

Received August 26, 2020, accepted September 9, 2020, date of publication September 14, 2020, date of current version September 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3023915

Studying the Effects of Compression in EEG-Based Wearable Sleep Monitoring Systems

DELAND HU LIU¹, (Student Member, IEEE), AND
SYED ANAS IMTIAZ², (Senior Member, IEEE)

¹Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78705, USA

²Wearable Technologies Laboratory, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K.

Corresponding author: Syed Anas Imtiaz (anas.imtiaz@imperial.ac.uk)

ABSTRACT Long-term sleep monitoring through the use of wearable EEG-based systems generates large volumes of data that need to be either locally stored or wireless transmitted. Compression of data can play a vital role to reduce the power consumption of these already resource-constrained systems. While compression methods can result in significantly reduced data storage and transmission requirements, the loss in signal information can have an impact on the algorithms used to extract the key sleep parameters. This paper studies the impact of six different state-of-the-art compression methods, including wavelet, SPIHT, filter and predictor-based methods, analysing their effects on the reconstructed signal quality particularly for automatic sleep staging applications. It looks at how the overall sleep staging accuracy as well as the detection accuracy of different sleep stages is reduced as a result of different EEG compression methods. It shows that the SPIHT and predictor-based compression methods outperform wavelet and filter-based methods in preserving the relevant signal features. It also shows that compression ratios of up to 65 can be achieved using the QSPIHT method with less than 10% loss in overall sleep staging accuracy.

INDEX TERMS Sleep disorders, electroencephalogram (EEG) compression, lossy compression, wearables, brain monitoring, sleep staging.

I. INTRODUCTION

Sleep disorders can severely affect the quality of life of those suffering from them, leading to abnormal sleep patterns that interfere with physical, mental and emotional functioning [1]. These disorders increase risks of diabetes, cardiovascular disease, stroke, and often correspond to increase in levels of mental distress and suicidal intentions [2]–[4]. It is estimated that more than 70 million adults suffer from different types of sleep disorders (e.g. insomnia) in the United States. Among those, at least 40 million suffer from long-term sleep disorders [5].

Diagnosis of sleep disorders is generally performed in specialized sleep clinics using an overnight sleep study known as polysomnography (PSG). This involves monitoring of brain activities using electroencephalogram (EEG), eye movements using electrooculogram (EOG), muscle movements using electromyogram (EMG), as well as other physiological parameters such as respiration and heart rate. These signals are then analyzed in blocks of 30-second epochs,

and each epoch is assigned one of the five stages of sleep (Wake, N1, N2, N3, REM) based on the criteria defined by the American Academy of Sleep Medicine (AASM) [6]. This process is known as sleep stage scoring or sleep staging and is an essential part of the diagnosis of various sleep disorders.

While PSG provides useful insights into sleep that are helpful for diagnosis, only a limited number of PSG recordings can be performed in sleep clinics. This is due to the cost of the study, the time taken to perform the analysis, as well as the limited number of specialized sleep clinics where these studies can be performed [7]. In addition, since most sleep studies are performed in unfamiliar environments, such as a hospitals or sleep clinics, this may lead to ‘first-night’ effects, evident in lower sleep-efficiency and decreased sleep time [8]. Home-based or ambulatory PSG is considered to be an alternative to in-clinic PSG to overcome some of these limitations. However, this requires patients to put on several electrodes at precise locations on their scalps and bodies, making it a difficult system to use in practice.

Recent research has focused on creating user-friendly wearable systems and development of automated sleep

The associate editor coordinating the review of this manuscript and approving it for publication was Ioannis Schizas¹.

staging methods to facilitate long-term sleep monitoring in these systems. These wearable systems allow capturing data over multiple nights in order to obtain more information about the patient's sleep problems [2], [9]–[12]. These systems use EEG signals recorded from a limited number of EEG channels to obtain the different stages of sleep automatically. Analysis and classification of signals obtained over long periods require methods that can extract different features from them and identify the sleep stages automatically. With the availability of better computing resources over the last decade, several methods for automatic sleep scoring have been published using various features of EEG signals together with different classification methods [13]–[18]. This ability to perform automatic sleep staging in comfortable environments (e.g. homes) improves scoring reliability, reduces time and cost of sleep staging, subsequently making sleep disorder diagnosis accessible to a larger population [19].

