Efficient IoT Big Data Streaming with Deep Learning-enabled Dynamics

Junhua Wong, Vincenzo Piuri, *Fellow, IEEE*, Fabio Scotti, *Senior Member, IEEE,* Qingxue Zhang, *Senior Member, IEEE*

***Abstract*— Internet of Medical Things (IoMT) is igniting many emerging smart health applications, by continuously streaming the big data for data-driven innovations. One critical obstacle in IoMT big data is the power hungriness of long-term data transmission. Targeting this challenge, we propose a novel framework called, IoMT Big-data Bayesian-backward Deep-encoder learning (IBBD), which mines deep autoencoder (AE) configurations for data sparsification and determines optimal trade-offs between information loss and power overhead. More specifically, the IBBD framework leverages an additional external Bayesian-backward loop that recommends AE configurations, on top of a traditional deep learning loop that executes and evaluate the AE quality. The IBBD recommendation is based on confidence to further minimize the regularized metrics that quantify the quality of AE configurations, and it further leverages regularization techniques to allow adjusting error-power tradeoffs in the mining process. We have conducted thorough experiments on a cardiac data streaming application and demonstrated the superiority of IBBD over the common practices like Discrete Wavelet Transform, and we have further generalized IBBD through validating the optimal AE configurations determined on one user to other users. This study is expected to greatly advance IoMT big data streaming practices towards precision medicine.**

***Index Terms*— Internet of Medical Things, Deep Learning, Data Mining, Regularization.**

I. Introduction

With advancements of electronics, wireless, and intelligent algorithms [1, 2], Internet of Medical Things (IoMT) [3, 4], with high efficiency, intelligence, reliability, connectivity, and more features, is attracting intensive interests for smart health applications [5-8]. With the potential to continuously stream human bio-dynamics to the cloud, it is expected that IoMT can greatly boost the big data-driven precision health practices.

IoMT, as illustrated in Fig. 1, is of great potential to capture and stream the bio-dynamics from multiple aspects, such as diverse bio-dynamics of the brain, cardiac, muscular, and skeletal systems. The big data captured can be streamed to the smart phone and/or cloud for further mining, thereby facilitating medical decision support, emergency alarm generation, lifestyle management and other data-driven smart health applications. Furthermore, the ubiquitous deployment of IoMT on the population-level practices is expected to bring even larger amounts of big data from the massive devices to the data mining facilities.

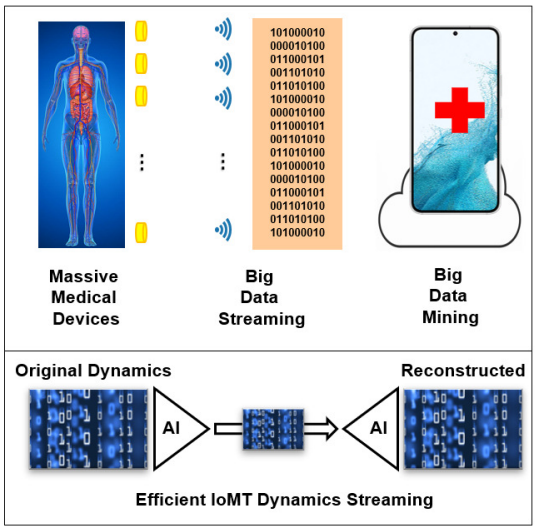

Fig. 1 An illustration of Internet of Medical Things (IoMT) for big data-driven precision health (top), and the deep learning-enabled dynamics sparsification for efficient big data streaming (bottom).

One critical obstacle of IoMT currently is the power hungriness of the IoMT system, introduced by the continuous data transmission from massive devices on each individual. Frequent charging IoMT is troublesome and impacting its deployment in people's daily lives. We in this study target this challenge and propose a new framework that leverages deep learning to sparsify the big IoMT data on the wearables and then reconstruct the bio-dynamics on the data mining facilities such as the smart phones and cloud centers, as shown in Fig. 1.

An IoMT device usually consists of a signal sensing module, a micro-controller module, a wireless module, and a power management module. It is well known that the wireless module often consumes a significant amount of power compared with other modules [9]. Especially, when IoMT is continuously streaming bio-dynamics, the device may only last for a few hours or even a shorter time. Motivated by this, we believe that minimizing the wireless transmission power is of great potential to maximize the active time of the IoMT system.

There have been previously reported studies on health monitor data compression [1]. Discrete Wavelet Transformation (DWT) has been a common practice in many studies [10-13], which firstly transforms the original signal to the time-frequency domain, and then selects out significant wavelet coefficients for transmission. The coefficients are used on the

Junhua Wong is with Purdue University School of Engineering and Technology, Indianapolis, IN46202, USA. E-mail: junhuaw@purdue.edu.
Vincenzo Piuri is with Department of Computer Science, University of Milan, Italy. E-mail: vincenzo.piuri@unimi.it.
Fabio Scotti is with Department of Computer Science, University of Milan, Italy. E-mail: fabio.scotti@unimi.it.
Qingxue Zhang is with Purdue University School of Engineering and Technology, Indianapolis, IN46202, USA. E-mail: qxzhang@purdue.edu.

receiver like a phone or cloud to reconstruct the original data. Lu *et. al* reported an Electrocardiogram (ECG) compression approach with DWT based on set partitioning in a hierarchical-trees algorithm. Sahambi *et. al* developed DWT-based ECG digital signal characterization system. Jha *et. al* [14] applied DWT after empirical decomposition for two-stage ECG signal compression. DWT has also been applied to other IoMT modalities, such as the Electroencephalogram [1, 15] and Electromyography [16].

Compressed Sensing (CS) is another common practice in various studies [13, 17-19], which uses a conversion matrix to transform the original signal to a new domain with sparse representations. On the receiver, the signal is reconstructed using the conversion matrix. Discrete Cosine Transform (DCT) has also been used in some studies [20-22], which applies the cosine waves as basic functions to perform signal projection and critical coefficient extraction.

These methods reported have noticeable impact on the area, and at the same time substantial efforts on furthering the field are still urged. More specifically, how can we learn the complex dynamics in the data to facilitate the data sparsification? Machine learning methods, therefore, are attracting more and more attentions by leveraging complex critical pattern mining for data compression. Qian *et. al* [2] reported a Kmeans-based dictionary approach to compress the ECG data. The deep autoencoder (AE) was reported to compress the ECG signal in [23]. Abdellatif *et. al* [24] applied AE to compress the EEG signal for wireless power reduction. The transfer learning has also been combined to AE for multimodal data compression in [25]. Wei *et. al* [26] reported a hybrid method based on both compressed sensing and deep learning for privacy-aware data transmission.

Interesting advancements have been made with these previous studies on machine learning and the most recent deep learning-based compression methods. Our research, however, targets a great challenge of deep learning-based intelligent compression in the field. More specifically, we will thoroughly investigate how the deep AE architecture design impacts the compression-induced power overhead and information loss. Further, we will investigate algorithms to determine optimal AE configurations to trade-off between the information loss and power overhead, thereby enabling error-power co-optimization. Our research will advance both theoretical and practical understanding of energy efficient IoMT.

We in this study propose a novel framework called, IoMT Big-data Bayesian-backward Deep-encoder Learning (IBBD), which aims to, by leveraging two levels of backward optimization, mine the AE configurations and determine the optimal error-power trade-offs. More specifically, the IBBD framework leverages an additional external Bayesian-backward [27, 28] loop, on top of a traditional deep AE learning loop that is based on the backward propagation of error gradients. As illustrated in Fig. 2, the proposed Bayesian-backward learning, based on the metrics of the current and previous AE configurations, generates the AE configuration recommendation for the next iteration of evaluation. The recommendation is based on confidence to further minimize the

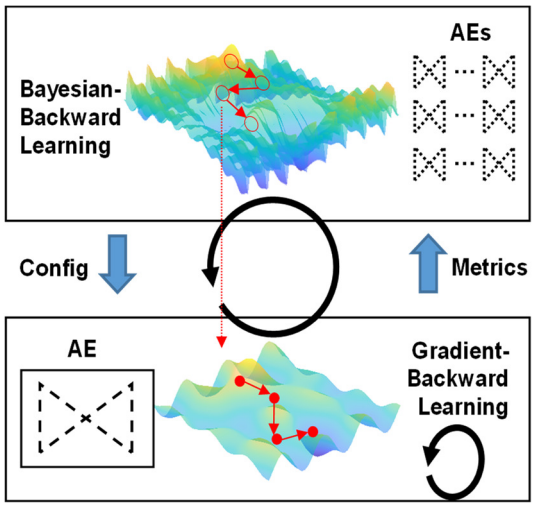

Fig. 2 The proposed novel framework called, IoMT Big-data Bayesian-backward Deep-encoder Learning (IBBD).
Note. IBBD recommends AE architectures with Bayesian learning (top), and executes evaluation with gradient-backward learning (bottom). Each red circle indicates an AE configuration; and each red dot indicates the learning trajectory in the configuration-constrained solution space.

regularized metrics that quantify the quality of AE configurations. The gradient-backward learning trains the configured AE architecture, and the metrics are afterwards generated that include information loss in terms of the signal reconstruction error, and the AE power consumption. Each red circle on top in Fig. 2, indicates an AE configuration that is expected to further minimize AE evaluation metrics; and each red dot in the bottom indicates the deep neural network learning trajectory in the configuration-constrained solution space.

The proposed IBBD framework enables multiple AE optimization tasks, such as best-effort-on-error optimization, best-effort-on-power optimization, and more importantly, optimal error-power co-optimization. We have conducted thorough experiments on ECG big data compression towards IoMT-empowered precision health, and demonstrated the effectiveness of the proposed framework for efficient IoMT big data streaming with the optimal deep learning of sparsification.

Our contributions are summarized as below:

(1) Proposing a novel IBBD framework that recommends AE configurations for evaluation with the top-level of optimization, and executes the recommendation with the bottom-level of AE learning;

(2) Designing a regularized Bayesian-backward deep encoding learning approach, to quantify the quality of AE configurations by combing metrics including information loss and power overhead, thereby enabling error-power co-optimization;

(3) Generalizing the proposed IBBD framework through validating the optimal AE configurations determined on one user to other users;

(4) Conducting experiments on ECG big data compression and streaming, and demonstrating that the optimal AEs are superior to the common practices like DWT, error-minimum AE and power-minimum AE.

