Improved ECG Denoising using CEEMAN based on Complexity Measure and Nonlocal Mean Approach

Punitkumar Bhavsar, *Member, IAENG*

Abstract— **Measurement of cardiac electrical activity also known as electrocardiogram (ECG) is prone to a variety of noise which deteriorates the accuracy of extracted clinical parameters. Raw ECG signals are non-liner and non-stationary in nature which limits the use of traditional filters for denoising. The complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), an advanced version of Ensemble Empirical Mode Decomposition (EEMD), has shown usefulness in denoising ECG. However, it suffers from initial noisy intrinsic mode functions (IMFs) which are difficult to identify and denoise. Studies have shown various methods from information theory perspective to identify such noise corrupted IMFs. However, none of the study could provide a consummate solution. This article proposes the use of sample entropy to identify the noisy IMFs and utilize the non-local mean smoothening technique to reduce the effect of noise. Proper identification of noisy IMFs followed by adaptive smoothing using the NLM algorithm improves denoising performance of the CEEMDAN algorithm. The proposed method is verified considering real ECG data from MIT-BIH. The comparative analysis from the results suggests the use of the proposed method for improved performance of denoising of ECG signals under the influence of higher noise.**

Index Terms— **CEEMDAN, NLM, ECG, Sample Entropy, Denoising algorithms**

I. INTRODUCTION

The process to extract the vital parameters from noisy biomedical signals presents challenges due to the overlap of noise with the information content[1]. In the case of electrocardiogram (ECG), high-frequency noise from muscles overlaps with the QRS complex, and low-frequency components are corrupted by the baseline wandering and movement artifacts[1]. Denoising ECG with the traditional brick type filters results in the removal of important information that limits the accurate estimation of the medical parameters. Hence, denoising based on the information content is critical for efficient ECG denoising.

Almost all the biomedical signals are nonlinear and non-stationary which requires adaptive denoising techniques based on the characteristic of noise[1]. Reliability of wearable sensing and remote health monitoring depends on the quality extracted parameters from the acquired ECG signals[2][3]. Accuracy of the estimated ECG parameters highly depends on the distortion from various artifacts and noise[4]. Preprocessing algorithms which reduce the effect of such distortions are considered the heart of wearable technologies.

Though wavelet based thresholding techniques have shown effectiveness in reduction of noise, they rely on manual tuning of threshold for better classification accuracy[5]. In addition, majority of these algorithms rely on the accurate detection of R peaks from noise corrupted ECG signals. Accurate detection of R peak location is a challenging task in case the signal is affected by high frequency components from muscle movements. Though R peak detection algorithms have shown significant improvement with the use of methodologies derived from information theory, they lack validation on the dataset from wearable ECG signals[6]. Current study utilizes the advantage of EMD and NLM algorithms to reduce the effect of noise while overcoming the limitation with the use of complexity measure of sample entropy. The effectiveness of the proposed method is verified with various performance measures and comparative analysis.

Empirical Mode Decomposition (EMD) based denoising is well known for such requirements as it offers the decomposition of the original signal into intrinsic mode functions (IMF) which are relatively easier to analyze[7]. Usually during the denoising process, the high-frequency noise components are removed from the initial IMFs whereas the trend component is removed based on the last IMF for signal reconstruction. Though EMD based reconstruction is error-free, it suffers from mode mixing where the same oscillations appear in more than one IMF which is difficult to understand for its precise physical interpretation [8][9]. Noise-assisted method of Ensemble EMD (EEMD) was proposed by[10] to reduce the mode mixing phenomena by exploiting the property of white noise. This method adds a finite amplitude white noise to the original signal before the process of decomposition. The decomposition using EMD is repeated for different realizations of white noise. The ensemble mean of each IMF across all the realizations significantly reduces the effect of noise and guarantees the convergence to the original signal. The process of ensemble cancels the effect of added white noise. In addition, the dyadic nature of the IMF reduces the likelihood of mode mixing [10]. Though EEMD significantly reduces the effect of mode-mixing, the number of IMFs for the same signal with different realizations of the white noise may differ in number. Moreover, the reconstruction is also not perfect due to residual noise. A modified EEMD named Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) which provides a complete reconstruction of the original signal with lesser computational requirements was originally proposed in [11]. The algorithm is described in more detail in section II(A).

The IMFs from the CEEMDAN decomposition are

arranged in the order from higher frequency IMFs to lower frequency IMFs. Usually, the higher frequency IMFs are noise dominant whereas lower frequency IMFs contain information. Studies have shown various methods and techniques from information theory to identify such noise dominant IMFs which are then denoised to improve the denoising performance [10-11]. For better denoising, Xu et al. used a correlation of IMFs with the original signal to identify the noise dominant IMFs for denoising with the use of wavelet thresholding technique[13]. The correlation-based techniques for the identification of noisy IMFs are limited to capture the linear relationships [14]. Polat & Nour used EMD decomposition followed by suppression of insignificant coefficients of DWT which is then followed by run-length coding in order to achieve better compression of ECG data [15]. This method utilizes the EMD decomposition for categorization of noisy components and DWT-based dead zone quantization for reduction of noise. Kabir & Shahnaz chose windowing of the first three high-frequency IMFs for denoising using adaptive soft thresholding from DWT[16]. The adopted method assumes the initial three IMFs are noise dominant. This approach may degrade the denoising performance if the actual number of noisy IMFs is more than three. In a similar study from Rakshit & Das to achieve higher denoising performance, initial three IMFs are selected and denoised using adaptive switching mean filtering operation[17].

While presumption of fixed number of initial IMFs for denoising reduces the computation complexity, it limits the likelihood for further improvement in case the noise is present in the IMF beyond the considered IMF number. CEEMDAN has been used in combination with other denoising techniques to provide hybrid solutions for better denoising performance [18][13]. These techniques have shown effectiveness in reduction of noise, however, they do not significantly remove the noise from the ECG. In this article, we propose a method to identify the number of noisy IMFs based on the measure of sample entropy (SampEn) and denoise them using a non-local mean algorithm. SampEn is a measure of complexity of sequence that does not depend on the information about the source of complexity[19] and removes the dependency on knowledge of noise source. It works on the similarity of the patterns within a sequence. Higher the complexity of a sequence higher is the dissimilarity of the patterns within a sequence. As a consequence, the SampEn measure provides a more reliable way for identification of noisy IMFs. After the identification procedure, the NLM algorithm has been used to reduce the effect of noise [20]. This algorithm works on the principle of averaging similar patches within the neighborhood region. NLM has proven its denoising performance in the area of image processing[21][22][23] due to its relatively simpler structure with better performance [21].The level of smoothening depends on the parameters like size of the patch, size of the neighborhood and the bandwidth. This article uses adaptive adjustment of these parameters for each IMFs separately. The parameters for the NLM algorithm are adjusted based on the standard deviation and SampEn of the IMFs. Such adaptive adjustment of the parameter shows improved denoising performance for the noisy ECG signals. This article proposes the methodology that utilizes SampEn as a reliable measure for detection of noisy IMFs. NLM

smoothing is used to reduce the effect of noise from the detected noisy IMFs.

Section II describes ECG attributes, CEEMDAN algorithm, NLM algorithm, and SampEn in more detail. The proposed methodology is discussed in more detail in Section III. The performance of the proposed method is evaluated with the use of standard performance evaluation metrics and presented in Section IV. Finally, discussion and conclusion of the proposed approach are presented in Section V and Section VI respectively.