Wearable systems are generally very small with limited processing and computing resources and a small power source. Hence it is imperative to reduce the size of data stored or wirelessly transmitted from these systems to limit power consumption. For long-term sleep monitoring over several weeks, large amount of EEG data is generated even from a single channel [20]. With a growing trend of wearable devices to support higher sampling frequencies and recording channels to improve diagnostic accuracy, overcoming the power and storage constraints becomes even more important [9]. Therefore, reduction in the quantity of data to be transmitted or stored, through compression, is desirable in wearable devices for long-term monitoring.

Compression encodes a sequence of data with smaller number of bits compared to its original bit length. Lossy EEG compression achieves higher compression ratio (CR) than lossless techniques by allowing an amount of loss of signal fidelity to be tolerated. This loss is commonly measured by percentage root-mean squared difference (PRD). A trade-off exists between CR and the loss in signal fidelity such that higher CRs result in larger PRDs in reconstructed signals. In the context of sleep staging, the overall sleep staging accuracy as well as the detection of different stages may deteriorate when neural features characteristic of sleep stages are distorted by the compression process. Hence, it is important to investigate how the performance of automated sleep stage detection is affected by different state-of-the-art lossy EEG compression algorithms at different compression levels.

This paper aims to study the effects of different EEG compression algorithms and compression ratios on the output of automatic sleep staging algorithms for use in long-term sleep monitoring systems. It presents a detailed analysis demonstrating the impact of state-of-the-art compression methods on the quality of EEG signals and subsequently their resulting impact on automatic sleep stage classification. Section II reviews specific compression algorithms that are suitable for EEG signals. Based on this review and the specific requirements of wearable sleep staging, six algorithms are selected. Section III describes the implementation

of these algorithms, the database used to study their performances, the sleep staging algorithm to characterise the effect of compression, and the evaluation metrics used to present the results. In Section IV, the results of compression are presented demonstrating the impact of different compression parameters on reconstructed signal quality and the reconstructed signal error at different compression ratios. Section IV also includes a runtime comparison between the compression algorithms and briefly describes how they perform on noisy EEG data. Finally, Section V presents the impact of different compression methods on the overall sleep staging accuracy as well as in the different sleep stages.

II. REVIEW OF COMPRESSION METHODS

There are hundreds of compression schemes that already exist in academic literature, many of them already quite widely used for various industrial and medical applications. In this paper, we have restricted our focus to compression methods that are relevant for the specific application of wearable EEG-based sleep staging. As a result, the methods reviewed in this section include those that are known to be appropriate for compressing EEG signals, have a high compression ratio, and can be implemented in hardware with relatively low computational complexity. In light of these requirements, this paper focuses on lossy compression methods, which provide higher compression ratio compared to that of lossless compression at the expense of non-exact reconstruction.