This study therefore can greatly advance efficient IoMT big data streaming with optimal deep learning of sparsification, and can be generalized from ECG to other health modalities.

II. Approaches

In this section we detail the proposed IBBD framework and theories of the algorithms, as well as thorough evaluations and comparisons for demonstrating effectiveness of IBBD.

A. Algorithmic Framework Architecture

As shown in Fig. 2, the top-level optimization in IBBD leverages the Bayesian-backward learning to iteratively recommend AE configurations to minimize the regularized AE cost function. The bottom-level executes and evaluates the AE configurations, yielding the metrics including the reconstruction error and encoding power overhead.

B. Fundamentals of Deep Nonlinear Encoding

As an unsupervised learning technique, AE firstly learns efficient data representations through an encoding function, and then reconstructs the input-referred output with a decoding function. As defined in (1), the AE learning process is minimizing the difference between the input and the reconstructed output, where ϕ and ψ correspond to the functions of the nonlinear encoder and decoder, respectively, χ_i^n is the $i-th$ sample in the $n-th$ instance of the training database, L and N give the length of the time-series and the total number of instances, respectively, and learning of ϕ and ψ is executed simultaneously.

$$\phi^{\#}, \psi^{\#} = arg \min_{\phi,\psi} \left(\sum_{n=1}^{N} \sqrt{\frac{\sum_{i=1}^{L} (\chi_i^n - (\phi \cdot \psi)\chi_i^n)^2}{L}} / N \right) \quad (1)$$

We here have leveraged the Root Mean Square Error (RMSE) in (1) to evaluate the reconstruction information loss. With minimizing this loss, the AE learning process finally results in two functions $\phi^{\#}$ and $\psi^{\#}$. These two functions are also coupled with the AE configuration $\boldsymbol{\Omega}$, which is a set of D parameters that determines the AE architecture and learning process.

For instance, one of the critical parameters is the depth of the nonlinear encoding function. Therefore, the encoder is a nested function as (2), where $\theta(\boldsymbol{\Omega})$ determines the AE architecture and learning strategy based on the configuration $\boldsymbol{\Omega}$, and $\phi(\theta)$ optimizes the neural connection weights based on the given AE architecture and learning strategy. Similarly, the decoder is also a nested function as (3) with the AE architecture and learning strategy given by the function $\vartheta(\boldsymbol{\Omega})$.

$$\phi = \phi(\theta(\boldsymbol{\Omega})), \boldsymbol{\Omega} = \{\omega_d | d = 1, \dots, D\} \quad (2)$$

$$\psi = \psi(\vartheta(\boldsymbol{\Omega})), \boldsymbol{\Omega} = \{\omega_d | d = 1, \dots, D\} \quad (3)$$

Therefore, it is essential to find out the optimal configuration $\boldsymbol{\Omega}$ to achieve maximum reduction of the information loss, which cannot be made by the AE learning process itself. It means, there is a hyperspace, as indicated in Fig. 2, which provides more dimensionalities by including both the $\boldsymbol{\Omega} \rightarrow$ loss space and the consequent $\{\theta(\boldsymbol{\Omega}), \vartheta(\boldsymbol{\Omega})\} \rightarrow RMSE$ subspace. Next, we will detail the proposed Bayesian-backward deep-encoder learning framework to facilitate the hyperspace optimization for intelligent AE configuration determination.

C. IBBD: IoMT Big-data Bayesian-backward Deep-encoder Learning

The proposed novel framework, IBBD, aims to determine the optimal AE configuration through Bayesian-backward deep-encoder learning. Here we will detail how IBBD, through modelling the configuration and loss with Gaussian distributions, can recommend AE configuration at the top-level optimization. The embedding of the bottom-level AE learning and evaluation of the recommended configuration will also be detailed to indicate how two levels of IBBD cooperate with each other. Further, considering diverse power overhead for different AE configurations, the IBBD framework can regulate the loss including the reconstruction error and the power overhead, realizing error-power co-optimization.

We here model the $\boldsymbol{\Omega} \rightarrow \zeta$ relationship with a Bayesian linear regression function [27] as (4), where $f(\boldsymbol{\Omega})$ is the loss prediction function, ε is the noise, and ζ is the actual loss corresponding to the AE configuration $\boldsymbol{\Omega}$. The prediction function $f(\boldsymbol{\Omega})$ is modeled as, based on the probability theory and random process [27], a Gaussian function distribution as (5), where $\mu(\boldsymbol{\Omega})$ and $\sigma^2(\boldsymbol{\Omega})$ are the mean and variance of the function $f(\boldsymbol{\Omega})$ in a Gaussian process $\mathcal{GP}$, respectively.

$$\zeta = f(\boldsymbol{\Omega}) + \varepsilon \quad (4)$$

$$f(\boldsymbol{\Omega}) \sim \mathcal{GP}(\mu(\boldsymbol{\Omega}), \sigma^2(\boldsymbol{\Omega})) \quad (5)$$

$$\boldsymbol{H_t^{\zeta}} = \{\zeta_1, \zeta_2, \dots, \zeta_t\} \quad (6)$$

$$\boldsymbol{H_t^{\Omega}} = \{\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \dots, \boldsymbol{\Omega}_t\} \quad (7)$$

$$\boldsymbol{H}_t = \{(\boldsymbol{\Omega}_1, \zeta_1), (\boldsymbol{\Omega}_2, \zeta_2), \dots, (\boldsymbol{\Omega}_t, \zeta_t)\} \quad (8)$$

The distribution of $f(\boldsymbol{\Omega})$ gives all possible functions to predict ζ from $\boldsymbol{\Omega}$, thereby statistically yielding the confidence of different AE configurations in terms of minimizing the AE loss. The optimal configuration is then recommended for AE execution and evaluation to generate the actual loss ζ, which serves as a new observation. Accumulatively, a set of observations $\boldsymbol{H_t^{\zeta}}$ as (6) can be established based on iterative recommendations $\boldsymbol{H_t^{\Omega}}$ as (7). The pairs of configuration and observation form a set $\boldsymbol{H}_t$ as (8), where t denotes the time or the iteration index.

In the context of efficient IoMT big data streaming, the power overhead of the encoding process is an important factor, and we want power reduction in wireless data communication to be significantly higher than the data sparsification overhead. Therefore, we further propose a regularized loss as (9), to co-consider the AE reconstruction error and the AE-decoder power overhead. The loss considers both $rRMSE$ and $P(\phi^{\#}(\theta(\boldsymbol{\Omega}))$, which correspond to the relative RMSE with the learned AE ($\phi^{\#}$ and $\psi^{\#}$ together), and the power consumption of the encoder ($\phi^{\#}$ only), respectively. $\boldsymbol{\chi^{DB}}$ is the testing database.

$$\zeta = \alpha * rRMSE\left(\phi^{\#}\left(\theta(\boldsymbol{\Omega})\right), \psi^{\#}\left(\vartheta(\boldsymbol{\Omega})\right), \boldsymbol{\chi^{DB}}\right) + \beta * P\left(\phi^{\#}\left(\theta(\boldsymbol{\Omega})\right)\right) \quad (9)$$

The propose regularized loss in (9) can therefore trade-off between the AE reconstruction error and the encoding power overhead. The factors α and λ are aiming to regularize the contribution of rRMSE and power, respectively, which are metrics that quantify the quality of AE configurations. The rRMSE, as (10), aims to normalize the RMSE in the loss, where $\boldsymbol{\chi^n}$ is the n-th time series in the database.

$$rRMSE\left(\phi^{\#}\left(\theta(\boldsymbol{\Omega})\right),\psi^{\#}\left(\vartheta(\boldsymbol{\Omega})\right),\boldsymbol{\chi^{DB}}\right) = \sum_{n=1}^{N}\frac{\sqrt{\frac{\sum_{i=1}^{L}\left(x_i^n-(\phi^{\#}\cdot\psi^{\#})x_i^n\right)}{L}}}{max(\boldsymbol{\chi^n})-min(\boldsymbol{\chi^n})}\Big/N \quad (10)$$

The Bayesian linear regression model [27] for AE architecture can be represented as (11), where $\upsilon(\cdot)$ is a basis function for space transformation on $\boldsymbol{\Omega}$, and $\boldsymbol{w}$ has a prior Gaussian distribution as (12) with a zero mean and $\sum_w$ as the covariance matrix. The mean and covariance of $f(\boldsymbol{\Omega})$ are then calculated as (13) and (14), respectively. By leveraging (11) and (13), (14) is further reduced to (15), where $K(\boldsymbol{\Omega},\boldsymbol{\Omega})$ concisely denotes the covariance operation, called a kernel function.

$$f(\boldsymbol{\Omega}) = \upsilon(\boldsymbol{\Omega})^T\boldsymbol{w} \quad (11)$$

$$\boldsymbol{w} = \mathcal{N}(0,\textstyle\sum_w) \quad (12)$$

$$\mu^{f(\boldsymbol{\Omega})} = \mathbb{E}[f(\boldsymbol{\Omega})] = 0 \quad (13)$$

$$\sigma^{2f(\boldsymbol{\Omega})} = \mathbb{E}[((f(\boldsymbol{\Omega})-\mu(\boldsymbol{\Omega}))((f(\boldsymbol{\Omega}')-\mu(\boldsymbol{\Omega}'))] \quad (14)$$

$$\sigma^{2f(\boldsymbol{\Omega})} = \upsilon(\boldsymbol{\Omega})^T\textstyle\sum_w\upsilon(\boldsymbol{\Omega}') = K(\boldsymbol{\Omega},\boldsymbol{\Omega}) \quad (15)$$

We here further introduce a squared exponential function as covariance, considering it yields a combination of infinite number of basis functions for the Bayesian regression model [29]. This more powerful kernel representing the covariance is as (16), where $\|\boldsymbol{\Omega}_p-\boldsymbol{\Omega}_q\|$ gives the Euclidean distance between two AE configurations $\boldsymbol{\Omega}_p$ and $\boldsymbol{\Omega}_q$, and l is a length scale. The prior distribution of $f(\boldsymbol{H_t^\Omega})$ is thus given as (17), where $\boldsymbol{H_t^\Omega}$ is a set of AE configurations already evaluated. Similarly, the prior distribution of $f(\boldsymbol{H^\Omega})$ is given in (18), where $\boldsymbol{H^\Omega}$ is the AE configuration space.