II. METHODOLOGY

A. ECG acquisition and attributes

ECG measurements are acquired by placing electrodes at specific locations on the surface of the body - the arms, legs and chest. The electrical depolarization and repolarization of the cardiac cells result in electrical potentials across these specific locations. Each depolarization creates a wave of electrical charge movement across the cell membrane to become less negative. The massive flow of the charge results in electrical current that can be detectable by the electrodes. Once the depolarization completes, the repolarization starts leading to reestablishment of the original condition again. Each cycle of depolarization and repolarization leads to development of varying electrical potential. These potentials are measured by the small electrode patches attached on the surface of the body. A typical pattern of ECG signal is shown in Fig. 1.
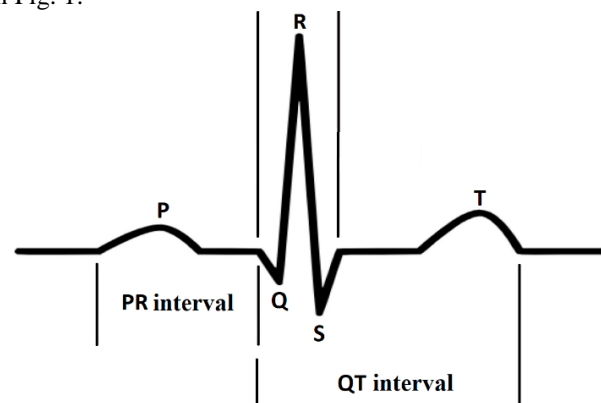

Fig. 1 Normal ECG signal and its component

TABLE I
Normal ECG parameters

Segment	Duration(ms)	Amplitude (mV)
P wave	80-100	0.05-0.25
PR interval	120-200	~
QRS complex	<120	1-1.5
QT interval	350-460	~
T wave	160	<0.5

The peak of the potential can be observed on the ECG signal called R peak. It represents the depolarization of the ventricles muscles. P waves and T waves represent the depolarization of atria and repolarization of ventricles respectively. Each of these waves occur at specific instances and specific durations for a healthy individual. Usual normal range of values for these parameters are shown in Table I. However, these parameters are affected by noise from i) power line interference 2) Electromyogram noise 3) baseline wander and 4) relative motion of electrodes on the body surface. Hence, the immediate post acquisition steps involve

signal denoising for parameter extraction to make further inference about cardiac health conditions.

B. CEEMDAN Algorithm

CEEMDAN is based on the EEMD procedure with an additional advantage of the removal of reconstruction error while maintaining the reduced effect of mode mixing. It also removes the problem of different numbers of modes that exist in EEMD by adaptively controlling the noise. The basis of the CEEMDAN is ensemble averaging of the IMFs added with white Gaussian noise. The addition or subtraction of the noise is adaptively controlled by the amplitude of signal or the residual of the signal during the decomposition process. This process gradually eliminates the reconstruction error by iterative procedure. The steps for the decomposition procedure are as follows [11].

1. Generate I number of noisy realizations of the original signal x(n) by adding white noise w(n) in the signal. The noisy realization of the signal is given by

$$x^i(n) = x(n) + \alpha_i\, w^i(n), \qquad i = 1,2,\dots,I \qquad (1)$$

Here, $w^i(n)$ is the ith realization of the white noise, $x^i(n)$ is referred to as the ith group of noisy signals. α_i is the parameter that controls the effect of the noise.

2. Decompose the noisy realizations with the use of the EMD algorithm. The first order IMF of $x(n)$ is obtained by performing the ensemble average of the first order IMF of all I realizations as follows

$$\overline{IMF_1}(n) = \frac{1}{I}\sum_{i=1}^{I} IMF_1^i(n) \qquad (2)$$

Where $IMF_1^i(n)$ is the first order IMF of ith realization of noisy $x(n)$.

3. Manipulate the first order residue by subtracting $\overline{IMF_1}(n)$ from the original signal $x(n)$

$$r_1(n) = x(n) - \overline{IMF_1}(n) \qquad (3)$$

Now, let the operator E$_i$(.) represent the ith order IMF of the signal.

4. Perform the decomposition of the realization $r_1(n) + \alpha_0 E_1(w^i(n)), i = 1,2,\dots,I$ till their first EMD mode to obtain second IMF of $x(n)$ as follows

$$\overline{IMF_2}(n) = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1(n) + \alpha_0 E_1(w^i(n))) \qquad (4)$$

5. Calculate kth order residue $r_k(n)$ as follows

$$r_k(n) = r_{k-1}(n) - \overline{IMF_k}(n), \qquad k = 2,3,\dots,K \qquad (5)$$

6. Perform the decomposition of $r_k(n) + \alpha_k E_k(w^i(n))$, $i = 1,2,\dots,I$ to obtain $k+1^{th}$ order IMF of $x(n)$

$$\overline{IMF_{k+1}}(n) = \sum_{i=1}^{I} E_1\left(r_k(n) + \alpha_0 E_k(w^i(n))\right) \qquad (6)$$

7. Repetition of Step 5 and Step 6 is performed until the situation where the further decomposition is not possible due to limitations of the condition of optima that more than one minimum or maximum must be present in the IMF. The final residue is calculated as

$$R(n) = x(n) - \sum_{k=1}^{K} \overline{IMF_k}(n) \qquad (7)$$

Here K is the total number of modes. Hence, the original signal is x(n) is expressed as

$$x(n) = R(n) + \sum_{k=1}^{K} \overline{IMF_k}(n) \qquad (8)$$

Equation (8) suggests the CEEMDAN provides an exact reconstruction of the signal in terms of IMFs and residue. Here α_i is used to control the SNR at each decomposition stage. In a study from Wu & Huang, it is suggested to use a small value of α_i for a signal dominated in high frequency whereas relatively larger value for a signal dominated in low frequency region[24]. On the other hand, Flandrin et al. suggested the same fixed value at each decomposition for a specific application[11]. In order to understand the decomposition from the CEEMDAN procedure, a simulated noisy signal of ECG is used and decomposed into 12 modes as shown in Fig. 2
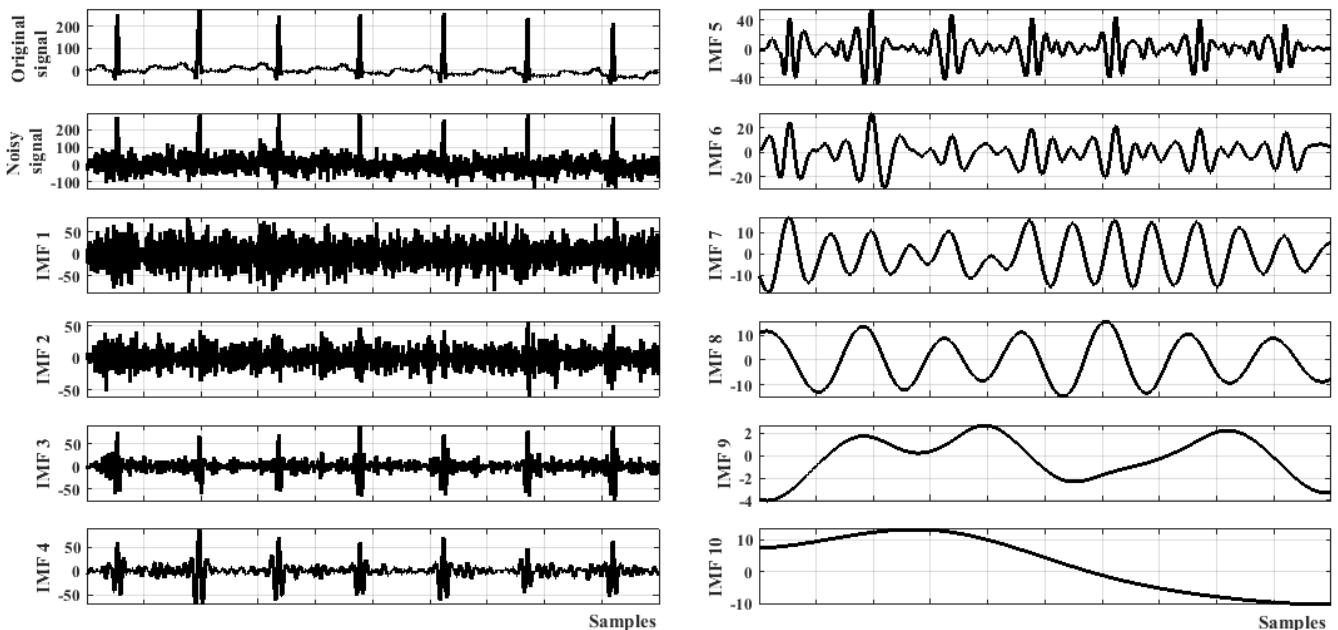

