, that is, conditions in which a systems determines the identity of a presented subject exploiting the availability of information taken from multiple users, and system performance is measured in term of identification rate (IR). In more detail, closed-set (CS) identification, in which it is
TABLE I
SUMMARY OF STATE-OF-THE-ART APPROACHES USING BIOMETRIC TRAITS ACQUIRED THROUGH WEARABLE DEVICES FOR AUTOMATIC PEOPLE RECOGNITION

| Signal | Paper | Subjects | Sessions | Device | Multiple conditions | Time for recognition | Feature | Comparator | Recognition modality | Performance |
|--------|-------|----------|----------|--------|---------------------|----------------------|---------|------------|---------------------|-------------|
| ECG    | [10]  | 5        | over 6 months | Commercial | Yes | 6 heartbeats | Wavelet | SVM | CS Identification | 70-100% |
|        | [34]  | 33       | over 6 weeks  | Commercial | Yes | 10 heartbeats | Learned | CNN | CS Identification | 95.9% |
|        | [30]  | 20       | 6 in 6 days   | Commercial | Yes | 3 heartbeats | Statistical | RF | CS Verification | 21.9% |
|        | [18]  | 56       | 2 in 2 days   | Prototype | No | 30 heartbeats | Time signal | L2 dist. | OS Verification | 21.5% |
|        | [51]  | 400      | over 17 months| Commercial | Yes | 5 min | Statistical | OC-SVM | OS Verification | >20% |
|        | [38]  | 7        | 6 in 50 days  | Prototype | Yes | 4 heartbeats | Statistical | RF | CS Identification | 90.97% |
|        | [52]  | 20       | > 20 days     | Commercial | Yes | 10 min | Learned | CNN | CS Identification | 55.8% |
|        | [53]  | 12       | 4 in 4 days   | Prototype | No | 8 s | Learned | CNN | CS Identification | 95.7% |
|        | [37]  | 100      | 3 in 17 days  | Prototype | No | 20 heartbeats | Learned | CNN+RNN | CS Identification | 87.1% |
|        | [14]  | 17       | 2 in 1 week   | Commercial | Yes | 10 s | Spectrogram | CNN | CS Identification | 94.9% |
|        | [40]  | 74       | over 5 days   | Commercial | Yes | 1 min | Statistical | CNN+RNN | CS Verification | 17.9% |
| SCG    | [44]  | 10       | over 1 week   | Prototype | No | 1 s | Spectrogram | GMM | CS Verification | 12.0% |
| ECG.PPG| [54]  | 25       | 2 in two days | Prototype | No | 1 min | Statistical | SVM | CS Identification | 92% |
|        | [33]  | 53       | 2 in 8 weeks  | Prototype | Yes | 1 heartbeat | Time Signal | L2 dist. | OS Verification | 13% |

assumed to know all the possible users which could access the system, therefore without the possibility of rejecting an unknown subject, has been always performed in the considered studies. Employed classifiers include support vector machine (SVM), random forest (RF), Gaussian mixture model (GMM), CNNs and recurrent neural networks (RNNs). Conversely, verification has been evaluated only in few works, and open-set (OS) conditions have been considered only in a subset of them [18], [33], [51]. For instance, CS verification has been taken into account in [44], estimating for each user a model depending on characteristics of the same impostors also considered for tests. Unfortunately, only OS conditions instead properly resemble operative scenarios, where data taken from potential impostors, considered to test a given recognition system, are not available during the enrolment of a legitimate user. The recognition results achieved in the performed works, normally expressed in terms of EER, are commonly quite high, always greater than 10%, mainly due to the significant differences between the characteristics of the physiological signals collected in different recording sessions. This aspect is easily noticeable when comparing the performance reported in papers exploiting SCG for biometric purposes: while very good recognition rates are given in [33], [45], [46], [47], and [48], where single-session databases have been considered, much worse results are shown in [44], where data collected during two distinct sessions are compared during the recognition process.

Given the exposed state of the art, the aim of the present study is to design a recognition system performing OS verification, able to achieve good recognition performance when comparing heart-related signals collected by wearable inertial sensors, over short time intervals, and in distinct sessions.

III. PROPOSED BIOMETRIC RECOGNITION SYSTEM

In order to perform a thorough analysis of the discriminative capabilities of SCG and GCG signals, several strategies are proposed in the BIOWISH system to derive, from the considered data, different parametric representations to be used for verification purposes. In more detail, in the following it is assumed that a generic recording can be expressed through the signals \( \{p_x, p_y, p_z\} \), each lasting \( S \) seconds and sampled at a rate \( Q \), with \( \{x, y, z\} \) being the three axes provided by an accelerometer or gyroscope sensor. The acquired data have to be first preprocessed as described in Section III-A. Then, different parametric representations, to be used when computing the dissimilarity distances between probe and enrolment samples upon which the verification process has to be carried out, are derived as detailed in Section III-B. In the performed tests, the considered feature extractors are trained according to the learning strategies outlined in Section III-C.

A. Preprocessing

The collected SCG and GCG recordings are first low-pass filtered retaining the subband below 25 Hz, relevant for heart activity, and then resampled at a rate \( R = 60 \) Hz to reduce the computational complexity of the subsequent processing. The resulting signals are then segmented into overlapping frames, using rectangular time windows of length \( L = 5 \) s with an overlap \( O = 80\% \), thus generating a set of \( K = \lceil 1 + \frac{1}{2} (\frac{S}{L} - 1) \rceil \) segments, each comprising \( 3 \cdot (L \cdot R) \) samples, from an original acquisition. The created frames are treated as individual samples in the considered systems, i.e., a verification is carried out comparing a query probe, consisting of a single frame lasting \( L \) seconds, with an enrolment set comprising \( K \) frames generated from an original recording of duration \( S \).

In more detail, a generic frame can be expressed through time-domain coefficients as a \( 3 \times (L \cdot R) = 3 \times 300 \) matrix \( V \), with each row of this matrix containing values corresponding to one of the three axes of the employed accelerometer or gyroscope. Alternatively, an expression in the time/frequency-domain can be derived as a \( 25 \times 41 \times 3 \) tensor \( W \), containing the logarithm of the power spectral densities (PSDs) obtained from the short-time Fourier transforms (STFT) computed over each of the three available axes, using a frequency resolution of 1 Hz over the subband of interest, and 41 Hamming time windows within the considered frame length \( L \).
B. Feature Extractors

A given frame, expressed through either \( V \) or \( W \), can be processed according to multiple approaches in order to derive different discriminative representations. The processing techniques considered to create templates from the treated SCG and GCG data have been selected taking in mind the need for a low computation complexity, since the proposed system should be designed in order to be implemented on wearable devices with limited computation capabilities. The possibilities evaluated in the performed tests are presented in the following, along with the methods adopted to compare frames acquired during the enrolment and verification phases.