$$cov\left(f(\boldsymbol{\Omega_p}),f(\boldsymbol{\Omega_q})\right) = K(\boldsymbol{\Omega_p},\boldsymbol{\Omega_q}) = \exp\left(\frac{-\|\boldsymbol{\Omega}_p-\boldsymbol{\Omega_q}\|}{2l^2}\right) \quad (16)$$

$$f(\boldsymbol{H_t^\Omega})|\boldsymbol{H_t^\Omega}\sim\mathcal{N}(0,K(\boldsymbol{H_t^\Omega},\boldsymbol{H_t^\Omega})) \quad (17)$$

$$f(\boldsymbol{H^\Omega})|\boldsymbol{H^\Omega}\sim\mathcal{N}(0,K(\boldsymbol{H^\Omega},\boldsymbol{H^\Omega})) \quad (18)$$

Further, the joint Gaussian distribution of the training outputs, $f(\boldsymbol{H_t^\Omega})$, and the testing outputs $f(\boldsymbol{H^\Omega})$, based on the prior, is represented as (19), where $K(\boldsymbol{H_t^\Omega},\boldsymbol{H_t^\Omega})$, $K(\boldsymbol{H_t^\Omega},\boldsymbol{H^\Omega})$, and $K(\boldsymbol{H^\Omega},\boldsymbol{H^\Omega})$ correspond to three covariance matrices. After considering the observations $\boldsymbol{H_t^\zeta}$ with noise, $\mathcal{N}(0,\sigma_n^2)$, we introduce the noise item to (19), yielding (20).

$$\begin{bmatrix}f(\boldsymbol{H_t^\Omega})\\ f(\boldsymbol{H^\Omega})\end{bmatrix}|\boldsymbol{H^{\Omega_t}},\boldsymbol{H^\Omega}\sim\mathcal{N}\left(0,\begin{bmatrix}K(\boldsymbol{H_t^\Omega},\boldsymbol{H_t^\Omega}) & K(\boldsymbol{H_t^\Omega},\boldsymbol{H^\Omega})\\ K(\boldsymbol{H^\Omega},\boldsymbol{H_t^\Omega}) & K(\boldsymbol{H^\Omega},\boldsymbol{H^\Omega})\end{bmatrix}\right) \quad (19)$$

$$\begin{bmatrix}\boldsymbol{H_t^\zeta}\\ f(\boldsymbol{H^\Omega})\end{bmatrix}|\boldsymbol{H_t^\Omega},\boldsymbol{H^\Omega}\sim \mathcal{N}\left(0,\begin{bmatrix}K(\boldsymbol{H_t^\Omega},\boldsymbol{H_t^\Omega})+\sigma_n^2I & K(\boldsymbol{H_t^\Omega},\boldsymbol{H^\Omega})\\ K(\boldsymbol{H^\Omega},\boldsymbol{H_t^\Omega}) & K(\boldsymbol{H^\Omega},\boldsymbol{H^\Omega})\end{bmatrix}\right) \quad (20)$$

Next, the conditional distribution, based on the principles of conditioning the joint Gaussian prior distribution on the observations [30], is calculated as (21) through (23), where $f(\boldsymbol{H^\Omega})|\boldsymbol{H_t^\Omega},\boldsymbol{H_t^\zeta},\boldsymbol{H^\Omega}$ is the posterior probability of $f(\boldsymbol{H^\Omega})$ given $\boldsymbol{H_t^\Omega},\boldsymbol{H_t^\zeta}$, and $\boldsymbol{H^\Omega}$, which correspond to AE configurations already evaluated, observed AE loss scores for these configurations, and the AE configuration space, respectively. t and $t+1$ correspond to the most recent iteration index in IBBD, and next iteration, respectively.

$$f(\boldsymbol{H^\Omega})|\boldsymbol{H_t^\Omega},\boldsymbol{H_t^\zeta},\boldsymbol{H^\Omega} = \mathcal{N}(\mu_{t+1}^{f(\boldsymbol{H^\Omega})},\textstyle\sum_{t+1}^{f(\boldsymbol{H^\Omega})}) \quad (21)$$

$$\mu_{t+1}^{f(\boldsymbol{H^\Omega})} = K(\boldsymbol{H^\Omega},\boldsymbol{H_t^\Omega})[K(\boldsymbol{H_t^\Omega},\boldsymbol{H_t^\Omega})+\sigma_n^2I]^{-1}\boldsymbol{H_t^\zeta} \quad (22)$$

$$\textstyle\sum_{t+1}^{f(\boldsymbol{H^\Omega})} = K(\boldsymbol{H^\Omega},\boldsymbol{H^\Omega}) - K(\boldsymbol{H^\Omega},\boldsymbol{H_t^\Omega})[K(\boldsymbol{H_t^\Omega},\boldsymbol{H_t^\Omega})+\sigma_n^2I]^{-1}K(\boldsymbol{H_t^\Omega},\boldsymbol{H^\Omega}) \quad (23)$$

The Bayesian-backward learning on the mean $\mu_{t+1}^{f(\boldsymbol{H^\Omega})}$ is calculated as (22), where the predicted mean of $f(\boldsymbol{H^\Omega})$ in the iteration $t+1$ is estimated with the information from AE configurations already evaluated, $\boldsymbol{H_t^\Omega}$, observed evaluation results, $\boldsymbol{H_t^\zeta}$, and the space $\boldsymbol{H^\Omega}$ itself. The covariance, $\sum_{t+1}^{f(\boldsymbol{H^\Omega})}$, is as (23), which only relates to $\boldsymbol{H_t^\Omega}$ and $\boldsymbol{H^\Omega}$, not $\boldsymbol{H_t^\zeta}$. And the estimation is performed by linking $\boldsymbol{H_t^\Omega}$ and $\boldsymbol{H^\Omega}$ through the kernel function in (16).

We can observe that the Gaussian kernel gives almost unity for any two AE configurations which are very close to each other (if setting the length scale l to be 1). The kernel result decreases when the distance of AE configurations increases.

Now, the function values $f(\boldsymbol{H^\Omega})$, corresponding to the AE configuration space $\boldsymbol{H^\Omega}$, can be sampled from the joint posterior distribution described in (21) through (23). To determine the optimal AE recommendation, $\boldsymbol{\Omega}_{t+1}$, with a minimum predicted function value, the confidence of different AE configurations is calculated as (24) by co-considering the mean and variance with a weighting factor λ. Here λ is selected as 2.756, corresponding to the 96% confidence level.

$$\boldsymbol{\Omega}_{t+1} = arg\min_{\boldsymbol{\Omega}\in\boldsymbol{H^\Omega}}\left(\mu_{t+1}^{f(\boldsymbol{H^\Omega})}(\boldsymbol{\Omega}) + \lambda\left(\textstyle\sum_{t+1}^{f(\boldsymbol{H^\Omega})}(\boldsymbol{\Omega})\right)^{1/2}\right) \quad (24)$$

The recommended AE configuration, $\boldsymbol{\Omega}_{t+1}$, is backward sent then from the top-level to the bottom-level of IBBD for execution and evaluation in the iteration $t+1$.

D. IBBD Regularization Mechanisms

To thoroughly investigate IBBD regularization mechanisms, we have taken into account three strategies, as (25) through (27), denoted as IBBD_r, IBBD_e, IBBD_g, respectively.

$$\zeta = rRMSE\left(\phi^{\#}\left(\theta(\boldsymbol{\Omega})\right), \psi^{\#}\left(\vartheta(\boldsymbol{\Omega})\right), \boldsymbol{\chi}^{\boldsymbol{DB}}\right) \quad (25)$$

$$\zeta = P\left(\phi^{\#}\left(\theta(\boldsymbol{\Omega})\right)\right) \quad (26)$$

$$\zeta = \alpha * rRMSE\left(\phi^{\#}\left(\theta(\boldsymbol{\Omega})\right), \psi^{\#}\left(\vartheta(\boldsymbol{\Omega})\right), \boldsymbol{\chi}^{\boldsymbol{DB}}\right) + \beta * P\left(\phi^{\#}\left(\theta(\boldsymbol{\Omega})\right)\right) \quad (27)$$

IBBD_r ($\alpha = 1, \beta = 0$) as (25) focuses on rRMSE minimization, which is expected to yield AE configurations with minimum information loss but possible un-optimized power overhead. IBBD_e ($\alpha = 0, \beta = 1$) as (26), instead, searches AE configurations with minimum power overhead, without co-optimizing information loss. These two regularization strategies provide two special scenarios to demonstrate IBBD learning processes and facilitate comparison with IBBD_g.

In contrast, IBBD_g as (27) co-optimizes both information loss and AE power overhead, mainly with the regularization factor β that is selected from a set of values. These thorough considerations will provide rich and interesting results to demonstrate the behavior and effectiveness of the proposed IBBD framework, which will be detailed in the results section.

E. IBBD Generalization

To conduct solid evaluation of the IBBD algorithm, we have firstly executed IBBD on the ECG data from one human subject, and then generalized the determined optimal AE configurations to other human subjects.

More specifically, we have taken into account multiple dimensionalities for the AE configuration $\boldsymbol{\Omega}$, including the depth of the encoder, the AE input width, the number of feature maps of each stage, the max-pooling size, and the learning rate. This huge AE space is expected to contain high-quality AE candidates with both low error and power overhead, which, at the same time, poses a big challenge in AE configuration determination. We therefore apply the proposed IBBD algorithm on this AE space, on one human subject, to determine the optimal candidates.

We will also analyze in detail the top ten optimal candidates for each IBBD_g regularization sub-strategy to demonstrate the consistent quality of candidates. For instance, for IBBD_g010, where $\alpha = 1$ and $\beta = 0.010$, we will evaluate ten optimal candidates, IBBD_g010_optA through IBBD_g010_optJ.