Fig. 2 Decomposition of ECG with SNR of 0 dB using CEEMDAN

A noise-free signal d(n) is shown in Fig. (2-a) whereas a noisy signal with SNR of 0 dB is shown in Fig. (2-b). The corresponding IMFs from 1 to 10 are shown in Fig. 2(c-l). Since CEEMDAN adds noise during the decomposition process, some noisy IMFs are generated as shown in Fig. 2(c-d). The number of noisy IMFs is not fixed and the determination requires the measure of information. In this article, the measure of sample entropy is used to identify such noisy IMF for denoising.

C. NLM Algorithm

Similar to the natural images, ECG signals are also of repetitive nature which makes the possibility for utilization of the NLM algorithm for denoising. Here, we use this algorithm considering application to one dimensional use case. The NLM algorithm tries to recover the true signal **d** from its noisy observations **x=d+n**, where **n** is additive white noise independent from true signal **d**. It estimates each true sample from its noisy observations by calculating the weighted average of the points in the predetermined search neighborhood. The algorithm assigns more weights when the observations are similar in the search neighborhood. If the observations are not similar, a relatively lesser weight is assigned while calculation of the average. In other words, similar observations are assigned higher weights in order to reduce the effect of noise. Averaging over N samples reduces the effect of noise by the factor of the square root of N. The estimate $\hat{d}(p)$ for the noisy sample $x(p)$ located at p in NLM is given by[20],

$$\hat{d}(p) = \frac{1}{\Omega(p)} \sum_{q \epsilon N(p)} w(p,q)x(q) \qquad (9)$$

where, $\Omega(p) = \sum_q w(p,q)$

$$w(p,q) = exp(-\frac{\sum_{\delta \epsilon P}(x(p+\delta) - x(q+\delta))^2}{2L_p\lambda^2})$$

Here p is the local samples around $x(p)$, λ is the bandwidth parameter and controls the level of smoothening, L_p is the patch width (2P+1) representing number of samples in the local sample p.

The graphical representation of the patches is shown in Fig. 3. Here p and q are the independent indexes representing the locations of patches of length L_p within $N(p)$. If the patches in the current neighborhood $N(p)$ are similar then $w(p,q)$ will be higher and vice versa. The size of the $N(p)$ is 2S+1. The larger $N(p)$ accommodates relatively more number of patches resulting in higher smoothening.
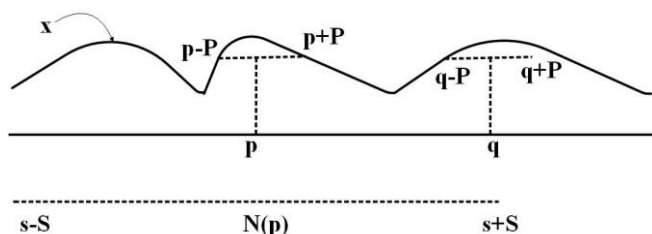

Fig. 3 Graphical representation of NLM parameters. The patch centered on p is compared with patches centered on q within the region of $N(p)$

One of the key advantages of the NLM algorithm is that the weighting $w(p,q)$ depends on the similarity of the patches and not on the location of p and q. It results in preservation of the edges of similar patches within the local structure $N(p)$ by the process of averaging which is not possible with typical filtering. Usually, larger $N(p)$ accommodates more similar patches resulting in more effective denoising. However, larger N scales to more computational requirements.

Parameter adjustment:

The NLM denoising requires the adjustment of three parameters that affect the filtering operation: the patch size (L_p), neighborhood sample size N(p), and bandwidth parameter λ. Generally, the patch size is termed by half-width R and the N(p) is specified as neighborhood half-width S. Based on this, the patch size and N(p) can be expressed as 2P+1 and 2S+1 respectively. Usually, the patch size is selected based on the smallest feature required to preserve while smoothening. For example, the patch size in the case of ECG denoising is selected as the size of the QRS complex[20], and N(p) is limited by the computational demand. Usually, a larger neighborhood half-width S improves the denoising performance, however, it is correlated and scaled with the computational demand. Larger S requires larger computational time. When N(p) is selected as the entire signal length, the algorithm considers the global comparison resulting in a truly non-local mean comparison. The bandwidth (λ) is an important parameter that controls the level of smoothening for different patches. Larger λ produces excessive smoothening resulting in different patches to look similar whereas smaller λ produces poor smoothening resulting in large fluctuations to remain present. Deliberate selection of λ is an important aspect of any effective denoising application. Unnecessary larger λ may damage the required details whereas unnecessary smaller λ may damage relatively larger features. In an investigation to identify overall reasonable choice of λ, Tracey & Miller proposed value of λ=0.6 for overall improved performance as inferred by the curve of SNRimpr vs. λ/σ [20].

One of the limitations of NLM denoising for ECG signals is relatively less averaging effect in the region of QRS-complex. Comparatively higher magnitude of the R wave from the nearby samples requires the smaller patch size limited to the R peak region for effective filtering operation. Larger patch size that includes the Q wave and S wave region results in low estimation of averaging effect. It is difficult to differentiate this region from bit-to-bit locations. The decomposition from CEEMDAN adaptively separates such high magnitude high frequency components available in different IMFs. Hence parameter adjustment for each IMFs based on specific attributes reduces the low averaging effect in the nearby region of R peak.

D. Sample Entropy

The sample entropy (SampEn) provides a measure of randomness without the need for the information about the source of randomness. Due to this key advantage, it finds applicability in several domains. Though it was initially developed for the use of physiological data analysis, it finds applications in medicine[25], physiology [26], earth science [27], device health monitoring [28], geology [29]. The larger the value of SampEn the more is the complexity or randomness in the time series and vice versa.

The method assesses time series for similar patches. Relatively frequent and similar patches result in lower value. In principle, SampEn (m,r,N) gives a measure of similarity on the basis that the two subseries of length m remain similar (within the tolerance set as ±r times the time series' standard

deviation) to the subseries of length m+1. The values closer to zero indicate low complexity and higher regularity in the time series. On the other side, higher values indicate relatively more randomness or the presence of significant amount of Gaussian noise [30]. Essentially, SampEn is the negative of the natural logarithm of conditional probability that two subseries with length m remain similar for length m+1. Unlike approximate entropy measure, SampEn does not consider the self-matching and improves the reliability of the measure by reducing potential bias. This makes SampEn an unbiased measure and eliminates its dependency on self-matching.

Following steps are performed to calculate the SampEn of a given time series[30].