1) Dictionary Learning: Dictionary learning has been successfully applied to several heart-related physiological traits \([55],[56]\). Techniques belonging to this category aim at finding a sparse representation of the input data by linearly combining the coefficients of the original representations. Sparsity is achieved learning dictionary atoms lying in a space whose dimensionality is greater than the one of the original input, thus allowing to achieve proper flexibility when handling the intra-class variability of the collected data. These approaches could be suitable for the intended BIOWISH system, since the learning process could be implemented in efficient ways, requiring limited computational complexity.

In the performed tests, we have evaluated the effectiveness of label consistent k-means singular value decomposition (LC-KSVD) \([57]\) and sparse-representation-based classification (SRC) \([58]\). As detailed in Section III-C, subject-specific dictionaries of SCG and GCG data are estimated using both techniques during the enrolment of each user, and applied to the probe samples collected during verification to estimate whether the considered data could belong to the legitimate identity. Sparse dictionary learning is here applied to the time/frequency-domain coefficients in \( W \), having observed through experimental tests that better performance are achieved in this way, with respect to considering the time-domain coefficients in \( V \).

2) Deep Learning: Deep learning techniques have been demonstrated to be extremely efficient at extracting discriminative information from several biometric traits \([59]\). Nevertheless, a major drawback limiting their use in wearable devices is given by the high computational complexity required by networks designed, for instance, to perform image recognition tasks \([60]\). For this reason, we have here evaluated whether networks with fairly low computational complexity can be exploited to perform SCG and GCG-based biometric recognition.

In more detail, we have designed two distinct networks to derive discriminative representations from either the time coefficients in \( V \) or the time/frequency elements in \( W \). Table II details a novel architecture, named WISHNET\(_T\), here proposed to extract a discriminative representation with 1024 coefficients from the time/frequency content \( W \) of a frame. The employed architecture comprises 4 convolutional (CONV) layers with rectified linear units (ReLUs) as non-linear activation functions, 3 max-pooling (MP) layers, and 2 dropout (DO) layers. Its size is lower than ResNet18.

Features in the time domain are also considered as inputs to another CNN, indicated as WISHNET\(_T\), and detailed in Table III, here designed and employed to extract a 128-coefficient template from the time-domain frame representation \( V \). A batch normalization (BN) layer is employed in WISHNET\(_T\) to handle the variability of the input data. The size of WISHNET\(_T\) is lower than extremely compact networks such as MobileNetv2.

It is worth mentioning that additional representations have been investigated in the performed tests, for instance using the time-domain coefficients \( V \) as input to long short-term memory (LSTM) RNNs, yet the results obtained in such scenarios are not comparable with those achieved through the representations detailed in this section, and are therefore not considered in the following for biometric recognition purposes.

C. Representation Learning

Different strategies have been used to train the employed dictionary learning and deep learning models. The adopted

---

**TABLE II**

| Layer | Filter | Pad | Input | Output |
|-------|--------|-----|-------|--------|
| Conv  | \( (3\times3\times3) \times 128 \) | \([1,2]\) | \( 25\times41\times3 \) | \( 25\times41\times128 \) |
| ReLu  | -      | -   | \( 25\times41\times128 \) | \( 25\times41\times128 \) |
| MP    | 2x2    | -   | \( 25\times41\times128 \) | \( 12\times20\times128 \) |
| DO    | -      | -   | \( 12\times20\times128 \) | \( 12\times20\times128 \) |
| Conv  | \( (3\times5\times128) \times 256 \) | \([1,2]\) | \( 12\times20\times128 \) | \( 12\times20\times256 \) |
| ReLu  | -      | -   | \( 12\times20\times256 \) | \( 12\times20\times256 \) |
| MP    | 2x2    | -   | \( 12\times20\times256 \) | \( 6\times10\times256 \) |
| Conv  | \( (3\times5\times256) \times 512 \) | \([1,2]\) | \( 6\times10\times256 \) | \( 6\times10\times512 \) |
| Relu  | -      | -   | \( 6\times10\times512 \) | \( 6\times10\times512 \) |
| MP    | 2x2    | -   | \( 6\times10\times512 \) | \( 3\times5\times512 \) |
| DO    | -      | -   | \( 3\times5\times512 \) | \( 3\times5\times512 \) |
| Conv  | \( (3\times5\times12) \times 1024 \) | \([0,0]\) | \( 5\times5\times512 \) | \( 1\times1\times1024 \) |
| ReLu  | -      | -   | \( 5\times5\times512 \) | \( 1\times1\times1024 \) |

**TABLE III**

| Layer | Filter | Pad | Input | Output |
|-------|--------|-----|-------|--------|
| Conv  | \( (3\times3\times1) \times 32 \) | \([0,0]\) | \( 3\times3\times1 \) | \( 1\times1\times32 \) |
| BN    | -      | -   | \( 1\times3\times32 \) | \( 1\times1\times32 \) |
| ReLu  | -      | -   | \( 1\times3\times32 \) | \( 1\times1\times32 \) |
| MP    | 1x2    | -   | \( 1\times3\times32 \) | \( 1\times1\times32 \) |
| DO    | -      | -   | \( 1\times1\times32 \) | \( 1\times1\times32 \) |
| Conv  | \( (1\times5\times32) \times 64 \) | \([0,0]\) | \( 1\times7\times32 \) | \( 1\times1\times64 \) |
| ReLu  | -      | -   | \( 1\times7\times32 \) | \( 1\times1\times64 \) |
| MP    | 1x2    | -   | \( 1\times7\times32 \) | \( 1\times1\times32 \) |
| DO    | -      | -   | \( 1\times1\times32 \) | \( 1\times1\times32 \) |
| Conv  | \( (1\times5\times64) \times 64 \) | \([0,0]\) | \( 1\times1\times32 \) | \( 1\times1\times64 \) |
| ReLu  | -      | -   | \( 1\times1\times32 \) | \( 1\times1\times64 \) |
| MP    | 1x2    | -   | \( 1\times1\times64 \) | \( 1\times1\times32 \) |
| Conv  | \( (1\times3\times64) \times 128 \) | \([0,0]\) | \( 1\times1\times64 \) | \( 1\times1\times128 \) |
| ReLu  | -      | -   | \( 1\times1\times128 \) | \( 1\times1\times128 \) |
strategies are designed in order to derive discriminative representations allowing to perform OS verification. Specifically, it is assumed that, in addition to data captured from a specific legitimate user, a set of acquisitions captured from other subjects is also made available during an enrolment process. These additional resources are exploited to train user-specific binary classifiers, able to discriminate between the legitimate subject and potential impostors. Such strategy is adopted to derive the desired representations for methods relying on both dictionary learning and deep learning. In the latter case, the networks are trained adding a dropout layer, and a softmax producing two outputs for the considered binary classification task, to the architectures in Tables II and II, and using a cross-entropy (CE) loss for the backpropagation learning algorithm.