Afterwards, we generalize the optimal AE candidate $\boldsymbol{\Omega}^{opt}$ to each $\boldsymbol{\chi}^{sub(s)}$, data of the subject s, among S human subjects and determine the AE quality with (28), where η is any metric and $\mathcal{M}$ is the corresponding metric function.

$$\eta = \sum_{s=1}^{S} \mathcal{M}\left(\phi^{\#}\left(\theta(\boldsymbol{\Omega}^{opt})\right), \psi^{\#}\left(\vartheta(\boldsymbol{\Omega}^{opt})\right), \boldsymbol{\chi}^{sub(s)}\right)/S \quad (28)$$

F. Framework Evaluation

We will also implement the common practice, DWT, to demonstrate the effectiveness of the proposed IBBD in terms of optimal AE configuration determination. Besides, the above-mentioned IBBD_r and IBBD_e, will provide interesting comparison with signal-metric optimization. Further, multiple sub-strategies of IBBD_g allow us to thoroughly evaluate the IBBD behavior and effectiveness. Last, the generalization from one human subject to other subjects will offer a solid justification of the IBBD framework.

III. Results

A. Experimental Setup

Diverse bio-dynamics are provided by the well-known MIT-BIH Arrythmia Database [31, 32], which has dozens of different types of arrythmia-related ECG abnormalities and is thus applied in our experiments. From each 30-min 360-Hz recording, we have selected 2^{19} samples that is about 24.27 minutes and can be easily segmented with a window size of 2^n. The ECG time series of ten high-fidelity recordings has been processed through a band-pass filter with a range of 0.5 Hz to 49.5 Hz, to remove the baseline wander and power-line interference. 75% and 25% of the segments of each subject is used for training and testing, respectively.

The regularization parameters evaluated for IBBD_g are chosen as: $\alpha = 1, \beta \in \{0.010, 0.035, 0.050, 0.100\}$. The four sub-strategies are denoted as IBBD_g010, IBBD_g035, IBBD_g050, and IBBD_g100, respectively. A huge AE space has been explored, based on the depth of the encoder ($\{2,4,6\}$), the AE input width ($\{128, 256, 512\}$), the number of feature maps of each stage ($\{2,3,4,5,6,7,8\}$), the max-pooling size (fixed at 2 to facilitate compression ratio determination when increasing the depth), and the learning rate (from 0.005 to 0.025).

The power consumption of the compression algorithms, including the AE-encoders and the DWT methods, is calculated based on the number of operations required and the power table of the ARM processor [33]. The Bluetooth data transmission power is based on the energy characterization table [34]. We in this study mainly focus on the algorithm design and validation, and in future, will implement the algorithms on the hardware for further experiments. Besides, IBBD enables a possibility of deactivating the Bluetooth module in enlarged no-transmission durations. The so-called 'sleeping mode' of Bluetooth can further minimize the significant static power consumed for maintaining connection, thereby greatly increasing the power saving. We in future will also investigate this great potential of further power saving offered by IBBD.

B. Single-Objective IBBD Learning

To investigate the behavior of IBBD and also provide interesting comparison, we have firstly evaluated the best-effort-on-error optimization, *i.e.*, IBBD_r, and the best-effort-on-power optimization, *i.e.*, IBBD_e. As shown in Fig. 3, the last 200 iterations are visualized in Fig. 3(a), and further zoomed versions are given in Fig. 3(a1) and Fig. 3(a2), for readability purpose. We can observe that IBBD either converges to the minimum error area (red dots) or the minimum power

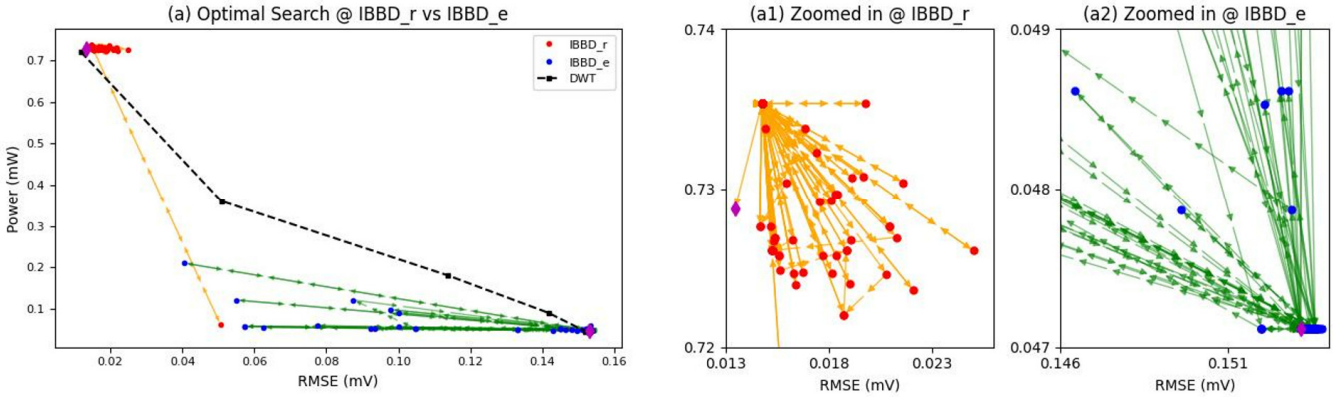

Fig. 3 Evaluations on the best-effort-on-error optimization, i.e., IBBD_r, and the best-effort-on-power optimization, i.e., IBBD_e, to investigate the behavior of IBBD and also provide interesting comparison. The effectiveness of IBBD will be mainly further demonstrated later in dual-objective learning.
Notes. The optimal configuration search processes for both are visualized in (a). The zoomed in version for IBBD_r is in (a1), and for IBBD_e is in (a2). Only last 200 iterations are visualized in (a), and further zoomed versions are given, for readability purpose. Diamonds indicate the best optimal configurations. We can observe that IBBD either converges to the minimum error area (red dots) or the minimum power area (red dots), indicating its learning effectiveness under different regularization strategies. The DWT method is also included for comparison, indicating similar optimal configurations as those determined by single-objective IBBD learning.

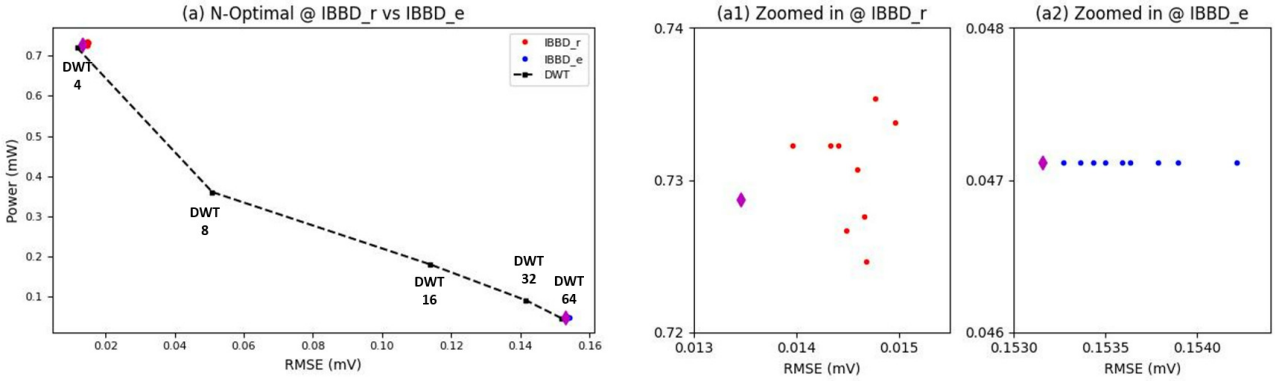

Fig. 4 The top N (N=10) optimal configurations of IBBD_r and IBBD_e. The consistent quality of configurations is shown in (a), and further illustrated with zoomed-in graphs in (a1) and (a2). Diamonds indicate the best optimal configurations.

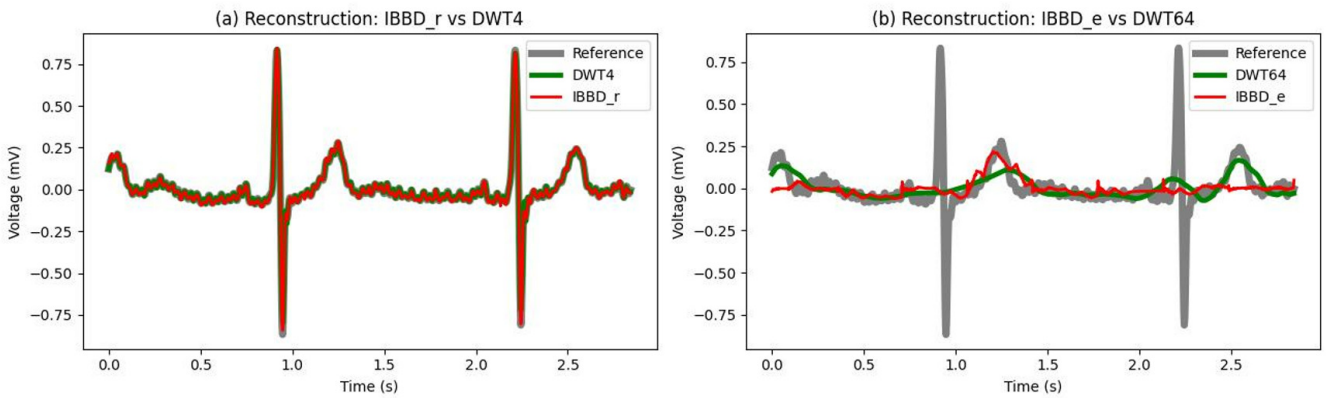

Fig. 5 ECG reconstruction comparison, indicating similar quality between single-objective IBBD and DWT. (a) IBBD_r and DWT4 both have little information loss when targeting error minimization. (b) IBBD_e and DWT64 sacrifice the reconstruction performance when minimizing power overhead.

area (blue dots), indicating its learning effectiveness under different regularization strategies. The DWT method is also included for comparison, indicating similar optimal configurations as those determined by the single-objective IBBD learning.