1. Let X represent the time series of uniformly sampled N elements as follow
$$X = \{x(1), x(2), \ldots, x(N)\}$$

2. X is divided into equally sized m dimensional vectors X(1), X(2), ..., X(N-m+1).
    where $X(i) = \{x(i), \ x(i+1), \ldots, x(i+m-1)\}$,
    $$i = 1,2,\ldots,N-m+1$$
    here, m is referred to as an embedded dimension with typical values of 1 or 2.

3. Assign the distance d_{ij} between X(i) and X(j) as follow
    $$d_{ij} = max(|X(i+k) - X(j+k)|),$$
    $$i, j = 1,2,\ldots,N-m+1 ; i \neq j$$
    $$k = 0,1,2,\ldots,m-1$$

4. Calculate the number of d_{ij} less than r and consider it as B_i. Here r is the distance parameter. Usually, it is selected between 0.1σ to 0.25σ. B_i represent the number of m dimensional vectors X(i) closer to X(j) by a distance of r. Normalize B_i by the number of vectors as follow
    $$B_i^m(r) = \frac{B_i}{N-m}, i = 1,2,\ldots,N-m+1; i \neq j$$

5. Calculate the average value of $B_i^m(r)$ as follow
    $$B^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} B_i^m(r) \qquad (10)$$

6. Increase the value of m by 1 and calculate $B^{m+1}(r)$ by following the steps from 2 to 5.

7. The value of sample entropy is calculated as follows
    $$SampEn(m,r,N) = \ln(B^m(r)) - \ln(B^{m+1}(r)) \quad (11)$$

Here, the value of SampEn depends on the selection of the parameters m,r and N. Generally, the statistical property of the SampEn is in line with the statistical property of signal when N>500. In this article, N=2000, m=2, and r=0.25σ are selected for the analysis.

III. PROPOSED METHOD

In this article, a novel method for denoising ECG signals is proposed which utilizes the measure of SampEn to identify the noisy IMF which are subsequently smoothened using the NLM algorithm. IMF is categorized as noisy based on the value and the specific pattern of SampEn. As discussed earlier, it is a measure of the orderly behavior of the time series. Usually, higher-order IMFs are information-rich and the corresponding SampEn decreases due to the repetitive pattern of the ECG signals. In order to identify the number of initial IMFs with dominant effect of noise, the IMF number is identified from which the first three consecutive decrease of SampEn values are observed. Three consecutive samples of SampEn are considered to avoid the ambiguity with two samples due to possible fluctuations from the influence of

noise as shown in Fig. 4. The IMFs after order two are observed with a consecutive decrease in SampEn which corresponds to normal orderly behavior which is in line with the theory of EMD. Consecutive decrease in the SampEn values represent orderly behavior in subsequent IMFs and reduced effect of noise. After the selection of noisy IMF, they are smoothened with the use of the NLM algorithm as discussed in Subsection II. Reconstruction from CEEMDAN decomposition is error-free, hence smoothening of higher-order noisy IMF reduces the effect of high-frequency noise adaptively. Eventually, the smoothened IMFs are combined with the other IMFs to get the overall denoised signal. The block-diagram representation of the algorithm is shown in Fig. 5.

The steps of the proposed algorithm are described below.

1. Calculate the IMFs of the signal which are to be processed by the CEEMDAN algorithm as described in Section II.

2. Calculate the SampEn of each IMF as described in Section 2.3. Select the parameters m=1 or 2, r=0.25σ of the selected IMF.

3. Arrange the SampEn values in accordance with the corresponding IMF number.

4. Identify the IMF number from which the SampEn consistently decreases for at least three consecutive samples. Consider each IMF before the start of such monotonous decrease in the SampEn as the noisy IMF.

5. Select the noisy IMF for smoothening using the NLM algorithm. Select the parameter of NLM smoothing based on the statistical property of the IMF as discussed in Section II.

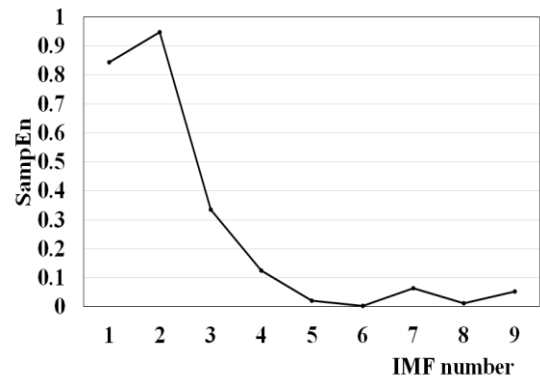6. Combine the smoothened IMF with the other IMFs to reconstruct the denoised ECG signal.

Fig. 4 Sample entropy of IMFs from CEEMDAN

IV. RESULT

A dataset from the MIT-BIH repository is used to evaluate the effectiveness and robustness of the proposed method. The performance of the algorithm is evaluated based on the following standard indicators.

a) Signal to noise ratio (SNR)
    Signal to noise ratio is defined as follows
    $$SNR_{dB} = 10\log\frac{\sum_{i=1}^{N} d(n)^2}{\sum_{i=1}^{N}(\hat{x}(n) - d(n))^2},$$
    Here, $\hat{x}(n)$ is the denoised signal, d(n) is the noise-free signal, N is the length of the signal. In the case of real measurements, d(n) is considered as the source data. Though the measured data contains noise, it facilitates the comparison of performance metrics for denoising

algorithms. Higher the SNR better is the denoising performance.

b) Improvement in SNR

Improvement in SNR is defined as follows

$$SNR_{impr}(dB) = 10log\frac{\sum_{i=1}^{N}(x(n)-d(n))^2}{\sum_{i=1}^{N}(\hat{x}(n)-d(n))^2}$$

c) Root mean square error (RMSE)

RMSE is defined as following

$$RMSE(\hat{x},d) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{x}(n)-d(n))^2}$$

Lower the RMSE better is the estimation performance of the method.

d) Improvement in RMSE

Improvement in RMSE is defined as follows

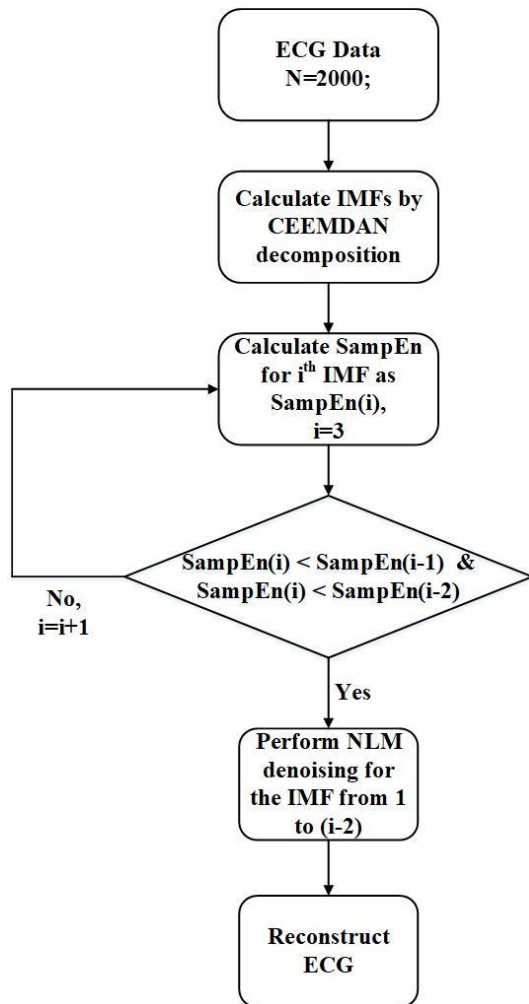$$RMSE_{impr} = (\frac{RMSE(x,d)-RMSE(\hat{x},d)}{RMSE(x,d)})$$