However, the aforementioned learning approach implies that a different model has to be trained for each enrolled user. In order to learn a discriminative representation that can be used for any person, a different learning strategy can alternatively be exploited. In more detail, the availability of acquisitions taken from a set of training subjects can be exploited to train networks according to a siamese strategy [61]. Such approach involves the simultaneous update of two parallel networks, with the aim of generating representations close in the considered space in case samples from the same subjects are used as inputs, and distant in case samples from different subjects are fed to the two networks. Instead of using a softmax with a CE training loss as in the standard classification task, this learning procedure simply computes an Euclidean (L2) distance between the representations obtained from the two parallel networks, and evaluates a contrastive loss function whose minimization is sought through the backpropagation learning algorithm. Resorting to siamese learning strategies has the advantage that the training process can be performed only once, with the resulting network usable for any user to be enrolled in the system, without the need for training a subject-specific network for every user, and therefore significantly simplifying the enrolment procedure. Furthermore, as it will be further detailed in Section IV, a siamese learning process can be designed in order to specifically learn stable representations with proper permanence, selecting appropriately the pairs used as inputs for the parallel networks.

The verification process is conducted as a standard binary classification task when learning subject-specific models. Conversely, the features extracted from a probe frame are compared against all the representations generated from the enrolment acquisition of the claimed identity, selecting the lowest computed distance as representative of the comparison process, when using networks trained with a siamese strategy.

### IV. EXPERIMENTAL TESTS

The database collected to evaluate the proposed system is presented in Section IV-A. Results achievable on it when using the proposed methods and state-of-the-art approaches are presented in Section IV-B. The feasibility of recognizing a person without having to know a priori the performed activity is considered in Section IV-D. An explainability analysis is eventually conducted in Section IV-E.

#### A. Employed Database and Settings

Data related to the heart activity of 31 subjects, 16 male and 15 female, with ages between 20 and 41 years and without evidence of cardiovascular diseases, have been collected and exploited to evaluate the effectiveness of the proposed approach. In more detail, SCG and GCG signals have been collected, for each involved subject, from five different wearable Xsens Xdot\(^1\) MEMS IMU sensors, each attached at different positions of the subjects’ chest using a hypoallergenic adhesive. The landmarks where the sensors have been placed, commonly adopted in medical applications, are shown in Fig. 2, and correspond to:

- the focus of auscultation of the pulmonary (P) valve, between the first and second rib;
- the auscultation focus of the aortic (A) valve, between the first and second ribs;
- the xiphoid (X) process on the lower part of the sternum;
- the mitral (M) valve auscultation focus, between the fourth and fifth ribs;
- the tricuspid (T) valve auscultation focus, between the third and fourth ribs.

Each sensor has a size of 36.30 × 30.35 × 10.80 mm, and a weight of 10.8 g, being therefore extremely light and comfortable to wear, and contains a three-axis accelerometer and a three-axis gyroscope. The acceleration signals have been used as SCG data, while the angular velocities as GCG traits. The sensors coordinate system has a body-fixed right-handed Cartesian reference, as shown in Fig. 2. All the IMUs have been synchronized before starting data acquisition, so that all

---

\(^1\) https://www.xsens.com/xsens-dot
the sensor data were time-synced. The algorithm synchronizes the clocks to an error smaller than $20 \mu s$, which decays to an error smaller than $1.8$ ms after 30 min [62]. The acquired data have been stored in the device memory at a sampling frequency $Q = 120$ Hz, and then downloaded to the laptop before being processed as described in Section III. During a recording session, four different ADL have been performed by each subject, that is, lying, standing, sitting, and walking, as shown in Fig. 3. Each activity has been performed for 120 s while quietly breathing. The existence of permanent characteristics within the considered data has been investigated in the performed tests thanks to the availability of two recording sessions, taken at an average time distance of 21 days, for the involved subjects.

The collected data have been used to learn the representations outlined in Section III. Specifically, since our aim is to test the recognition performance achievable in open-set verification scenarios, subject-specific binary classifiers have been trained using, for each user, data taken from the first recording session of the considered individual as positive class, and data taken from other 15 randomly-selected subjects, out of the available ones, as negative class. The recognition rates attainable through these subject-specific models have been estimated considering, for each user, the recordings from the second session as genuine verification probes, and data collected from the remaining 15 subjects, not involved during training, as impostor attempts. The proposed WISHNET$_{TF}$ and WISHNET$_{CE}$ networks have been trained using stochastic gradient descent with momentum (SGDM), a learning rate of 0.0005, and a batch size of 16. The networks trained with cross-entropy loss function to perform binary classifications are mentioned in the following as WISHNET$_{TF}^{CE}$ and WISHNET$_{CE}^{CE}$.

The representations relying on siamese training have been instead obtained by selecting, for each subject considered as a genuine user, other 15 subjects out of the available ones for training purposes, while a disjoint set with data from the remaining 15 subjects have been used to estimate the false acceptance rate (FAR). In more detail, the considered networks are trained exploiting only the data from the 15 randomly-selected subjects, with pairs whose distance has to be minimized defined by taking two frames extracted from recordings of different sessions of the same subject, and pairs whose distance has to be maximized defined by taking two frames from recordings of different subjects. Following this approach, the networks should explicitly extract permanent characteristics from the considered data, and the learned representations can be thus used also for subjects not involved during the training stage. Network training has been performed using SGDM, a batch size of 64 pairs, and a 0.0001 learning rate. The feature extractors relying on siamese strategy and contrastive loss are reported in the following as WISHNET$_{TF}^{CE}$ and WISHNET$_{CE}^{CE}$.

### B. Comparison With State-of-The-Art Approaches

The recognition results obtained when considering SCG or GCG signals made of a single frame, therefore lasting $L = 5$ s, as verification probe, are reported in terms of EERs in Table IV. The results are referred to comparisons of recordings captured during different acquisition sessions, yet considering the same activity and the same sensor location in both enrolment and verification stages. Cross-validation tests have been conducted changing, at each iteration, the sets of 15 subjects employed to learn the adopted representations. Furthermore, for each considered activity and subject, $S = 90$ s from the first recording session have been randomly extracted from the available recordings and employed as enrolment data.