There are three main types of lossy EEG compression methods: (1) Transform-based (2) Filter-based and (3) Predictor-based. Transform-based compression involves transformation of the signal to a different domain, where the sparsity of the signal can be exploited by retaining only the most significant components through a thresholding stage [21]. Majority of transform-based compression methods in current literature uses discrete wavelet transform (DWT) pre-processing [22]–[27] to get a sparse representation of the signal. Higgins *et al.* [25] proposed a modified JPEG2000 compression algorithm that consisted of a DWT pre-processing stage, followed by thresholding and uniform quantization of wavelet coefficients. They used Cohen-Daubechies-Feaveueau 9/7 (CDF9/7) wavelet to compute the DWT coefficients of the EEG data, and added a hard thresholding stage in which the threshold was tuned to study the trade-off between CR and seizure detection accuracy. In comparison to using an arithmetic encoder (AC), they showed higher CRs at set seizure detection accuracies using the Set Partitioning in Hierarchical Trees (SPIHT) encoding algorithm. This compression method can be coined as the Lossy SPIHT algorithm. In a separate study [22], they also studied the processor load of the Lossy SPIHT algorithm on a 50MHz Blackfin BF537 device, and summarized SPIHT's nature as an embedded coder and its advantages of low computational complexity. Higgins *et al.* [26] proposed that scalar quantization of wavelet coefficients coupled with a lossless SPIHT stage, further improved CR-PRD balance, compared to directly encoding coefficients with standard

lossy SPIHT without prior quantization (Lossy SPIHT method). Higgins *et al.* coined this new compression algorithm the QSPIHT. In [28], Cardenas-Barrera *et al.* used the same compression architecture as in [25], but proposed wavelet packet transform (WPT) pre-processing and a run-length coder (RLC) for encoding and achieved a peak CR of 9.13 at a PRD of 5.25%. Adopting a different transform, Birvinskas *et al.* [29] applied fast fixed-point DCT transforms to EEG data and truncated last DCT coefficients of each frame for lossy compression. The examined DCT transforms included Chen DCT, Loeffler DCT and BinDCT, since the high speed and low computational complexity of these transforms are attractive candidates for low-power, embedded systems. In [23], Nguyen *et al.* adopted a similar compression architecture as in [25], but applied DWT on 2D arranged EEG and used an adaptive arithmetic coder (ACC). The proposed method with 2D DWT and AAC outperformed the SPIHT-based method described in [22] slightly at a higher CR for a given PRD.

Predictor-based compression methods often consist of two stages. A predictor, such as a neural network, in the first stage estimates the value of the current sample based on those of several past samples, and only residual errors from these estimations and header information that describes the predictor model are transmitted. Several predictive models have been examined for EEG compression, including autoregressive model (AR) [30], artificial neural networks [31], [32] and recursive-least-squares predictor [26]. Sriraam *et al.* presented near-lossless predictor-based compression methods using: Single layer perceptron (SLP), Multi-layer perceptron (MLP), Elman network, Autoregressive model and Finite impulse response (FIR) filter [31]. In each method, residual errors from a predictor were uniformly quantized, and passed into an AC encoder. Among the examined methods, one with SLP model achieved the best CR-PRD balance. In other studies, the authors also examined the concept of context-based bias cancellation error modelling (CBNLC) to improve compression gains by removing the systematic bias of residual errors after quantization [32], [33]. Among the five predictor-based methods in [31], all achieved slightly higher CR at given PRDs with bias cancellation.

Filter-based compression methods involve exploiting the sparsity of signals in subbands by compressing localised spectral content. Bazan-Prieto *et al.* [34] presented a method in which EEG input was first decomposed into clinically meaningful subbands with Nearly-Perfect Reconstruction Cosine Modulated Filter Banks (N-PR CMFB). Subband coefficients were then thresholded, uniformly quantized and passed through RLC. This algorithm was further refined in [35] with a retained energy based coding. Given pre-defined PRDs, global thresholds were computed to truncate subband samples until the retained energy in subbands corresponded to the pre-defined PRD levels.

In order to compare their performance and impact, different compression methods from each of the three types were selected in this paper. From existing literature on

the transform-based methods, it can be summarized that QSPIHT, Lossy SPIHT and modified JPEG2000 EEG compression algorithms with WPT and DWT pre-processing achieve a desirable CR (a minimum of 5) at low levels of signal distortion of around 7% – 10% PRD. Importantly, SPIHT is a computationally simple embedded coder with low processor load, and hence is suitable for implementation on low-powered wearable devices [22]. Predictor-based techniques are generally more useful in the context of near-lossless and lossless EEG compression. To study their effects on compression and sleep staging accuracy, the algorithm, described in [31], [32], using a single layer perceptron (SLP) predictor followed by a uniform quantizer and an arithmetic encoder was implemented. Finally, this paper followed [34], [35] and implemented filter-based compression with M-channel Nearly-Perfect Reconstruction Cosine-Modulated Filter Banks (N-PR CMFB) to analyze its performance in the context of automatic sleep staging for long-term sleep monitoring.