Further, we have illustrated the top N (N=10) optimal configurations in Fig. 4. The consistent quality of configurations is shown in Fig. 4(a), and further illustrated with zoomed-in graphs in Fig. 4(a1) and Fig. 4(a2). Diamonds indicate best optimal configurations. The distribution of IBBD_r is well concentrated and the best optimum is very close to other optimums. Similar quality is observed from IBBD_e.

The best optimums of IBBD_r and IBBD_e have been compared further in Fig. 5 and Table I. In Fig. 5, IBBD_r and DWT4 (compression ratio 4) both have little information loss, when targeting error minimization. IBBD_e and DWT64

TABLE I. Summary of the best optimum of IBBD_r and the best optimum of IBBD_e, indicating a similar quality as DWT.

	CR	nLayer	nWidth	nMap	nKernel	nPooling	nLearning	RMSE	Power
IBBD_r	4	2	128	5	14	2	0.02	0.013	0.729
DWT4	4							0.012	0.720
IBBD_e	64	6	128	2	2	2	0.005	0.153	0.047
DWT64	64							0.152	0.045

Notes. CR: compression ratio; nLayer: depth of encoder (and decoder); nWidth: input width; nMap: number of feature maps of each layer; nKernel: convolutional filter size; nPooling: max-pooling size; nLearning: learning rate; units of RMSE and Power: mV and mW.

sacrifice the reconstruction performance when minimizing power overhead. We will later in dual-objective IBBD learning demonstrate the superiority of IBBD_g. Table I gives both details of both IBBD configurations, indicating similar RMSE and power for IBBD and DWT. We have also found that, IBBD_r prefers a smaller compression ratio and a large convolutional kernel size, which are important to extract the patterns and minimize information loss.

C. Dual-Objective IBBD Learning

The regularized Bayesian-backward deep encoding learning can further co-optimize the error and power, to provide AE configurations superior to DWT. We here have thoroughly evaluated four co-optimization regulation strategies: IBBD_g010, IBBD_g035, IBBD_g050, and IBBD_g100, with gradually increasing emphasis on power optimization.

As shown in Fig. 6, when the regularization factor is only 0.010, the corresponding algorithm IBBD_g010 still mainly targets error minimization. When increasing the factor to 0 .035, IBBD_g035 pushes the search process both downwards and rightwards, thereby providing AE configurations with error-power co-optimization. Further increasing the factor to 0.050 and then to 0.100, the IBBD_g050 and IBBD_g100 have more emphasis on energy optimization, and the latter one results in very attractive AE configurations.

The top N (N=10) optimums of each strategy are further given in Fig. 7, to demonstrate the consistent quality of four IBBG_g strategies on error-power co-optimization, with different levels of emphasis on power. We find that, IBBD_g035 and IBBD_g050 have very similar top N

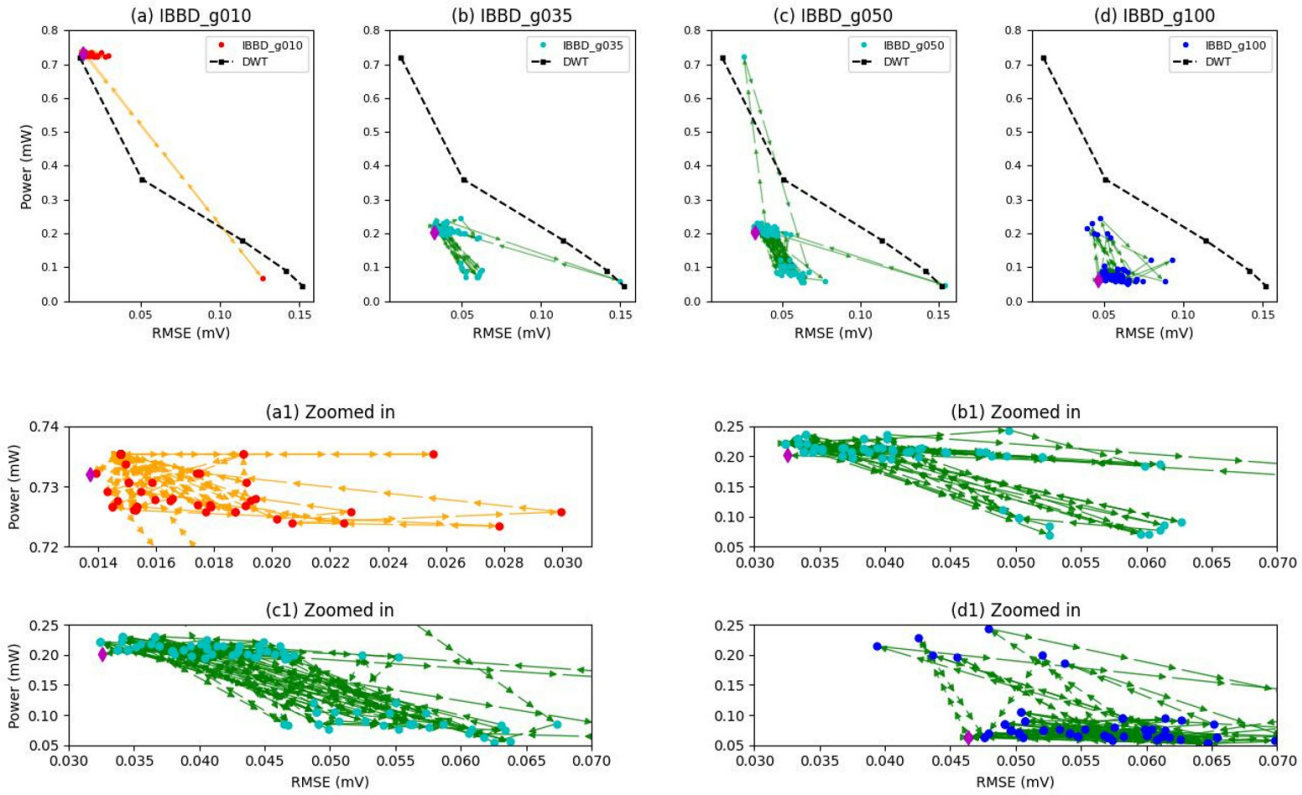

Fig. 6 Evaluation of IBBD_g with different regularization strategies, *i.e.*, IBBD_g010, IBBD_g035, IBBD_g050, and IBBD_g100, with gradually increasing emphasis on power optimization. Both big pictures, (a) to (d), and zoomed in illustrations, (a1) to (d1), are given. Diamonds indicate the best optimal configurations. When the regularization factor is only 0.010, the corresponding algorithm IBBD_g010 still mainly targets error minimization. When increasing the factor to 0.035, IBBD_g035pushes the search process both downwards and rightwards, thereby providing AE configurations with error-power co-optimization. Further increasing the factor to 0.050 and then to 0.100, the IBBD_g050 and IBBD_g100 have more emphasis on energy optimization, and the latter one results in very attractive AE configurations.

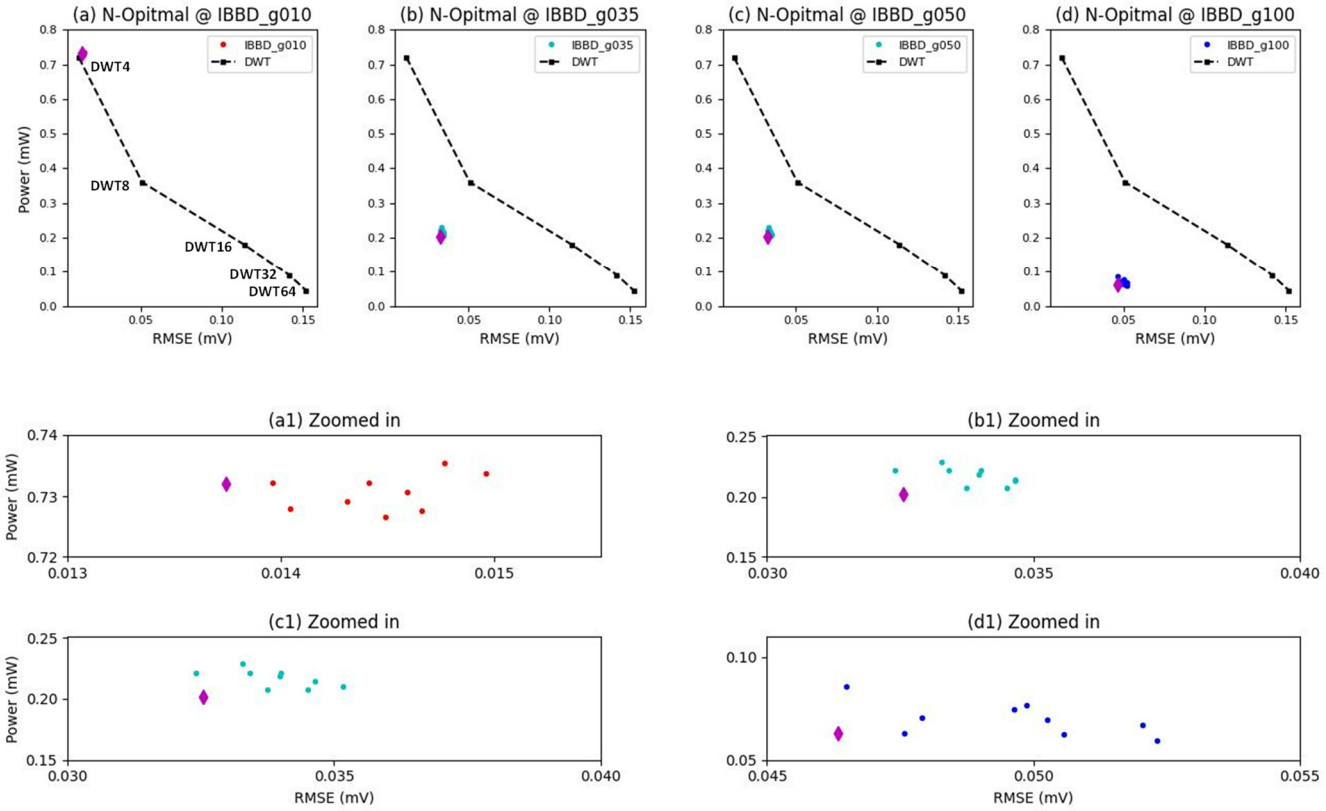

Fig. 7 Top *N* (*N*=10) optimums of each strategy are further illustrated to demonstrate the consistent quality of four IBBG_g strategies on error-power co-optimization, with different levels of emphasis on power. Diamonds indicate the best optimal configurations.