In order to perform a comparative analysis between the recognition rates achievable using the proposed approaches, and those attainable resorting to methods already proposed in literature, we have implemented the approaches described in [44], [45], [46], and [47], and applied them to the collected multi-session database. In more detail, since all the aforementioned approaches have been designed for CS conditions, with the only exception of [45], some modifications have been performed on the original approaches in order to apply them to the considered OS verification scenario. For the approach in [44], the background GMM model is estimated for each individual using data from other 15 random subjects, with the FAR then estimated using data from a disjoint set of individuals. As for the works in [46] and [47], the employed ResNet50 and SVM have been trained for binary classification, similarly to what has been done with the proposed dictionary learning methods, and the proposed networks trained with a CE loss function. The parameters characterizing the considered approaches, such as the number of Gaussians used in the GMM model of [44], have been empirically selected among a set of possible values, with the aim of maximizing the achievable recognition performance.

From the reported results it can be observed, first of all, that permanent characteristics can be actually derived from SCG and GCG signals of limited duration, and exploited to recognize a subject during acquisition sessions carried out some time after an enrolment. Yet, as expected, the obtained recognition rates are notably worse than those reported in [45], [46], [47], where single-session databases have been considered. SCG data commonly allow achieving better performance than GCG ones. However, these latter attain rates comparable, if not better, to SCG in several configuration, testifying the usefulness of exploiting these inertial measurements in addition to SCG, a possibility till now suggested only in [48].

The obtained results also show that the methods here proposed outperform the current state-of-the-art approaches in most of the considered scenarios. The results reported in bold in Table IV correspond to performance with a significant (at a 5% type I error) improvement over state-of-the-art methods, that is, the mean EERs of these latter are outside the 95% confidence intervals of the best-performing approach. The networks here proposed in Tables II and III commonly guarantee the best recognition rates among the considered strategies. In more detail, it can be noticed that training the employed networks with a siamese strategy is preferable on most occasions, with respect to the usage of binary cross-entropy learning. A further advantage of adopting a siamese strategy to train the proposed networks is also given by the time efficiency this approach guarantees when carrying out the enrolment of a subject. Table V reports an analysis on the computational complexity, expressed in terms of time required.
to perform the enrolment and verification stages associated with the considered processing methods, using a desktop equipped with an Intel i7 3.30 GHz CPU, 48 GB RAM and a Nvidia TITAN V GPU. As mentioned in Section III-C, adopting a siamese learning strategy allows to train a single network valid for multiple subjects, instead of training a dedicated network for each subject as done with binary classifiers. An enrolment procedure would therefore consist simply in the propagation of the collected frames through the network, without the need for further training. The task complexity is therefore moved at system level, without affecting the enrolment of each user, during which the trained networks are simply used as feature extractors. For the sake of completeness, the time required to train the proposed WISHNET\(_{CT}\) and WISHNET\(_{CF}\) networks in a siamese strategy are 240 s and 210 s, respectively. The best results are achieved for subjects in lying conditions, while the

| TABLE IV | RECOGNITION PERFORMANCE, IN TERMS OF EER (MEAN AND 95% CONFIDENCE INTERVAL, IN %), FOR A SINGLE FRAME USED AS VERIFICATION PROBE |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Signal Pos  | GMM \([44]\) | Correlation ResNet50 \([46]\) | SVM \([47]\) | LC-KVSD | SRC | WISHNET\(_{CT}\) P | WISHNET\(_{CT}\) M | WISHNET\(_{CF}\) P | WISHNET\(_{CF}\) M |
| P | 13.0 ± 1.0 | 16.5 ± 1.3 | 23.3 ± 1.9 | 10.4 ± 1.0 | 11.1 ± 0.9 | 8.3 ± 0.8 | 10.2 ± 0.9 | 8.1 ± 0.7 | 11.2 ± 0.9 | 14.8 ± 1.5 |
| A | 17.5 ± 1.2 | 21.5 ± 1.9 | 25.2 ± 2.1 | 10.4 ± 1.3 | 11.9 ± 1.1 | 10.4 ± 1.0 | 11.6 ± 1.1 | 7.8 ± 0.7 | 14.3 ± 1.2 | 10.7 ± 1.0 |
| B | 14.6 ± 1.2 | 20.8 ± 2.2 | 22.1 ± 2.1 | 7.5 ± 0.8 | 11.4 ± 0.9 | 7.7 ± 0.8 | 12.6 ± 1.0 | 8.7 ± 0.8 | 11.2 ± 0.9 | 17.5 ± 1.2 |
| T | 13.5 ± 1.1 | 17.1 ± 1.3 | 23.6 ± 1.9 | 8.1 ± 0.9 | 7.3 ± 0.9 | 5.6 ± 0.7 | 8.1 ± 0.8 | 7.2 ± 0.7 | 11.7 ± 1.0 | 12.6 ± 1.1 |

| TABLE V | COMPUTATIONAL COMPLEXITY, IN TERMS OF TIME REQUIRED TO PERFORM THE ENROLMENT AND VERIFICATION STAGES, ASSOCIATED WITH THE CONSIDERED PROCESSING METHODS (MEAN AND 95% CONFIDENCE INTERVAL) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Phase           | GMM \([44]\) | Correlation | ResNet50 \([46]\) | SVM \([47]\) | LC-KVSD | SRC | WISHNET\(_{CT}\) P | WISHNET\(_{CT}\) M | WISHNET\(_{CF}\) P | WISHNET\(_{CF}\) M |
| Enrolment       | 37 ± 8 ms | 37 ± 8 ms | 273 ± 12 s | 185 ± 15 ms | 63 ± 12 ms | 3.4 ± 1 s | 221 ± 10 s | 12 ± 2 ms | 157 ± 6 s | 13 ± 3 ms |
| Verification    | 0.4 ± 0.2 ms | 0.7 ± 0.3 ms | 2.8 ± 0.7 ms | 0.4 ± 0.2 ms | 0.6 ± 0.3 ms | 1.1 ± 0.5 ms | 0.2 ± 0.1 ms | 0.2 ± 0.1 ms | 0.1 ± 0.1 ms | 0.2 ± 0.1 ms |

Values in bold refer to results better than state of the art in a statistically significant sense \((p < 0.05)\).
best sensor positions are the ones located at the left side of the chest, that is, at the pulmonary and the mitral valves.