Fig. 5 Block-diagram representation of the proposed method

A. Analysis of ECG with Normal Sinus Rhythm

In this analysis, the recorded ECG data of arrhythmia from MIT-BIH is considered [30]. The database consists of ECG recording of 47 subjects for the duration of 30 minutes at the sampling rate of 360 Hz. The datasets named 100m, 101m, 102m, 117m, and 118m are considered for the denoising performance analysis of the proposed algorithm. The excerpt of consecutive normal sinus rhythm for at least 10 bits are

considered here. Fig. 6(a) shows the excerpt from 100m with a continuous line in black and noisy data (with 5 dB SNR) with a dotted line in black. Fig. 6(b-d) shows denoising using CEEMDAN, NLM, and the proposed method respectively. Though the real data also contains noise, it is considered to be the original signal for comparative analysis. Here, the parameters for the NLM algorithm are tuned considering the same methodology as described in [20]. The mean of SNRimpr versus λ/σ for different noise levels is shown in Fig. 7. It is observed that relative increase in the bandwidth parameter for different input SNR produces variations in SNRimpr. Improvement in output SNR is observed when input SNR is poor, particularly when λ/σ is in the range between 0.2 to 0.5 sigma. The highest improvement of almost 10 dB can be seen in the case when input SNR is -5 dB. On the other side, the decrement in SNRimpr is observed for the cases when input SNR is more than 5 dB. It is indicative of deterioration of SNRimpr when the signal is relatively less affected by the noise. In order to achieve higher improvement for the low SNR, λ=0.5σ is selected for the NLM algorithm. The same value of λ is chosen for the proposed method for comparative analysis of the results. The change in the other two parameters, local patch size (P) and the neighborhood (M), do not result in a significant change in the SNRimpr. Hence, similar values as from Tracey and Miller, M=2000 and P=10, are selected [20]. The same values of the parameters are used for smoothening of noisy modes from CEEMDAN in the proposed method except for the parameter bandwidth λ which is separately tuned to adjust for the large fluctuations within the modes. Through visual inspection from Fig. 6, it is evident that denoising using the NLM algorithm and the proposed method are better compared to denoising using CEEMDAN alone. For quantitative evaluation of performance, Table II shows improvement in SNR and improvement in RMSE for the selected excerpt of the data presented in Fig. 6. The SNRimpr from the CEEMDAN method, NLM algorithm, and the proposed method is 4.51, 5.50, and 7.89 respectively. The proposed method shows 74.9% and 43.4% better improvement in the SNR compared to CEEMDAN and NLM algorithms, respectively. The RMSE is calculated by considering the error between the original signal and the denoised signal. The RSME of the noisy signal is 0.21 whereas the RSME of the denoised signal from the proposed method is 0.08. The percentage improvement in the RMSE by the proposed method is 33.3% and 27.3% higher compared to that observed for CEEMDAN and NLM algorithms alone respectively.

TABLE II
Results of performance for denoising methods with input SNR of 5 dB

Algorithm	SNR of a noisy signal (dB)	SNR_{impr} (dB)	RMS error of a noisy signal	RMS error of a denoised signal	Improve-ment in RMSE (%)
CEEMDAN	5	4.51	0.21	0.12	42.8
NLM	5	5.50	0.21	0.11	47.6
Proposed method	5	7.89	0.21	0.08	61.9

Now, the analysis of the entire data set is considered to evaluate the performance of the proposed algorithm at different levels of input SNR. The input SNR is controlled by the noise amplitude. In order to consider the performance at the population level, the mean and standard deviation of the performance metric are calculated. Fig. 8 shows the mean of SNRimpr with different levels of input SNR.
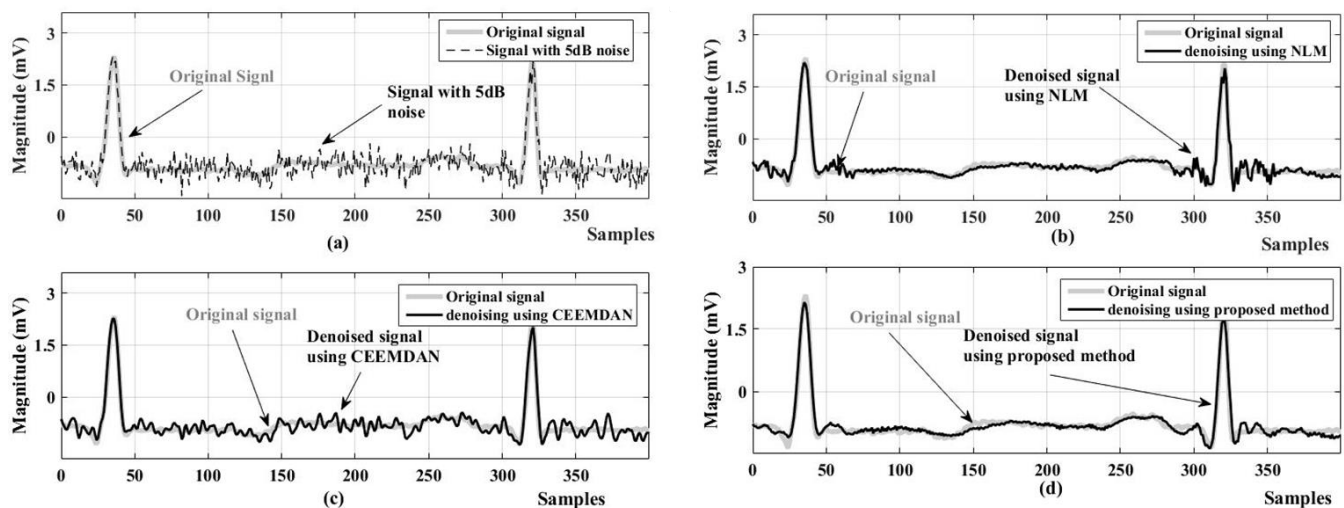
Fig. 6 (a) Original signal and the noisy signal (b) denoising using NLM algorithm (c) denoising using CEEMDAN (d) denoising using the proposed method
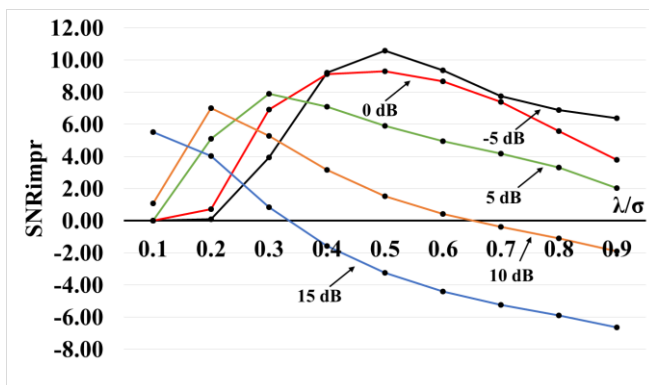