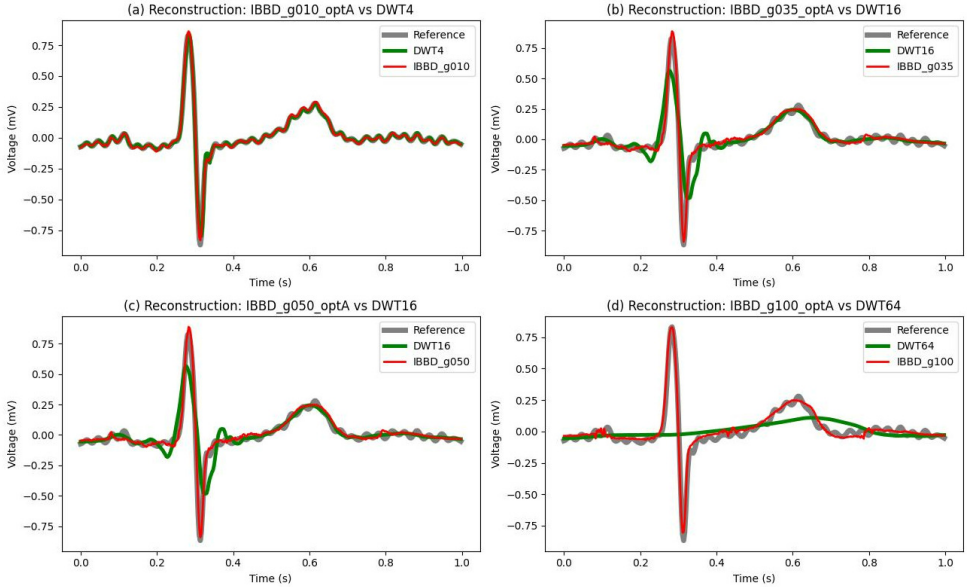

Fig. 8 ECG reconstruction comparison of four IBBD_g regularization strategies, indicating, during error-power co-optimization, larger regularization factors on power bring increasing information loss. But the signal reconstruction quality for IBBD_g decreases much slower than DWT, demonstrating the effectiveness of the proposed IBBD framework. Especially, in (d), the IBBD_g100 has substantially better reconstruction than DWT64.

optimums. IBBD_g100 has dramatically different top *N* optimums with more emphasis on power optimization.

ECG reconstruction comparison of four IBBD_g regularization strategies, as shown in Fig. 8, indicates, during error-power co-optimization, larger regularization factors on power bring increasing information loss. But the signal reconstruction quality for IBBD_g decreases much slower than DWT, demonstrating the effectiveness of the proposed IBBD framework.

The best optimum under each of four strategies is summarized in TABLE II. There are several interesting findings here. Firstly, IBBD_g010_optA is has similar RMSE and power as DWT4. Secondly, IBBD_g035_optA and IBBD_g050_optA have a same AE configuration, with both lower power and

TABLE II. Summary of the best optimum for each of four IBBD_g strategies, which illustrate the superiority to DWT on error-power co-optimization.

	CR	nLayer	nWidth	nMap	nKernel	nPooling	nLearning	RMSE	Power
IBBD_g010_optA	4	2	128	7	14	2	0.02	0.014	0.732
DWT4	4							0.012	0.720
IBBD_g035_optA	16	4	128	5	12	2	0.005	0.033	0.202
IBBD_g050_optA	16	4	128	5	12	2	0.005	0.033	0.202
DWT8	8							0.051	0.360
DWT16	16							0.114	0.180
IBBD_g100_optA	64	6	128	5	8	2	0.005	0.046	0.063
DWT32	32							0.142	0.090
DWT64	64							0.152	0.045

Notes. optA corresponds to the best optimum of each strategy.

lower RMSE, than those of DWT8. Though DWT16 has slightly lower power, its RMSE is much higher. Thirdly, IBBD_g100_optA has both lower power and lower RMSE than DWT32, and though DWT64 has slightly lower power, its RMSE is substantially high (0.152 mV compared to 0.046 mV of IBBD_g100_optA).

Therefore, we conclude IBBD_g035_optA (or IBBD_g500_optA) provides an AE configuration with moderate RMSE and power, and IBBD_g100_optA further decreases power by 3.2 times (0.202 mW to 0.063 mW), with only an increase in RMSE of 39%. Further, with more emphasis on power optimization, IBBD_g prefers to recommend AE configurations with a larger depth, a smaller amount of feature maps, a smaller convolutional kernel size, and a smaller learning rate. These indicate IBBD_g puts more attention on smaller but appropriate model sizes, and a smaller learning rate.

D. Quality of Optimal Candidates

To illustrate the ECG reconstruction quality of IBBD_g035 and IBBD_g100 strategies, we have in Fig. 9 and Fig. 10 further demonstrated four optimums (optB to optE) for each. In Fig. 9, the illustrations of other four optimums of IBBD_g035, still indicate greatly better signal reconstruction morphologies than DWT16. In Fig. 10, IBBD_g100, with a slight sacrifice of performance, actually provides substantially better signal reconstruction morphologies than DWT64.

TABLE III and TABLE IV summarize five optimums for IBBD_g035 and IBBD_g100, respectively. In the former table, the five optimums all have both lower power (no more than 0.222 mW) and lower RMSE (no more than 0.034 mV), than DWT8 (0.360 mW in power and 0.051 mV in RMSE). Again, DWT16 obtains slighter lower power with a significant increase in RMSE (from 0.051 mV to 0.114 mV). The numbers of

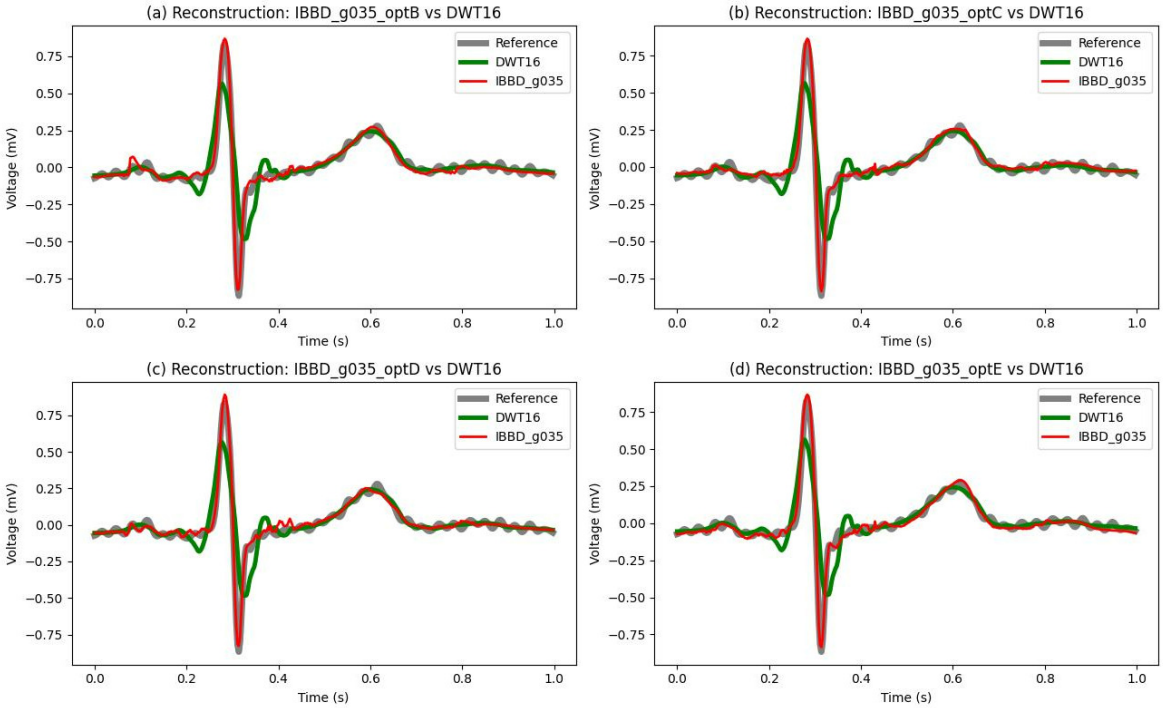

Fig. 9 Further illustrations of other four optimums of IBBD_g035, indicating greatly better signal reconstruction morphologies than those of DWT.

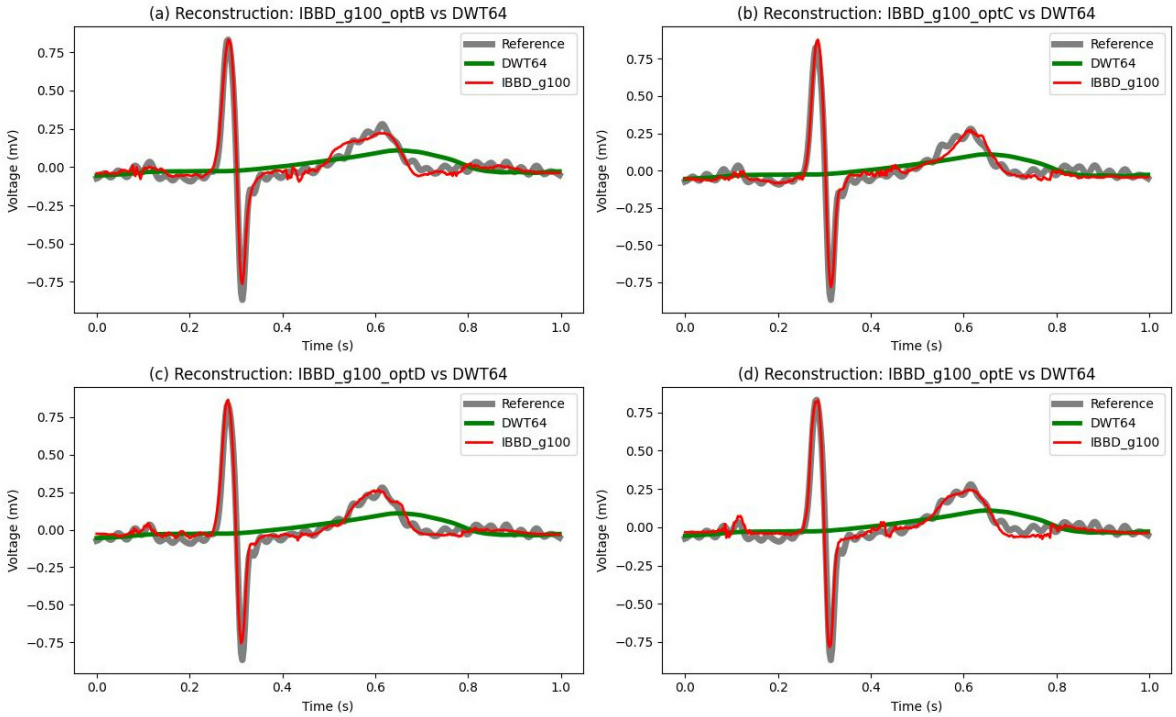

Fig. 10 Further illustrations of other four optimums of IBBD_g100, indicating substantially better signal reconstruction morphologies than those of DWT64.