C. Exploiting Multi-Biometric Approaches

Even though the results reported in Table IV are comparable with those achieved with other heart-based biometric traits when evaluated over multi-session databases, as shown by the summary in Table I, the achieved EERs are not good enough to justify the adoption in real-world scenario, regardless of the specific processing method employed. Nevertheless, the availability of several effective approaches, and of multiple sources of information, could be exploited to improve the attainable recognition performance by resorting to multi-biometric approaches. In more detail, we have performed tests by combining, through a sum operator, the scores provided by the proposed dictionary learning approaches, and those obtained through the networks here proposed and trained with a siamese strategy. The obtained results are shown in Table VI, where also the recognition rates achieved when jointly exploiting SCG and GCG data are reported. The notable performance improvement achieved through multi-biometric approaches verify the usefulness of designing multiple processing techniques to deal with data having significant intra-class variability, and the importance of investigating the discriminative capabilities of GCG signals, in addition to those of SCG, an aspect that has been still neglected, if not in [48]. Another major motivation supporting the joint exploitation of SCG and GCG data derives from the fact that both signals can be easily captured through a single device, as for the commercial product used in our tests.

Further improvements in recognition rates can be accomplished considering the availability of more than a single frame during verification. Fig. 4 shows the recognition performance achievable for increasing durations of the probes. The reported EERs have been obtained taking the average of the scores obtained for consecutive frames, and considering signals collected by a sensor placed at the pulmonary valve, that is, the position guaranteeing the best verification results as shown in Table VI. The obtained results demonstrate that EERs less than 5% can be achieved when exploiting recordings lasting 20 s and collected while carrying out all the considered activities, with very low error rates obtained especially for subjects in lying conditions. The use of IMUs to capture heart-induced vibrations on a subject’s chest is therefore an effective solution to recognize a user even several days after the initial enrolment.

D. Dealing With Multiple Activities

The results so far presented have been obtained when comparing signals captured from subjects performing the same activity during both enrolment and verification. However, as already observed in [48], it would be desirable to have the ability to recognize a person without having to know a priori the performed activity. Unfortunately, as shown in [48], the signals collected during distinct activities are significantly different, making it hard to recognize a person using SCG or GCG collected in conditions different than the ones carried out during the enrolment. In order to handle such differences, in [48] it has been suggested to train classifiers using data collected while performing all the admissible activities. Although such approach is feasible, following this strategy implies a significant deterioration of the achievable recognition performance, with respect to scenarios focused on a specific activity [48]. In order to preserve the capability of attaining high recognition rates, it is here proposed to perform a different approach relying on a two-step user recognition process, during which the activity a user is performing has to be first determined, and then a verification model specifically learned for that condition is employed. Actually, human activity recognition (HAR) tasks have long been studied, and the results attained in several literature works, exploiting inertial measurements for such aim, suggest that such approach could be actually feasible and effective.

In order to verify such hypothesis when exploiting data captured by a single wearable device placed on a subject’s chest,

---

**TABLE VI**

| Signal | Pos. | Activity | Lying | Sitting | Standing | Walking |
|--------|------|----------|-------|---------|----------|---------|
| SCG    | P    | 4.1±0.5  | 6.4±0.7 | 5.1±0.6 | 3.2±0.4  |         |
|        | A    | 6.1±0.7  | 7.2±0.8 | 8.2±0.9 | 3.2±0.4  |         |
|        | X    | 4.9±0.6  | 5.4±0.6 | 7.3±0.8 | 4.6±0.6  |         |
|        | M    | 3.1±0.5  | 5.7±0.7 | 5.1±0.7 | 2.8±0.3  |         |
|        | T    | 4.8±0.6  | 4.3±0.5 | 6.0±0.8 | 3.4±0.4  |         |
| GCG    | P    | 4.5±0.6  | 12.1±1.2 | 12.1±1.6 | 8.4±0.9  |         |
|        | A    | 7.0±0.8  | 13.6±1.4 | 5.5±0.8 | 7.4±0.8  |         |
|        | X    | 8.1±0.9  | 11.1±1.2 | 10.9±1.2 | 7.4±0.8  |         |
|        | M    | 4.2±0.5  | 3.4±0.4 | 4.9±0.7 | 8.1±0.8  |         |
|        | T    | 3.5±0.4  | 9.4±1.0 | 5.5±0.8 | 4.2±0.5  |         |
| SCG+GCG| P   | 1.3±0.3  | 6.1±0.8  | 4.9±0.6 | 3.0±0.5  |         |
|        | A   | 4.3±0.5  | 6.9±0.7  | 3.6±0.4 | 2.4±0.4  |         |
|        | X   | 3.1±0.5  | 5.2±0.7  | 5.5±0.7 | 2.2±0.4  |         |
|        | M   | 1.9±0.4  | 2.9±0.5  | 2.3±0.4 | 2.5±0.4  |         |
|        | T   | 2.3±0.4  | 3.9±0.6  | 3.1±0.5 | 1.8±0.4  |         |

Best results for each activity in bold.

---

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
TABLE VII
OVERALL ACCURACY (MEAN AND 95% CONFIDENCE INTERVAL, IN %) OF THE HAR TASK, USING A SINGLE FRAME AS VERIFICATION PROBE

| Sign. Pos. | SVM | ResNet50 | LSTM | SRC | WISHNET$_{CF}$ | WISHNET$_{TF}$ |
|------------|-----|----------|------|-----|----------------|----------------|
| P          | 87.6±0.8 | 86.8±0.8 | 77.7±1.3 | 85.9±0.9 | 86.9±0.9 | 85.7±0.9 |
| A          | 85.1±0.8 | 96.0±0.4 | 80.0±1.1 | 86.5±0.8 | 87.5±0.8 | 86.8±0.9 |
| M          | 95.1±0.4 | 95.7±0.4 | 80.0±1.0 | 96.1±0.3 | 95.9±0.4 | 89.7±0.8 |
| T          | 87.8±0.7 | 92.7±0.7 | 84.7±1.0 | 91.0±0.6 | 91.1±0.6 | 85.2±0.7 |

TABLE VIII
OVERALL ACCURACY (MEAN AND 95% CONFIDENCE INTERVAL, IN %) OF THE HAR TASK, USING A SINGLE FRAME AS VERIFICATION PROBE AND JOINTLY EXPLOITING MULTIPLE REPRESENTATIONS

| Signal       | Position       | Pulmonary Aortic Xiphoid Mitral Tricuspid |
|--------------|----------------|------------------------------------------|
| SCG          | 89.1±0.8       | 96.2±0.4 90.0±0.7 96.5±0.4 93.8±0.5     |
| GCC          | 85.8±0.9       | 79.4±1.2 82.1±1.0 78.0±1.5 88.2±0.8     |
| SCG+GCC      | 93.7±0.4       | 96.7±0.4 91.5±0.6 97.0±0.3 96.9±0.3     |

Table VIII reports the classification accuracies achievable when combining the scores obtained using classifiers based on ResNet50, SRC, and the proposed WISHNET$_{CF}$ network. Following this approach, very accurate HAR can be accomplished for sensors placed at all the considered positions, especially when combining the information derived from both SCG and GCG data. As for biometric purposes, the recognition capabilities could be further enhanced by considering recognition probes longer than 5 s. Fig. 5 shows the activity classification accuracy achievable by taking decisions relying on the mean of the scores computed for consecutive frames, considering signals recorded at the pulmonary valve. The reported plots show that good recognition accuracy can be achieved when exploiting SCG and GCG probes with enough duration, thus allowing to select the suited model to perform biometric recognition.