Fig. 7 Variation in SNRimpr with parameter λ for a different level of input SNR
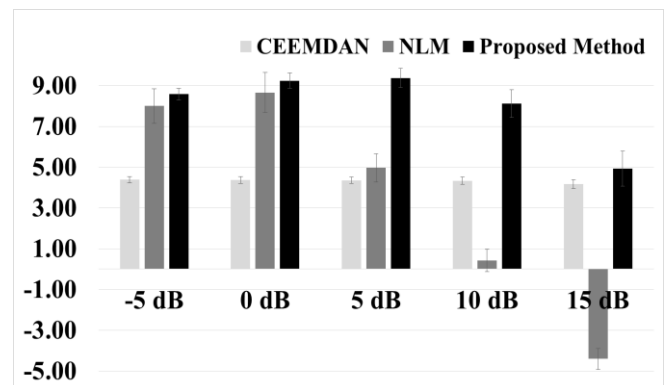

Fig. 8 Mean of Improvement in output SNR with different levels of input SNR

It is observed that the mean of SNRimpr from the proposed method is 8.58 for -5 dB of input SNR which is 48.9% and 6.7% higher than CEEMDAN and NLM algorithms respectively. Similarly, it is 52.8% and 6.3% higher for 0 dB, 53.5%, and 47.1% higher for 5 dB, 46.6%, and 94.7% for 10 dB of input SNR compared to the denoising using CEEMDAN and NLM algorithm respectively. In addition, the proposed method shows improved denoising in case of 15dB of input SNR compared to NLM denoising alone. NLM denoising produced deteriorated quality of original ECG signal beyond 15dB of input SNR. The percentage improvement in the output SNR reduces with the increase in the input SNR of the signal. The proposed algorithm provides comparatively better performance when the signal is having a large influence of noise. Table III shows the improvement in RMSE for different levels of input SNR. The observed improvement in RMSE at -5dB of input SNR from the proposed method is 63.0% which is 36.8% and 4.5% higher compared to improvement in RMSE from CEEMDAN and NLM denoising methods respectively. Similarly, it is 39.6% and 4.2% higher for 0 dB, 40.2%, and 52.2% higher for 5 dB compared to the CEEMDAN and NLM denoising methods respectively. The negative value of improvement from NLM denoising for 15 dB input SNR shows an increase in distortion for the set values of the parameters.

TABLE III
Improvement in RMSE with different input SNR

Algorithm	-5 dB	0 dB	5 dB	10 dB	15 dB
CEEMDAN	0.40	0.40	0.39	0.39	0.38
NLM	0.60	0.63	0.43	0.05	-0.66
Proposed	0.63	0.65	0.66	0.61	0.43

B. Analysis of ECG with Atrial Fibrillation

Atrial fibrillation, often known as AFib or AF prevails in the majority of cardiac arrhythmias. An arrhythmia is a condition in which the heart beats in an irregular pattern, either too slowly or too rapidly. When a person has AF, their heart's atria (two upper chambers) beat irregularly, resulting in inadequate blood flow from atria to the ventricles (the lower two chambers). Because of the uneven beat, the ventricles do not have enough time to fill completely which results in inadequate blood flow to the lungs and other regions of the body. This scenario can sometimes results in the production of blood clots, which can raise the risk of a brain stroke. Almost one third population with AF is asymptomatic. If the detection and treatment is not started in the early stage, it may result in formation of blood clots increasing the risk of brain stroke.

C. Analysis of ECG with Atrial Fibrillation

Atrial fibrillation, often known as AFib or AF prevails in the majority of cardiac arrhythmias. An arrhythmia is a condition in which the heart beats in an irregular pattern, either too slowly or too rapidly. When a person has AF, their heart's atria (two upper chambers) beat irregularly, resulting in inadequate blood flow from atria to the ventricles (the lower two chambers). Because of the uneven beat, the ventricles do not have enough time to fill completely which results in inadequate blood flow to the lungs and other regions of the body. This scenario can sometimes results in the production of blood clots, which can raise the risk of a brain stroke. Almost one third population with AF is asymptomatic. If the detection and treatment is not started in the early stage, it may result in formation of blood clots increasing the risk of brain stroke.

Here, we consider the ECG data containing the episodes of AF. The frequency content and the time domain characteristics of ECG during AF condition is different from the normal sinus rhythm [31]. Fig. 9 shows the typical pattern of ECG recording with AF case. It is evident that the RR interval as well as the shape of different waveforms (P,Q,R,S) have changed leading to variations in parameters. Precise estimation of the ECG parameters during this condition requires minimum interference from noise. Hence, publicly available MIT-BIH atrial fibrillation dataset [32] is used to test the effectiveness of the proposed denoising methodology against different levels of input noise. The dataset consists of two ECG signals recorded for 10 hours duration at a sampling of 250 Hz with resolution within 10 millivolt of range. Each rhythm is manually annotated for all 23 recordings. Here, we have considered the portion of the recording where an episode of atrial fibrillation is present and analyzed the performance of the proposed algorithm. Table IV shows the comparison of the performance of the proposed method for the case of 4 dB SNR. The proposed method shows almost 53.4% and 36% increase in the output SNR compared to CEEMDAN and NLM algorithms alone.

TABLE IV
Results of performance for denoising methods with input SNR of 4 dB

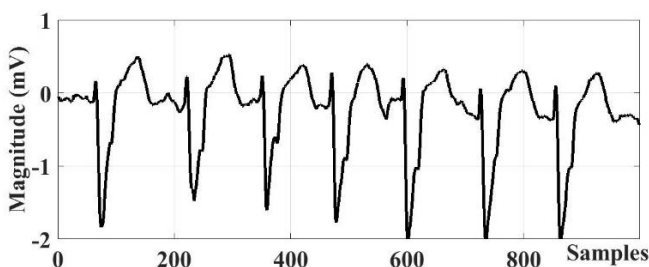| Algorithm	SNR of a noisy signal (dB)	SNR_{impr} (dB)	Improvement in RMSE (%)
CEEMDAN	4	4.27	38.9%
NLM	4	5.03	43.9%
Proposed method	4	7.90	59.7%

Fig. 9 Sample ECG with AF episode

Further analysis considering different input SNR is shown in Table V. The proposed method has shown 26.1% and 21.0% performance improvement in the presence of -5 dB input SNR compared to CEEMDAN and NLM algorithm alone respectively. Similarly, almost 75% and 25.4% improvement is observed in the presence of 5 dB input SNR compared to

CEEMDAN and NLM respectively. An improvement of performance with 57.7% (compared to CEEMDAN alone) and 10% (compared to NLM alone) is observed in the presence of 10dB of input SNR. The proposed algorithm shows comparable performance with NLM smoothening when the input SNR is 15 dB.

TABLE V
Improvement in output SNR with different input SNR

Algorithm	-5 dB	0 dB	5 dB	10 dB	15 dB
CEEMDAN	4.28	4.26	4.27	4.12	3.65
NLM	4.46	7.27	5.98	5.91	6.35
Proposed method	5.40	8.84	7.50	6.50	6.37

Next, we analyze the performance comparison considering the measure of improvement in the RMSE for the selected AF dataset. Table VI shows the improvement in RMSE for different levels of input SNR. The observed improvement in RMSE at -5dB of input SNR from the proposed method is 45.0% which is 18.3% and 7.1% higher compared to improvement in RMSE from CEEMDAN and NLM denoising methods respectively. Similarly, it is 64.1% and 12.3% higher for 0 dB, 48.7%, 16.0% higher for 5 dB, and 36.8% and 6.1% higher for 10 dB compared to the CEEMDAN and NLM denoising methods respectively.