TABLE III. Summary of five optimums of IBBD_g035, with both lower power and lower RMSE than DWT8, and much lower RMSE than DWT16.

	CR	nLayer	nWidth	nMap	nKernel	nPooling	nLearning	RMSE	Power
IBBD_g035_optA	16	4	128	5	12	2	0.005	0.033	0.202
IBBD_g035_optB	16	4	128	8	10	2	0.01	0.032	0.222
IBBD_g035_optC	16	4	128	7	8	2	0.025	0.034	0.207
IBBD_g035_optD	16	4	128	7	8	2	0.02	0.034	0.207
IBBD_g035_optE	16	4	128	8	10	2	0.005	0.033	0.222
DWT8	8							0.051	0.360
DWT16	16							0.114	0.180

TABLE IV. Summary of five optimums of IBBD_g100, with both lower power and RMSE than DWT32, and substantially lower RMSE than DWT64.

	CR	nLayer	nWidth	nMap	nKernel	nPooling	nLearning	RMSE	Power
IBBD_g100_optA	64	6	128	5	8	2	0.005	0.046	0.063
IBBD_g100_optB	64	6	128	5	8	2	0.01	0.048	0.063
IBBD_g100_optC	64	6	128	5	12	2	0.025	0.048	0.071
IBBD_g100_optD	64	6	128	4	12	2	0.005	0.051	0.063
IBBD_g100_optE	64	6	128	5	6	2	0.025	0.052	0.059
DWT32	32							0.142	0.090
DWT64	64							0.152	0.045

feature maps, the convolutional filter sizes, and the learning rates, have shown some diversity. In the latter table, similar findings can be made and IBBD_g100 enables very attractive error-power co-optimization, by further decreasing the number of feature maps, and/or the convolutional filter size. The large reconstruction error of DWT64 in Fig. 9 is also reflected in TABLE IV, meaning that its direct drop of decomposed signal details [35, 36], causes big signal distortions.

E. IBBD Generalization

We have conducted generalization evaluations of IBBD_g035_optA and IBBD_g100_optA to ECG reconstruction on other human subjects, which, as shown in Fig. 11, demonstrate the determined AE configurations are much better than DWT16 and DWT64, respectively. In TABLE V, we have summarized the performance of each subject, and provided the average RMSE and average power on all subjects.

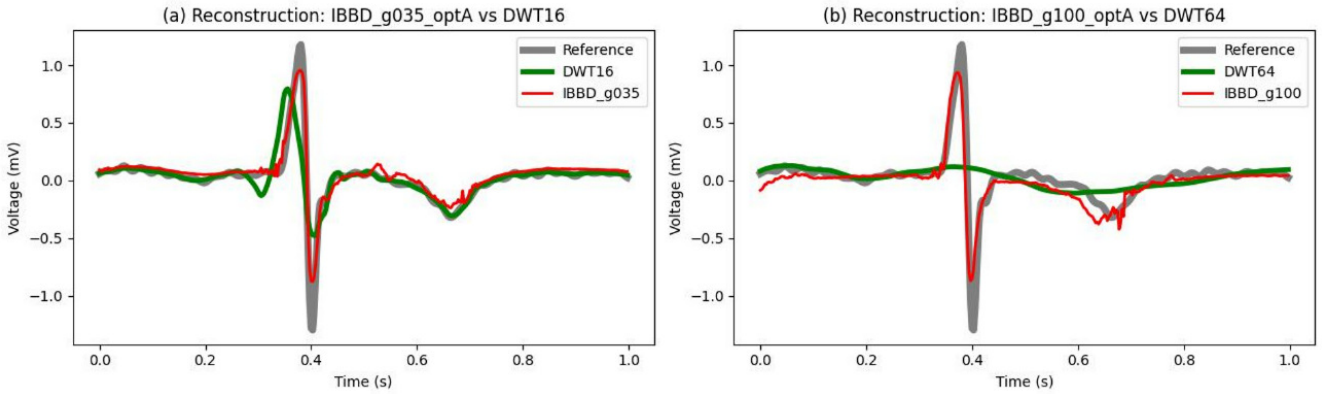

Fig. 11 Generalization evaluation of IBBD_g035_optA and IBBD_g100_optA to ECG reconstruction of other human subjects, demonstrating the determined AE configurations are much better than DWT16.

TABLE V. Summary of AE quality on all subjects, indicating IBBG_g provides configurations with very attractive error-power co-optimization. DWT methods either have both higher power and higher RMSE, or have slightly lower power but a great sacrifice in RMSE.

	RMSE_ S1	RMSE_ S2	RMSE_ S3	RMSE_ S4	RMSE_ S5	RMSE_ S6	RMSE_ S7	RMSE_ S8	RMSE_ S9	RMSE_ S10	RMSE Average	Power Average
IBBD_g035_optA	0.037	0.071	0.025	0.032	0.054	0.027	0.061	0.034	0.033	0.057	0.043	0.202
DWT8	0.052	0.087	0.033	0.027	0.050	0.024	0.065	0.038	0.037	0.051	0.046	0.360
DWT16	0.091	0.161	0.059	0.050	0.079	0.043	0.121	0.087	0.050	0.114	0.086	0.180
IBBD_g100_optA	0.055	0.133	0.035	0.053	0.062	0.051	0.099	0.050	0.049	0.055	0.064	0.063
DWT32	0.112	0.201	0.074	0.082	0.094	0.050	0.155	0.130	0.064	0.142	0.111	0.090
DWT64	0.121	0.220	0.081	0.099	0.104	0.053	0.181	0.151	0.071	0.152	0.123	0.045

Notes. S1: subject 1 that has been leveraged for AE optimums determination with IBBD; S2 to S10: unseen subjects.

The results on all subjects have effectively shown similar findings as detailed previously on the single subject, meaning that the optimal AE configurations can be well generalized to other unseen subjects.

More specifically, IBBD_g035_optA, on average, has power and RMSE as 0.202 mW and 0.043 mV, respectively, which are smaller than 0.360 mW and 0.046 mV, respectively, for DWT8. DWT16 decreases the power slightly to 0.180 mW, but increases RMSE greatly to 0.086 mV. Further, IBBD_g100_optA offers power as low as 0.063 mW, with RMSE only increased to 0.064 mV. DWT32, instead, has both higher power and higher RMSE, at 0.090 mW and 0.111 mV, respectively. DWT64 slightly decreases the power to 0.045 mW, but sacrifices a lot in RMSE to 0.123 mW.

IV. Conclusion

We proposed and evaluated a novel framework, IBBD, denoting IoMT Big data Bayesian-backward Deep-encoder learning, to mine the optimal deep autoencoder configurations that can effectively trade-off between information loss during data sparsification and power overhead. The regularized Bayesian-backward deep encoding learning algorithm, at the top-level, quantifies the quality of AE configurations by combing metrics including information loss and power overhead, thereby generating optimal AE recommendations. The AE learning is executed at the bottom-level for configuration evaluation. The proposed algorithm has provided different optimums based on the settings, and allows the user to apply different strategies of error-power trade-off. Evaluated on the cardiac data streaming application, IBBG has demonstrated AE configurations with very attractive error-power co-optimization, compared with the common practice – DWT methods that either have both higher power and higher RMSE, or have slightly lower power but a great sacrifice in RMSE. This study will greatly benefit IoMT big data streaming towards data driven-precision medicine.

References

[1] A. K. Idrees, S. K. Idrees, R. Couturier, and T. Ali-Yahiya, "An Edge-Fog Computing Enabled Lossless EEG Data Compression with Epileptic Seizure Detection in IoMT Networks," *IEEE Internet of Things Journal,* 2022.

[2] J. Qian, P. Tiwari, S. P. Gochhayat, and H. M. Pandey, "A noble double-dictionary-based ECG compression technique for IoTH," *IEEE Internet of Things Journal,* vol. 7, no. 10, pp. 10160-10170, 2020.

[3] Q. Zhang, D. Arney, J. M. Goldman, E. M. Isselbacher, and A. A. Armoundas, "Design Implementation and Evaluation of a Mobile Continuous Blood Oxygen Saturation Monitoring System," *Sensors,* vol. 20, no. 22, p. 6581, 2020.

[4] J. Zou and Q. Zhang, "Brain Visual Dynamics Decoding with Deep Learning & Edge Computing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 2022.

[5] M. Alkhodari, H. F. Jelinek, N. Werghi, L. J. Hadjileontiadis, and A. H. Khandoker, "Estimating Left Ventricle Ejection Fraction Levels Using Circadian Heart Rate Variability Features and Support Vector Regression Models," *IEEE Journal of Biomedical and Health Informatics,* vol. 25, no. 3, pp. 746-754, 2020.

[6] W. Zhang, L. Yu, L. Ye, W. Zhuang, and F. Ma, "ECG signal classification with deep learning for heart disease identification," in *2018 International Conference on Big Data and Artificial Intelligence (BDAI)*, 2018: IEEE, pp. 47-51.