III. MATERIAL AND METHODS

This section describes the EEG database used, the different EEG compression methods that are being analyzed, the sleep staging algorithm used as the reference to study the impact of compression, and the metrics used to evaluate and compare the performance of each compression algorithm.

A. DATABASE

In this paper, overnight EEG recordings from the DREAMS Subjects Database [36] were used to study the effect of compression on the reconstructed signal and its subsequent impact on automatic sleep staging accuracy. The database contains 20 whole-night polysomnography (PSG) recordings from healthy subjects. These recordings were visually scored by a sleep expert at 30-s epochs using the AASM criteria into one of the five sleep stages (Wake, N1, N2, N3, REM) [6]. All of the analysis was performed using data from a single EEG channel (FP1-A2), sampled at 200Hz. The use of this frontal EEG channel for wearable automatic sleep staging has been previously discussed in detail [12], [19].

B. WAVELET-BASED COMPRESSION

Wavelet transform allows multi-resolution analysis by providing both time and frequency localization of a signal. The ability to compute time-frequency representation is useful for analysis of non-stationary biomedical signals such as the EEG. The wavelet-based EEG compression algorithm in this paper follows the architecture of the modified JPEG2000 Part 1 algorithm proposed in [25]. The core components include segmentation, DWT pre-processing, uniform quantization, thresholding and entropy encoding. Thresholding retains the most significant wavelet coefficients representative of higher energy components, and sets to zero those coefficients with magnitude below the selected threshold. Hence, the threshold level is the parameter that controls the PRD and CR of each frame. Two variants of this

algorithm were used, one involving the use of DWT (DWT-based compression) and the other using WPT (WPT-based compression) for pre-processing operations [28]. Both of these pre-processing methods are briefly explained as follows.

1) DISCRETE WAVELET TRANSFORM (DWT)

DWT decomposes a signal $x(n)$ into a set of basis functions, also known as wavelets. The wavelet family is constructed from translation and scaling of a mother wavelet ψ , where translation captures signal's variability over time and scaling extracts frequency information of the signal [37].

In the discrete domain, DWT is computed:

$$x(n) = \sum_i \sum_j c_{i,j}(\psi(2^i n - j)) \quad i, j \in Z \quad (1)$$

where $c_{i,j}$ are the wavelet coefficients, ψ the mother wavelet and i, j are integer scale and shift parameters respectively.

DWT coefficients $c_{i,j}$ are computed through a subband coding algorithm in which the signal $x(n)$ is decomposed into frequency subbands recursively using digital high-pass $g(n)$ and low-pass $h(n)$ filters. The wavelet filter coefficients are uniquely associated with a mother wavelet. At the first level of the subband coding binary tree, $x(n)$ is passed into half-band $g(n)$ and $h(n)$, where high and low frequency components are extracted respectively. Filtered signals are then downsampled by two to follow Nyquist's rule. Hence, filtered signal after each level of decomposition is double the frequency resolution and half the time resolution compared to signal from a level above. High-pass filtered and low-pass filtered samples are collected as detail coefficients (CDs) and approximation coefficients (CAs) respectively [37]. The same process is repeated recursively on the CAs in the next level until the pre-defined decomposition level is reached. Because the magnitude of DWT coefficients at each level of decomposition are representative of the signal's energy at different subbands and time intervals, the DWT coefficients provide an alternative representation of the original signal $x(n)$, providing good localization of signal's energy in both time and frequency.