TABLE VI
Improvement in RMSE with different input SNR

Algorithm	-5 dB	0 dB	5 dB	10 dB	15 dB
CEEMDAN	0.38	0.39	0.39	0.38	0.34
NLM	0.42	0.57	0.50	0.49	-0.16
Proposed method	0.45	0.64	0.58	0.52	-0.03

D. Verification of the Proposed Method for Improvement in the Performance

The proposed method selects the IMF for denoising based on the specific pattern of the SampEn across IMFs. SampEn is calculated for each IMF by the normalization procedure as follows.

$$\text{SampEn}_{norm} = \frac{\text{SampEn} - SampEn_{min}}{\text{SampEn}_{max} - \text{SampEn}_{min}}$$

Here, SampEn_{norm} is the normalized value of SampEn across IMFs from a particular decomposition, $SampEn_{min}$ and SampEn_{max} are the minimum and maximum SampEn across IMFs from a particular decomposition respectively. Fig. 10 shows the behavior of SampEn and SNRimpr with IMF number. It contains the average value of SNRimpr (on y1 axis), normalized SampEn (on y2 axis), and IMF number (on the x-axis). The input SNR of 0 dB, 5 dB, 10 dB, and 15 dB are selected to analyze behavior of SNRimpr with denoising from the proposed method as shown in Fig. 10 (a)-(d) respectively.

It is observed that the SNRimpr increases with the order of the IMF for initial few IMFs. For the case with input SNR of 0 dB, the SNRimpr increases with smoothening by NLM till the inclusion of IMFs from one to four; afterward it gradually decreases with the inclusion of further modes of IMFs. In Fig. 10(a), the improvement in the SNR is observed till denoising of IMFs up to fourth level. Further smoothening beyond the fourth IMF decreases the denoising performance as observed from the decreased SNRimpr. The corresponding

SampEn increases up to the third IMF representing a possible high influence of noise. A significant drop in SampEn is observed beyond the third IMF. A similar trend is also observed for the input SNR of 5 dB as shown in Fig. 10(b), however, the corresponding SampEn starts to drop after the second IMF. Almost similar behavior of SampEn is also observed for the input SNR of 10 dB and 15 dB as shown in Fig. 10(c-d), however, the decrease in the SampEn starts from the first IMF in this case.
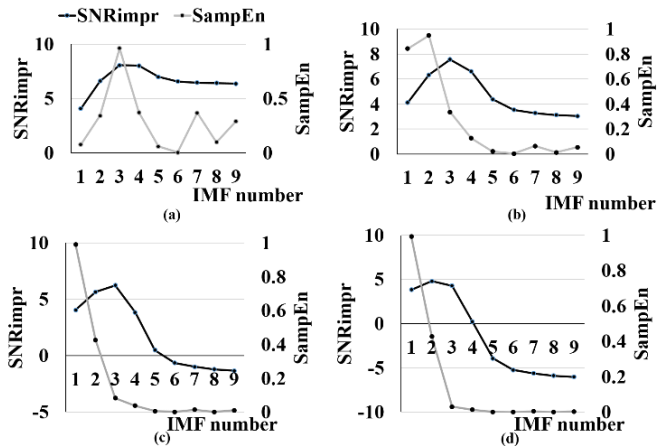

Fig. 10 Analysis of Improvement in SNR with denoising IMFs (a) Improvement in SNR with 0 dB input noise-y1 axis, SampEn of IMFs-Y2 axis (b)Improvement in SNR with 5 dB input noise-y1 axis, SampEn of IMFs-Y2 axis (c)Improvement in SNR with 10 dB input noise-y1 axis, SampEn of IMFs-Y2 axis (d)Improvement in SNR with 15 dB input noise-y1 axis, SampEn of IMFs-Y2 axis

It should be noted that the increase in SNRimpr is not significant compared to the cases with input SNR of 0 dB and 5 dB. Moreover, SampEn has also become almost zero when denoising is used from the fourth IMF onwards as observed from Fig. 10(c) and Fig. 10(d). It is observed that the maximum mean of SNRimpr for 0 dB, 5 dB, 10 dB, and 15 dB is 8.1, 7.6, 6.2, and 4.8 respectively. It decreases with increased input SNR. Intuitively, a signal with high SNR contains less noise which results in a corresponding drop for further improvement. It is also observed that the mean of SNRimpr drops significantly with smoothening of more IMFs for higher input SNR as inferred from Fig. 10(c) and Fig. 10(d). The negative values represent the start of distortion if the denoising is continued for the higher order IMFs. Hence, it is important to decide the proper choice of a number of IMFs for denoising. This study utilizes the specific pattern of consecutive decrease in SampEn. The IMFs after such a pattern are considered as information-IMF and excluded from the step of smoothening with the use of the NLM algorithm. For example, the number of initial IMFs, identified to have dominant noise, are 3, 2, 1, and 1 as observed from Fig. 10 (a-d) respectively.

V. DISCUSSION

This article describes the novel method of ECG denoising by smoothening of noise corrupted IMFs from the CEEMDAN procedure with the use of the well-known NLM algorithm. The CEEMDAN decomposition results in initial few IMFs with dominant noise components. These IMFs lie usually in the lower modes for ECG signals. In order to identify such noise dominant IMFs, a particular pattern in the values of SampEn is identified. The IMFs with larger SampEn are

considered to be noisy till the drop in SampEn is observed for three consecutive IMFs.

Here, the noise is assumed to have additive white Gaussian nature. The proposed method utilizes the NLM algorithm which can effectively reduce such noise by averaging the similar patches of a signal over a predefined neighborhood. In order to verify the selected number of noisy IMFs, SNRimpr versus IMF number is analyzed. The improvement in the SNR output becomes insignificant after the identified drop pattern in SampEn. In addition, it is also observed that the initial number of IMFs required for efficient denoising are more for lower input SNR and less for higher input SNR. It suggests the requirement of denoising based on the level of noise in the signal instead of denoising of the fixed number of first few initial IMFs. The effectiveness of the proposed method is tested with the datasets comprising normal sinus rhythm (without abnormality in ECG) as well as rhythm with atrial fibrillation (abnormal ECG patterns). Both these rhythms have different attributes and variations in power spectral density. The methodology produced almost comparable performance for both the cases at different input SNR.

The proposed method shows superior performance compared to denoising using NLM or CEEMDAN alone. The analysis from the performance index indicates the usefulness of the method for ECG with high level of contamination from the noise. Generally, the signals from wearable ECG applications endured from high interference from the muscle noise and movement artifacts that require denoising due to poor SNR. Incorporation of such state of the art denoising methodologies can improve the estimates of the health parameters. Though the method has better denoising performance, it requires additional computational resources to accommodate the extra layers of processing. Hence, it is beneficial to use this method in the scenarios of high noise contamination.

VI. CONCLUSION

This article presents a novel ECG denoising method based on the selective smoothening of IMFs with the use of NLM denoising. The selection of noisy IMF is based on the specific pattern of the SampEn measure. The proposed method of denoising provides better performance compared to CEEMDAN and NLM denoising methods alone. The comparative analysis suggests that the proposed method is a suitable choice for denoising ECG signal especially when the signal has low SNR i.e. when the signal is under the heavy influence of noise. Though this method improves the denoising performance from either NLM or CEEMDAN alone, it comes at an additional cost of computational requirements for denoising. Apart from the additional computational requirement, it also uses the measure of SampEn which acts as a decision layer between CEEMDAN and NLM techniques. The future study will investigate how to render faster computational methods as well as better methodologies from an information theory perspective to reduce these shortfalls.

REFERENCES

[1] R.M.Rangayyan, *Biomedical Signal Analysis*. John Wiley & Sons, 2015.
[2] I. Zagan, V. G. Gaitan, A. I. Petrariu, and A. Brezulianu, "Healthcare IoT m-greenCARDIO remote cardiac monitoring system - concept, theory of operation and implementation," *Adv. Electr. Comput. Eng.*,

vol. 17, no. 2, pp. 23–30, 2017, doi: 10.4316/AECE.2017.02004.