In order to evaluate the reliability of the obtained results, we have also performed additional tests on the public PAMAP2 database [65], which comprises SCG and GCG data captured from a device placed on the chest of nine subjects performing multiple ADL during a single session. Given the low number of available subjects, a leave-one-out strategy has been employed to divide the treated data into disjoint training and tests subsets. Leveraging on the combined use of SCG and GCG signals lasting 5 s, as for the tests whose results are reported in Table VIII, and considering the same four ADL covered in our database, performance analogous to the ones achieved on our data have been obtained, with a mean accuracy of 93.1%, despite the availability of a much smaller number of subjects upon which the training can be conducted. It is yet worth observing that the accuracy of a HAR task is affected by the number of admissible ADL that a subject is expected to perform. The data in the PAMAP2 database have been used to investigate this aspect, performing tests considering four activities (cycling, nordic walking, ascending and descending stairs) in addition to the four ones in our data, and also other four additional activities (running, rope jumping, vacuum cleaning, ironing) for a total of 12 ADL. A performance worsening, as expected, is observed in these cases, with an accuracy decrease to 88.2% when considering 8 activities, and to 85.4% for 12 ADL. The design of techniques able to carry out highly efficient SCG- and GCG-based HAR in scenarios with a high number of possible ADL is therefore a research topic that could bring relevant benefits for biometric recognition applications too. About this, it is worth remarking that HAR can be performed much more efficiently when using accelerometer data, with respect to alternatives comprising signals such as ECG or PPG [66], an advantage that should

we have performed tests on the collected database comprising data captured during four distinct ADL, and using processing techniques proposed in literature for HAR purposes [63], [64], such as SVM, transfer learning with consolidated networks such as ResNet50, and LSTM networks. We have also applied to HAR the approaches here employed for biometric recognition, that is, SRC, WISHNET$_{CF}$, and WISHNET$_{TF}$. For the sake of compactness, the results obtained with LC-KSVD have been omitted, being worse than what is obtained with SRC. As in standard multi-class problems, the considered networks have been trained with a CE loss function. Cross-validation has been carried out by randomly selecting, at each of 5 iterations, 20 subjects out of the available ones to collect the data upon which the training process is carried out, with samples taken from the remaining subjects used as testing probes.

The obtained results, expressed in terms of overall accuracy, are shown in Table VII for both SCG and GCG data, and sensors placed at different locations. Despite being designed for biometric recognition, the WISHNET$_{CF}$ network achieves fairly good results, especially when considering SCG data captured at the mitral valve. Nevertheless, it is evident that the joint exploitation of multiple techniques is necessary to attain performance that make the proposed two-step biometric recognition process effective. Errors in the activity recognition task could in fact imply severe degradation of the overall verification performance.
further support the interest in mechanical measurements of the heart activity for biometric recognition purposes.

E. Explainability Analysis

In order to provide a comprehensive analysis of the considered SCG and GCG biometric traits, we have used XAI techniques [67] to derive some insights regarding aspects exploited by the employed approaches when taking decisions for both biometric people recognition and activity recognition. Within the context of biometric recognition systems, most of the current contributions on explainability have used post hoc approaches, whose aim is to provide visual explanations to the predictions of a trained model [68]. Adopting this approach in the considered scenario, we have here employed Guided Gradient-weighted Class Activation Mapping (Grad-CAM) [69] to analyze the decisions of the employed networks.

The conducted evaluation has been threefold, that is, we have performed an analysis in the time, frequency, and spatial domains to characterize the input contributions to the taken predictions. The purpose of the time analysis is to investigate whether distinct parts of a cardiac cycle could have different influences on the performed predictions. To this aim, we have applied Guided Grad-CAM to the proposed network WISHNET\textsubscript{CE}, when used for both biometric recognition and activity recognition purposes, separately evaluating each considered ADL. The outcomes obtained from different samples have been averaged, after having segmented each heartbeat depending on the peaks in the SCG signal, corresponding to aortic openings, and normalized the activity maps to have values in the range
Fig. 8. Explainability analysis in the frequency (left) and spatial (right) domains, for the biometric recognition task.

Fig. 9. Explainability analysis in the frequency (left) and spatial (right) domains, for the HAR task.

[0;1] for each contribution. Fig. 6 shows the results on SCG and GCG data for biometric recognition, while Fig. 7 is referred to HAR. In addition to show the obtained behaviors, the figures also report the standard deviation of the normalized significance values computed across a cardiac cycle. This parameter can be used to evaluate the existence of portions with a relevance significantly greater than the others: the lower the standard deviation, the more homogeneous the contribution of different portions to the final prediction. The behaviors observed in all the considered conditions seem suggesting that, when performing biometric recognition, information from the whole cardiac cycle is exploited to perform a prediction, while specific portions may have an impact greater than the others when carrying out HAR. This conclusion is coherent with what has been observed in [70], where it has been shown that all the components of a heartbeat should be taken into account to reliably perform biometric recognition using ECG signals. It can be also observed that the contributions in SCG signals are typically more homogeneous than those derived from GCG data.

The analysis on the frequency and spatial contents have been conducted applying Guided Grad-CAM to the proposed network WISHNET<sub>CE</sub>. The outcomes obtained from different samples have been averaged for each of the 25 frequency bins available in W, and for each of the three axis of the considered inertial measurements, after having normalized to the [0; 1] range the activity maps of each contribution. Fig. 8 shows the results on SCG and GCG data for biometric recognition, while Fig. 9 is referred to HAR. From the obtained results it can be seen that high frequency contents are less important than contributions at lower frequencies for static conditions such as lying, standing, and sitting. When the performed activity involves movements, as in walking conditions, the high frequency contents is significantly relevant, suggesting that the used IMUs retain characteristics related to the subject behavior, in addition to information associated to the cardiac activity. The analysis on the spatial domain shows that the z-axis is the component providing more information in both the considered tasks, even though contributions from all the three components are relevant to take decisions, especially for biometric recognition.