2) WAVELET PACKET TRANSFORM (WPT)

WPT decomposes a signal into finer equal width frequency subbands via a full binary tree. Both CAs and CDs are passed into the filter banks at the next level of decomposition. Each wavelet packet orthonormal basis is formed by an arbitrary combination of bandpass filtering operations on CAs and CDs, and each produces a different set of disjoint subspaces that cover the signal's frequency domain. Examples of wavelet packets include Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT). Hence, the DWT basis is just one example of the wavelet packet that can be formed from wavelet packet decomposition [38]. The flexibility in multi-resolution analysis from WPT can be advantageous in compression, because an optimal wavelet packet

(best basis) for compression may be found for each frame of signal [39].

C. SPIHT-BASED COMPRESSION

Set Partitioning in Hierarchical Trees (SPIHT) were first proposed by Said and Pearlman as an efficient encoder of wavelet coefficients in image compression [40]. The SPIHT algorithm first encodes the most significant bits of the most significant wavelet coefficients, thus allowing direct control of the compression ratio. If the signal encoding or transmission process is interrupted at any point, the signal can be reconstructed to a level of fidelity appropriate to the number of bits received. Briefly, SPIHT relies on the principles of (1) set partitioning sorting algorithm and (2) ordered bit plane transmission [41]. In the sorting pass, the sorting algorithm is carried out to efficiently determine the *significance* of each wavelet coefficient distributed in the temporal orientation tree. The sorting algorithm follows a set partitioning rule and significance checking is performed through magnitude tests against a significance threshold 2^n . Significant coefficients are selected after each sorting pass. Threshold values in powers of 2 allow coefficients to be encoded as binary numbers through progressive bit plane analysis. After each sorting pass, the refinement pass is carried out where ordered bit plane transmission is performed by transmitting the n th most significant bit of significant coefficients found in the previous pass. This ensures the most significant remaining bits of significant coefficients are transmitted first. The process repeats with n decremented by one and replicates until n reaches 1. The same process and set partitioning rule run at the decoder side [40], [41].

Based on existing literature [22], [25], [26], two variants of SPIHT-based compression algorithms are implemented in this paper: Lossy SPIHT and QSPIHT. Both techniques consist of segmentation and DWT pre-processing stages, and differ in methods of introducing loss.

1) LOSSY SPIHT

In this technique, CR and PRD are controlled by setting the minimum significance threshold $2^{n_{min}}$ in the SPIHT encoding process. In lossless SPIHT, $n_{min} = 1$, and increasing n_{min} is equivalent to terminating the encoding process early. The number of quantization bits is set to the bit resolution of each original wavelet coefficient.

2) QSPIHT

In this technique, SPIHT operates in lossless mode and is used as an entropy encoder. CR and PRD are controlled by varying the quantization levels during the uniform quantization of wavelet coefficients.

D. PREDICTOR-BASED COMPRESSION

In predictor-based compression, a predictor model is used to estimate the current signal sample value based on its past values. The number of past samples used for each estimation is equivalent to the predictor order (p). The differences between

the original and the predicted samples, also called the residues and are generally of lower magnitudes than the original samples, are computed. These residues are then quantized and subsequently encoded before transmission. Compression is achieved when residues of lower magnitudes than the original samples can be encoded with shorter bit strings [30]. Apart from the encoded residues, a predictor's parameter settings and selected p number of samples are also transmitted. At the receiver, an identical predictor is used to reconstruct the signal using the transmitted parameter settings. The prediction procedure is repeated and transmitted residues are added to predicted samples at the receiver to recover the original samples with some loss from quantization.

This paper followed the algorithm described in [31], [32], and implemented a predictor-based compression technique that consists of a Single Layer Perceptron (SLP) predictor followed by a uniform quantizer and an arithmetic encoder. This is because [31], [32] concluded that compression with SLP achieves superior CR-PRD balance and better preserves EEG diagnostic information compared to compression techniques using other predictors such as Multi-Layered Perceptrons (MLP) and Autoregressive models (AR).