[7] C. Venkatesan, P. Karthigaikumar, and S. Satheeskumaran, "Mobile cloud computing for ECG telemonitoring and real-time coronary heart disease risk detection," *Biomedical Signal Processing and Control,* vol. 44, pp. 138-145, 2018.

[8] S. C. Mukhopadhyay, S. K. S. Tyagi, N. K. Suryadevara, V. Piuri, F. Scotti, and S. Zeadally, "Artificial Intelligence-based Sensors for Next Generation IoT Applications: A Review," *IEEE Sensors Journal,* 2021.

[9] Q. Zhang *et al.*, "IEEE Access Special Section Editorial: Smart Health Sensing and Computational Intelligence: From Big Data to Big Impacts," *IEEE Access,* vol. 9, pp. 30452-30455, 2021.

[10] Z. Lu, D. Y. Kim, and W. A. Pearlman, "Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm," *IEEE transactions on Biomedical Engineering,* vol. 47, no. 7, pp. 849-856, 2000.

[11] J. Sahambi, S. Tandon, and R. Bhatt, "Using wavelet transforms for ECG characterization. An on-line digital signal processing system," *IEEE Engineering in Medicine and Biology Magazine,* vol. 16, no. 1, pp. 77-83, 1997.

[12] D. Mitra, H. Zanddizari, and S. Rajan, "Investigation of kronecker-based recovery of compressed ecg signal," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 6, pp. 3642-3653, 2019.

[13] V. Izadi, P. K. Shahri, and H. Ahani, "A compressed-sensing-based compressor for ECG," *Biomedical engineering letters,* vol. 10, no. 2, pp. 299-307, 2020.

[14] C. Jha and M. Kolekar, "Empirical mode decomposition and wavelet transform based ECG data compression scheme," *IRBM,* vol. 42, no. 1, pp. 65-72, 2021.

[15] B. Nguyen, W. Ma, and D. Tran, "Impact of lossy data compression techniques on EEG-based pattern recognition systems," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018: IEEE, pp. 2308-2313.

[16] G. Cisotto, A. V. Guglielmi, L. Badia, and A. Zanella, "Joint compression of EEG and EMG signals for wireless biometrics," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018: IEEE, pp. 1-6.

[17] F. Pareschi *et al.*, "Energy analysis of decoders for rakeness-based compressed sensing of ECG signals," *IEEE transactions on biomedical circuits and systems,* vol. 11, no. 6, pp. 1278-1289, 2017.

[18] E. Balestrieri, L. De Vito, F. Picariello, and I. Tudosa, "A novel method for compressed sensing based sampling of ECG signals in medical-IoT era," in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2019: IEEE, pp. 1-6.

[19] J. Šaliga, I. Andráš, P. Dolinský, L. Michaeli, O. Kováč, and J. Kromka, "ECG compressed sensing method with high compression ratio and dynamic model reconstruction," *Measurement,* vol. 183, p. 109803, 2021.

[20] N. Soni, I. Saini, and B. Singh, "Morphologically Robust Discrete Cosine Transform based Lossless ECG Compression with Access Control Quality," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 2018: IEEE, pp. 289-293.

[21] A. Pandey, B. S. Saini, B. Singh, and N. Sood, "Quality controlled ECG data compression based on 2D discrete cosine coefficient filtering and iterative JPEG2000 encoding," *Measurement,* vol. 152, p. 107252, 2020.

[22] A. F. Hussein, A. K. AlZubaidi, A. Al-Bayaty, and Q. A. Habash, "An IoT real-time biometric authentication system based on ECG fiducial extracted features using discrete cosine transform," *arXiv preprint arXiv:1708.08189,* 2017.

[23] L. Sun, Z. Zhong, Z. Qu, and N. Xiong, "PerAE: An Effective Personalized AutoEncoder for ECG-Based Biometric in Augmented Reality System," *IEEE Journal of Biomedical and Health Informatics,* vol. 26, no. 6, pp. 2435-2446, 2022.

[24] A. A. Abdellatif, M. G. Khafagy, A. Mohamed, and C.-F. Chiasserini, "EEG-based transceiver design with data decomposition for healthcare IoT applications," *IEEE Internet of Things Journal,* vol. 5, no. 5, pp. 3569-3579, 2018.

[25] I. R. Panneerselvam, "Transfer learning autoencoder used for compressing multimodal biosignal," *Multimedia Tools and Applications,* vol. 81, no. 13, pp. 17547-17565, 2022.

[26] T. Wei, S. Liu, and X. Du, "Learning-Based Efficient Sparse Sensing and Recovery for Privacy-Aware IoMT," *IEEE Internet of Things Journal,* 2022.

[27] R. B. Gramacy, *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. Chapman and Hall/CRC, 2020.

[28] C. K. Williams and D. Barber, "Bayesian classification with Gaussian processes," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 20, no. 12, pp. 1342-1351, 1998.

[29] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural computation,* vol. 15, no. 7, pp. 1667-1689, 2003.

[30] C. Williams and C. Rasmussen, "Gaussian processes for regression," *Advances in neural information processing systems,* vol. 8, 1995.

[31] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation,* vol. 101, no. 23, pp. e215-e220, 2000.

[32] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine,* vol. 20, no. 3, pp. 45-50, 2001.

[33] S. Nikolaidis, "Instruction-level energy characterization of an ARM processor," in *Proceedings of the 2nd MARLOW workshop*, 2003.

[34] R. Balani, "Energy consumption analysis for bluetooth, wifi and cellular networks," *Online Httpnesl Ee Ucla Edufwdocumentsreports2007PowerAnalysis Pdf,* 2007.

[35] R. S. Pathak, *The wavelet transform*. Springer Science & Business Media, 2009.

[36] D. K. Alves, R. L. Ribeiro, F. B. Costa, and T. O. A. Rocha, "Real-time wavelet-based grid impedance estimation method," *IEEE Transactions on Industrial Electronics,* vol. 66, no. 10, pp. 8263-8265, 2018.

Junhua Wong is a graduate student in Department of Electrical and Computer Engineering, Purdue School of Engineering and Technology, Indianapolis, USA. His research area includes smart health, rehabilitation, deep learning, machine learning, big data, and efficient computing technologies.

Vincenzo Piuri (Fellow, IEEE) received the Ph.D. degree in computer engineering from Politecnico di Milano, Italy, in 1989. He is a Full Professor of computer engineering at the Università degli Studi di Milano, Italy, since 2000. He has also been an Associate Professor at Politecnico di Milano, Italy, and a Visiting Professor at The University of Texas at Austin, USA, and a Visiting Researcher at the George Mason University, USA.

His research interests include artificial intelligence, computational intelligence, intelligent systems, machine learning, pattern analysis and recognition, signal and image processing, biometrics, intelligent measurement systems, industrial applications, digital processing configurations, fault tolerance, dependability, and cloud computing infrastructures. Original results have been published in more than 400 articles in international journals, proceedings of international conferences, books, and book chapters.

He is a Fellow of the IEEE, a Distinguished Scientist of ACM and a Senior Member of INNS. He is IEEE Region 8 Director-elect (2021-22). He has been IEEE Vice President for Technical Activities (2015), IEEE Director, President of the IEEE Systems Council and the IEEE Computational Intelligence Society, Vice President for Education of the IEEE Biometrics Council, Vice President for Publications of the IEEE Instrumentation and Measurement Society and the IEEE Systems Council, and Vice President for Membership of the IEEE Computational Intelligence Society. He is an Associate Editor of the IEEE Transactions on Cloud Computing. He has been the Editor-in-Chief of the IEEE Systems Journal (2013-19) and an AE of IEEE Transactions on Computers, IEEE Transactions on Neural Networks, IEEE Transactions on Instrumentation and Measurement, and IEEE Access.

He received the IEEE Instrumentation and Measurement Society Technical Award (2002) and the IEEE TAB Hall of Honor (2019). He is an Honorary Professor at Obuda University, Hungary; Guangdong University of Petro□hemi□al Te□hnology, China; Northeastern University, China; Muroran Institute of Te□hnology, Japan; Amity University, India; and Galgotias University, India.

Fabio Scotti (Senior Member, IEEE) received the Ph.D. degree in computer engineering from the Politecnico di Milano, Milan, Italy, in 2003. He was an Assistant Professor at the Department of Information Technologies, Università degli Studi di Milano, Italy (2002-2015). He was an Associate Professor at the Department of Computer Science, Università degli Studi di Milano, Italy (2015-2020). He is a Full Professor at the Università degli Studi di Milano, Italy since 2020.

Original results have been published in over 150 papers in international journals, proceedings of international conferences, books, book chapters, and patents. His current research interests include biometric systems, machine learning and computational intelligence, signal and image processing, theory and applications of neural networks, three-dimensional reconstruction, industrial applications, intelligent measurement systems, and high-level system design.

He is an Associate Editor of the IEEE Transactions on Human-Machine Systems and the IEEE Open Journal of Signal Processing. He is serving as Book Editor (Area Editor, section Less-constrained Biometrics) of the Encyclopedia of Cryptography, Security, and Privacy (3rd Edition), Springer. He has been an Associate Editor of the IEEE Transactions on Information Forensics and Security, Soft Computing (Springer) and a Guest Coeditor for the IEEE Transactions on Instrumentation and Measurement.

Qingxue Zhang (Senior Member, IEEE) has over fifteen years' experience in both academia and industry, with his postdoc research at Harvard, products R&D in ICT, and Ph.D. research at University of Texas at Dallas. He started the faculty position in Purdue University School of Engineering & Technology, Indianapolis, USA, in 2018. He is directing the Ubiquitous Intelligence Lab, with research interests including Deep Learning, IoT/Wearable, Edge Computing, and Brain-inspired Learning, targeting smart health, home, and world applications.

He is a recipient of the prestigious USA NSF CAREER Award. He serves as USA NSF/NIH/NIST Panelists, IEEE workshop chairs, AE for IEEE Access, and committee for multiple IEEE conferences. He received the Featured Journal Article Award in IEEE Access, the Best Paper Award in UEMCON2017, the Early-Career Travel Award in AHA2020, the Favorite Professor Award in 2019, and the Google Cloud Award in 2019. He is a member of the IEEE Sensors Council, Computer Society, Communications Society, and Signal Processing Society, and also is a senior member of IEEE.