V. CONCLUSION

In this paper, a biometric recognition approach relying on wearable inertial sensors to detect mechanical heart activity, namely BIOWISH, has been proposed and tested. Several representations have been considered to extract discriminative information from the considered SCG and GCG signals, resorting to traditional machine learning algorithms such as those relying on dictionary learning, as well as introducing novel networks to process data in both time or time/frequency domains. The proposed systems have been tested considering recordings collected through commercial and inexpensive wearable IMUs from subjects carrying out four different activities, and placing sensors at different positions to evaluate the configuration guaranteeing the best verification performance.

It has been observed that the networks here proposed are able to achieve good verification performance, especially when a siamese strategy is adopted for training. The joint use of both SCG and GCG data, and the exploitation of recordings lasting approximately 20 s, allow to reach rather high recognition performance, especially for subjects in lying conditions and data acquired through sensors placed in correspondence to the pulmonary valve. In more detail, the achieved OS verification rates are the best ones reported in literature, for biometric systems based on heart-related traits acquired through wearable devices, when considering tests carried out on multisession databases.

A two-stage process, requiring to determine the carried out activity before extracting person-specific characteristics, has been also proposed to perform automatic people recognition without...
having to know a priori the performed activity. Also in this case, the joint usage of SCG and GCG data proved to be essential to achieve high performance, even though further improvements may be required for effectively handling a significant number of distinct activities. An explainability analysis has also been carried out on the considered heart-based inertial measurements, revealing some insights regarding aspects taken into account by the proposed neural networks when performing their predictions for biometric people recognition and activity recognition tasks. The obtained results testify that heart-induced vibrations, measured through IMU wearable devices placed on a subject’s chest, could represent a reliable biometric identifier to be exploited in automatic verification systems.

REFERENCES

[1] A. Ometov et al., “A survey on wearable technology: History, state-of-the-art and current challenges,” Comput. Netw., vol. 193, 2021, Art. no. 108074.

[10] C. Ye, B. V. Kumar, and M. T. Coimbra, “Human identification based on chest, could represent a reliable biometric identifier to be explored in automatic verification systems.

[21] T. J. Taji, “Gyrocardiography: A new non-invasive monitoring method for the assessment of cardiac mechanics and the estimation of hemodynamic variables,” Sci. Rep., vol. 7, 2017, Art. no. 6823.

[2] J. Hebert, B. Lewis, H. Cai, K. Venkatasubramanian, M. Provost, and K. Charlebois, “Ballistocardiogram-based authentication using convolutional neural networks,” 2018, arXiv:1807.03216.

[3] E. Piciucco, E. Di Lascio, E. Maiorana, S. Santini, and P. Campisi, “Biometric authentication using wearable devices: A survey,” ACM Trans. Intell. Syst. Tech., vol. 19, no. 10, pp. 3569–3580, May 2019.

[4] A. Pouryayevali, S. Hari, and D. Hatzinakos, “Pattern mining approach used in sensor-based biometric recognition: A review,” IEEE Sensors J., vol. 19, no. 10, pp. 3569–3580, May 2019.

[5] K. Revett and S. T. de Magalhães, “Cognitive biometrics: Challenges for the future,” in Proc. Int. Conf. Geometric Secur. Saf. Sustainability, 2010, pp. 79–86.

[6] A. Sundararajan, A. Sarwat, and A. Pons, “A survey on modality characteristics, performance evaluation metrics, and security for traditional and wearable biometric systems,” ACM Comput. Surv., vol. 52, no. 2, pp. 1–36, 2019.

[7] S. Khan, S. Parkinson, L. Grant, N. Liu, and S. McGuire, “Biometric systems utilising health data from wearable devices: Applications and future challenges in computer security,” ACM Comput. Sur., vol. 53, no. 4, pp. 1–29, 2020.

[8] E. Maiorana, “A survey on biometric recognition using wearable devices,” Pattern Recognit., vol. 156, pp. 29–37, 2022.

[9] C. Ye, B. V. Kumar, and M. T. Coimbra, “Human identification based on ECG signals from wearable health monitoring devices,” in Proc. 4th Int. Symp. Appl. Sci. Biomed. Commun. Technol., 2011, Art. no. 25.

[10] J. Chauhan, H. Nguyen, C. Wang, and B. Johnson, “I think, therefore I am: Usability and security of authentication using brainwaves,” in Proc. Int. Conf. Financial Cryptography Data Secur., 2013, pp. 1–16.

[11] J. Chauhan, Y. Hu, S. Seneviratne, M. M. Al-Ameen, M. A. Seneviratne, and Y. Lee, “BreathPrint: Breathing-acoustics-based user authentication,” in Proc. 15th Annu. Int. Conf. Mobile Syst. Appl. Serv., 2017, pp. 278–291.

[12] K. Suzuki, K. Shimizu, and K. Ognuchi, “Feasible personal identification by eye blinking using wear able device,” in Proc. IEEE 15th Int. Colloq. Signal Process. Appl., 2019, pp. 266–269.

[13] E. Picuicco, E. Di Lascio, E. Maiorana, S. Santini, and P. Campisi, “Biometric recognition using wearable devices in real-life settings,” Pattern Recognit. Lett., vol. 146, pp. 260–266, 2021.

[14] F. Agrafioti, F. Bui, and D. Hatzinakos, “Secure telemedicine: Biometrics for remote and continuous patient verification,” J. Comput. Netw. Commun., vol. 2012, 2012, Art. no. 924791.

[15] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, “On evaluating ECG biometric systems: Session-dependence and body posture,” IEEE Trans. Inf. Forensics Secur., vol. 9, no. 11, pp. 2002–2013, Nov. 2014.

[16] A. Rathore, Z. Li, W. Zhu, Z. Jin, and Z. Jin, “A survey on heart biometries,” ACM Comput. Surv., vol. 53, no. 6, 2020, Art. no. 144.

[17] J. Sancho, A. Alesanco, and J. García, “Biometric authentication using the PPG: A long-term feasibility study,” Sensors, vol. 18, no. 5, 2018, Art. no. 1525.

[18] J. Blasco, T. M. Chen, J. Tapiador, and P. Peris-Lopez, “A survey of wearable biometric recognition systems,” ACM Comput. Surv., vol. 49, no. 3, pp. 1–35, 2016.

[19] M. Bugdol and A. Mitas, “Multimodal biometric system combining ECG and sound signals,” Pattern Recognit. Lett., vol. 38, pp. 107–112, 2014.