E. FILTER-BASED COMPRESSION

Information in EEG signals are often not well distributed across the spectrum, and may be concentrated to specific frequency subbands [42]. Hence, filter-based compression is a less generalized technique for EEG compression compared to wavelet and SPIHT-based compression methods. It can be applied to decompose EEG signal into uniformly distributed frequency bands, and subband coefficients of small amplitudes that characterize clinically insignificant subbands can be truncated for compression [34].

Based on [34], [35], this paper implemented a M-channel Nearly-Perfect Reconstruction Cosine-Modulated Filter Banks (N-PR CMFB)-based compression technique. The pipeline of this compression method is identical to that of the modified JPEG2000 Part 1 algorithm proposed in [25]. The core components of the pipeline are signal segmentation, signal decomposition, uniform quantization, thresholding and entropy encoding. The only difference is that the input signal segments are decomposed through N-PR CMFB instead of wavelet transforms.

F. SLEEP STAGING ALGORITHM

The automatic sleep staging algorithm presented in [12], [19] is being used to study the impact of compression on automatic sleep staging. This algorithm, shown in Fig. 1, works by extracting several spectral features and uses a set of contextually-drive decision trees to classify a 30-s epoch into one of five sleep stages (Wake, N1, N2, N3 and REM). It uses data from one EEG channel (Fp1-A2), and extracts different spectral features including the relative power, power ratios, and spectral edge frequencies in various frequency bands. The classification stage takes into account the last sleep stage and uses two kinds of tests (*core* and *peripheral*)

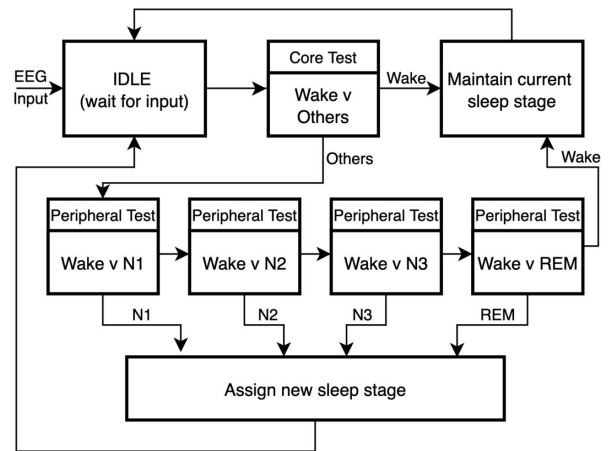


FIGURE 1. An overview of the automatic sleep staging algorithm where Core Test is a one-vs-all decision tree and Peripheral Test is one-vs-one decision tree [12].

to determine the next sleep stage. The *core* test consists of a *one-vs-all* decision tree that is used to determine whether a state transition is required. Only if it is required, a series of *one-vs-one* decision trees are used to determine the next sleep stage. As an example, if the last classified sleep stage is Wake then a *Wake-vs-Others* core test is used to establish whether the current epoch under analysis is still Wake or one of the other sleep stages. If it is deemed to be Wake, no further computations are required. If however, a state transition is needed then *Wake-vs-N1*, *Wake-vs-N2*, *Wake-vs-N3*, and *Wake-vs-REM* peripheral tests are used to find the new sleep stage. This approach of using a combination of small decision trees in a contextual manner based on the last sleep stage is highly beneficial for use in wearable systems where computing resources are severely constrained and has been validated with an ultra-low power integrated system presented in [12].

G. PERFORMANCE METRICS

In order to evaluate the compression performance, reconstruction accuracy of signal after compression, and the sleep staging accuracy after compression, several different metrics have been used.