[3] L. C. Chen, S. K. Swee, and T. S. Chiang, "Assessment of stroke severity level using emg, eeg and ecg for virtual rehabilitation," *Eng. Lett.*, vol. 29, no. 2, pp. 438–451, 2021.

[4] A. Secerbegovic, A. Gogic, N. Suljanovic, M. Zajc, and A. Mujcic, "Computational balancing between wearable sensor and smartphone towards energy- efficient remote healthcare monitoring," *Adv. Electr. Comput. Eng.*, vol. 18, no. 4, pp. 3–10, 2018, doi: 10.4316/AECE.2018.04001.

[5] B. Arvinti, A. Isar, and M. Costache, "Effective wavelet algorithm for an automatic detection of atrial fibrillation," *Adv. Electr. Comput. Eng.*, vol. 20, no. 3, pp. 49–56, 2020, doi: 10.4316/AECE.2020.03006.

[6] M. Z. Suboh, R. Jaafar, N. A. Nayan, and N. H. Harun, "Shannon energy application for detection of ECG R-peak using bandpass filter and stockwell transform methods," *Adv. Electr. Comput. Eng.*, vol. 20, no. 3, pp. 41–48, 2020, doi: 10.4316/AECE.2020.03005.

[7] N. E. Huang *et al.*, "The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998, doi: 10.1098/rspa.1998.0193.

[8] Y. Gao, G. Ge, Z. Sheng, and E. Sang, "Analysis and solution to the mode mixing phenomenon in EMD," *Proc. - 1st Int. Congr. Image Signal Process. CISP 2008*, vol. 5, pp. 223–227, 2008, doi: 10.1109/CISP.2008.193.

[9] X. Dao, M. Gao, and C. Cheng, "Mode adaptive demixing based on EMD in FMCW system," *Eng. Lett.*, vol. 27, no. 1, pp. 226–233, 2019.

[10] K. M. Chang, "Ensemble empirical mode decomposition: A Noise-Assited," *Biomed. Tech.*, vol. 55, no. 1, pp. 193–201, 2010.

[11] P. Flandrin, E. Torres, and M. A. Colominas, "A COMPLETE ENSEMBLE EMPIRICAL MODE DECOMPOSITION Laboratorio de Se ˜ nales y Din ´ amicas no Lineales , Universidad Nacional de Entre R ´ Laboratoire de Physique (UMR CNRS 5672), Ecole Normale Sup ` erieure de Lyon , France," pp. 4144–4147, 2011.

[12] S. Zhang *et al.*, "An Adaptive CEEMDAN Thresholding Denoising Method Optimized by Nonlocal Means Algorithm," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6891–6903, 2020, doi: 10.1109/TIM.2020.2978570.

[13] Y. Xu, M. Luo, T. Li, and G. Song, "ECG signal De-noising and baseline wander correction based on CEEMDAN and wavelet threshold," *Sensors (Switzerland)*, vol. 17, no. 12, 2017, doi: 10.3390/s17122754.

[14] J. Van Loco, M. Elskens, C. Croux, and H. Beernaert, "Linearity of calibration curves: Use and misuse of the correlation coefficient," *Accredit. Qual. Assur.*, vol. 7, no. 7, pp. 281–285, 2002, doi: 10.1007/s00769-002-0487-6.

[15] K. Polat and M. Nour, "Epileptic Seizure Detection Based on New Hybrid Models with Electroencephalogram Signals," *Irbm*, vol. 41, no. 6, pp. 331–353, 2020, doi: 10.1016/j.irbm.2020.06.008.

[16] M. A. Kabir and C. Shahnaz, "Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains," *Biomed. Signal Process. Control*, vol. 7, no. 5, pp. 481–489, 2012, doi: 10.1016/j.bspc.2011.11.003.

[17] M. Rakshit and S. Das, "An efficient ECG denoising methodology using empirical mode decomposition and adaptive switching mean filter," *Biomed. Signal Process. Control*, vol. 40, pp. 140–148, 2018, doi: 10.1016/j.bspc.2017.09.020.

[18] M. Blanco-Velasco, B. Weng, and K. E. Barner, "ECG signal denoising and baseline wander correction based on the empirical mode decomposition," *Comput. Biol. Med.*, vol. 38, no. 1, pp. 1–13, 2008, doi: 10.1016/j.compbiomed.2007.06.003.

[19] A. Delgado-Bonal and A. Marshak, *Approximate entropy and sample entropy: A comprehensive tutorial*, vol. 21, no. 6. 2019.

[20] B. H. Tracey and E. L. Miller, "Nonlocal means denoising of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2383–2386, 2012, doi: 10.1109/TBME.2012.2208964.

[21] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," *Proc. - 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005*, vol. II, no. 0, pp. 60–65, 2005, doi: 10.1109/CVPR.2005.38.

[22] J. Yang *et al.*, "Local statistics and non-local mean filter for speckle noise reduction in medical ultrasound image," *Neurocomputing*, vol. 195, pp. 88–95, 2016, doi: 10.1016/j.neucom.2015.05.140.

[23] C. K. Toa, K. S. Sim, Z. Y. Lim, and C. P. Lim, "Magnetic resonance imaging noise filtering using adaptive polynomial-fit non-local means," *Eng. Lett.*, vol. 27, no. 3, pp. 527–540, 2019.

[24] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2009, doi: 10.1142/S1793536909000047.

[25] D. E. Lake, J. S. Richman, M. Pamela Griffin, and J. Randall Moorman, "Sample entropy analysis of neonatal heart rate variability," *Am. J. Physiol. - Regul. Integr. Comp. Physiol.*, vol. 283, no. 3 52-3, pp. 789–797, 2002, doi: 10.1152/ajpregu.00069.2002.

[26] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy maturity in premature infants Physiological time-series analysis using approximate entropy and sample entropy," *Am. J. Physiol. Hear. Circ. Physiol.*, vol. 278, pp. H2039–H2049, 2000.

[27] D. Wang *et al.*, "Sample entropy-based adaptive wavelet de-noising approach for meteorologic and hydrologic time series," *J. Geophys. Res.*, vol. 119, no. 14, pp. 8726–8740, 2014, doi: 10.1002/2014JD021869.

[28] A. Widodo, M. C. Shim, W. Caesarendra, and B. S. Yang, "Intelligent prognostics for battery health monitoring based on sample entropy," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11763–11769, 2011, doi: 10.1016/j.eswa.2011.03.063.

[29] D. Mourenas, A. V. Artemyev, and X. -J. Zhang, "Dynamical Properties of Peak and Time-Integrated Geomagnetic Events Inferred From Sample Entropy," *J. Geophys. Res. Sp. Phys.*, vol. 125, no. 2, 2020, doi: 10.1029/2019ja027599.

[30] J. S. Richman, J. Randall Moorman, J. Randall, and M. Physi, "Downloaded from www.physiology.org/journal/ajpheart by ${individualUser," *Am J Physiol Hear. Circ Physiol*, vol. 278, pp. 2039–2049, 2000, [Online]. Available: www.physiology.org/journal/ajpheart.

[31] P. Melzi *et al.*, "Analyzing artificial intelligence systems for the prediction of atrial fibrillation from sinus-rhythm ECGs including demographics and feature visualization," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, 2021, doi: 10.1038/s41598-021-02179-1.

[32] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," *Circulation*, vol. 101, no. 23, 2000, doi: 10.1161/01.cir.101.23.e215.