[20] D. Rai, H. Thakkar, S. Rajput, J. Santamaria, C. Bhatt, and F. Roca, “A comprehensive review on seismocardiogram: Current advancements on acquisition, annotation, and applications,” Mathematics, vol. 9, 2021, Art. no. 2243.

[21] T. J. Taji, “Gyrocardiography: A new non-invasive monitoring method for the assessment of cardiac mechanics and the estimation of hemodynamic variables,” Sci. Rep., vol. 7, 2017, Art. no. 6823.

[22] D. C. F. M. C. Romano and E. Schena, “Comparison between chest-worn accelerometer and gyroscope performance for heart rate and respiratory rate monitoring,” Biosensors, vol. 12, no. 10, 2022, Art. no. 834.

[23] M. Rahmani, R. Berkvens, and M. Weyn, “Chest-worn inertial sensors: A survey of applications and methods,” Sensors, vol. 21, no. 8, 2021, Art. no. 2875.

[24] F. Agrafioti, J. Gao, and D. Hatzinakos, “Heart biometrics: Theory, methods and applications,” in Biometrics, J. Yang Ed. London, U.K.: InTech, 2011.

[25] M. Martinho, A. Fred, and H. I. Alçado da Silva, “Towards continuous user recognition by exploring physiological multimodality: An electrocardiogram (ECG) and blood volume pulse (BVP) approach,” in Proc. Int. Symp. Sens. Instrum. IoT Era, 2018, pp. 1–6.

[26] D. Bai, J. Yang, D. Zhu, and X. Wang, “Chest-worn inertial sensors for multi-modal biometric verification,” Sensors, vol. 18, no. 9, 2018, Art. no. 2782.

[27] D. Hwang, B. Taha, D. Lee, and D. Hatzinakos, “Evaluation of the time acquisition, annotation, and applications,” in Proc. 15th Int. Conf. Digit. Signal Process., 2013, pp. 1–5.
A. Bui, Z. Yu, and F. Bui, “A biometric modality based on the seismocardiogram (SCG),” in Proc. Int. Conf. Workshop Comput. Commun., 2015, pp. 1–7.

P.-Y. Hsu, P.-H. Hsu, and H.-L. Liu, “Exploring seismocardiogram biometrics with wavelet transform,” in Proc. IEEE Int. Conf. Pattern Recognit., 2021, pp. 4450–4457.

P.-Y. Hsu, P.-H. Hsu, T.-H. Lee, and H.-L. Liu, “Motion artifact resilient SCG-based biometric authentication using machine learning,” in Proc. IEEE Eng. Med. Biol. Soc., 2021, pp. 144–147.

E. Maiorana and C. Massaroni, “Biometric recognition based on heart-induced chest vibrations,” in Proc. IEEE Int. Workshop Biometrics Forensics, 2021, pp. 1–6.

J. Hernandez, D. McDuff, and R. Picard, “Bioinsights: Extracting personal data from “still” wearable motion sensors,” in Proc. IEEE Int. Conf. Wearable Implantable Body Sensor Netws., 2015, pp. 1–6.

O. I. Ihan, “Ballistocardiography and seismocardiography: A review of recent advances,” IEEE J. Biomed. Health Inform., vol. 19, no. 4, pp. 1414–1427, Jul. 2015.

S. Vhaduri and C. Poellabauer, “Multi-modal biometric-based implicit authentication of wearable device users,” IEEE Trans. Inf. Forensics Secur., vol. 14, no. 12, pp. 3116–3125, Dec. 2019.

G. Retinas, P. Filtinis, N. Efthymiou, E. Theodosis, A. Zlatintsi, and P. Maragos, “Person identification using deep convolutional neural networks on short-term signals from wearable sensors,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2020, pp. 3657–3661.

E. Lee, A. Ho, Y. Wang, C. Huang, and C. Lee, “Cross-domain adaptation for biometric identification using photoplethysmogram,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2020, pp. 1289–1293.

A. Diaz Alonso, C. Travieso, and M. D. J. B. Alonso, “Biometric personal identification system using biomedical sensors,” in Proc. 2nd Int. Conf. Commun. Control Intell. Syst., 2016, pp. 104–109.

S. M. Mathews, “Dictionary and deep learning algorithms with applications to remote health monitoring systems,” Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Delaware, Newark, DE, 2017.

J. Xu, G. Yang, K. Wang, Y. Huang, H. Liu, and Y. Yin, “Structural sparse representation with class-specific dictionary for ECG biometric recognition,” Pattern Recognit. Lett., vol. 135, pp. 44–49, 2020.

Z. Jiang, Z. Lin, and L. Davis, “Learning a discriminative dictionary for sparse coding via label consistent K-SVD,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2011, pp. 1697–1704.

J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, Feb. 2009.

K. Sundararajan and D. Woodard, “Deep learning for biometrics: A survey,” ACM Comput. Surv., vol. 51, no. 3, Art. no. 65, 2019.

O. Russakovsky et al., “Imagenet large scale visual recognition challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.

G. Koch, R. Zemel, and R. Salakhutdinov, “Siamese neural networks for one-shot image recognition,” in Proc. ICML Deep Learn. Workshop, 2015, pp. 1–8.

E. D’Alcala’, J. Voerman, J. Konrath, and A. Vydhyvanathan, “Xsens DOt wearable sensor platform white paper,” White Paper, 2021.

E. Ramanujam, T. Perumal, and S. Thinnaguru, “Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review,” IEEE Sensors J., vol. 21, no. 12, Jun. 2021.

S. Mathews, C. Kambhamettu, and K. Barner, “Centralized class specific dictionary learning for wearable sensors based physical activity recognition,” in Proc. Annu. Conf. Inf. Sci. Syst., 2017, pp. 1–6.

A. Reiss and D. Stricker, “Introducing a new benchmarked dataset for activity monitoring,” in Proc. IEEE Int. Symp. Wearable Comput., 2012, pp. 108–109.

M. Afsal Arani, D. Costa, and E. Shihab, “Human activity recognition: A comparative study to assess the contribution level of accelerometer, ECG, and PPG signals,” Sensors, vol. 21, no. 21, 2021, Art. no. 6997.

A. Arrieta et al., “Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” Inf. Fusion, vol. 58, pp. 82–115, 2020.

P. Neto et al., “Explainable biometrics in the age of deep learning,” 2022, arXiv:2208.0550.

R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 618–626.

J. Pinto and J. Cardoso, “Explaining ECG biometrics: Is it all in the QRS?,” in Proc. Int. Conf. Biometrics Special Int. Group, 2020, pp. 1–5.