1) COMPRESSION RATIO

Compression Ratio (CR) measures the efficiency of the compression process, and is computed as the ratio of the original data size with respect to the compression data size.

$$\begin{aligned}
 CR &= \frac{\text{Original data size}}{\text{Compressed data size}} \\
 &= \frac{L_o \times r}{b_c} \tag{2}
 \end{aligned}$$

where L_o is the length of the input EEG signal into the compression algorithm in samples, r the bit resolution of each original sample, and b_c the number of transmitted bits to represent the compressed signal. The larger the CR the larger the compression gains. To compute the CRs for different compression methods, the CR for each EEG signal frame

was determined, and the average CR for each subject and the entire database was computed.

2) PERCENTAGE ROOT-MEAN SQUARED DISTORTION

Percentage Root-mean squared Distortion (PRD) is a standard measure of average signal distortion between two signals, and is defined as follows.

$$PRD = \sqrt{\frac{\sum_{k=1}^N (x_k - \bar{x}_k)^2}{\sum_{k=1}^N (x_k)^2}} \times 100\% \quad (3)$$

where x_k and \bar{x}_k are the k th sample of the original and reconstructed signal respectively, and N is the length of the window over which PRD is calculated. As a measure of loss of signal fidelity, the lower the PRD, the higher the reconstruction accuracy. Similar to CR, PRD for each frame was computed and the average PRD was determined for the whole database. The CR-PRD tradeoff was then studied by examining the PRDs of reconstructed signals at given CRs.

3) SLEEP STAGING PERFORMANCE

The performance of the algorithm for classifying different sleep stages with the reconstructed signal is evaluated using the metrics recommended in [43].

Sleep Staging Accuracy (SSA) measures the overall fraction of epochs with true detections across all sleep stages.

$$Accuracy = \frac{\text{No. of true detections in stage X}}{\text{Total no. of epochs in stage X}} \quad (4)$$

Sensitivity (SEN) measures the fraction of epochs that are correctly classified in each specific sleep stage X . It addresses the variability of detection performances between sleep stages, and can be calculated as:

$$\begin{aligned} Sensitivity &= \frac{\text{No. of true detections in stage X}}{\text{Total no. of epochs in stage X}} \\ &= \frac{TP}{TP + FN} \end{aligned} \quad (5)$$

Selectivity (SEL) measures the fraction of correct detections of any sleep stage X with respect to the total number of automatic detections of that stage.

$$\begin{aligned} Selectivity &= \frac{\text{No. of true detections in stage X}}{\text{Total no. of detections of stage X}} \\ &= \frac{TP}{TP + FP} \end{aligned} \quad (6)$$

An epoch is labelled as true positive (TP) if it is correctly identified by the algorithm in a given stage, false positive (FP) if it is incorrectly identified as the stage of interest and False Negative (FN) if incorrectly rejected as the stage of interest.

IV. COMPRESSION RESULTS

A. COMPRESSION PARAMETERS

The performance of compression algorithms is greatly influenced by different compression parameters that need to be optimised for a given application. These parameters include segment length (N), mother wavelet function,

wavelet decomposition level (L), thresholding technique and encoder choice, etc. These must be selected carefully for the optimal CR-PRD balance on the test database, the DREAMS Subjects database [28]. In this section, the effects of each of these parameters on CR and PRD are analyzed both theoretically and empirically to inform design choices. This section also compares the runtime of studied compression algorithms and briefly describes how these algorithms perform on noisy EEG signals.

1) SEGMENT LENGTH

Segment (frame) length (N) selection presents a trade-off between CR and computational complexity [8]. Larger N increases CR at the expense of compression operations over a larger number of samples, which increases computational power to compress each frame of data. Nonetheless, higher CR means lower number of bits which, in turn, results in lower power required to store or transmit the compressed signal. In line with results reported in other papers [8], [25], [28], as well as experimenting with different values of N , it was empirically determined that $N = 1024$ results in the most optimal CR-PRD balance.

2) MOTHER WAVELET FUNCTION

Since basis functions from different wavelet families will capture the different characteristics of a signal's features, compression using different mother wavelets result in different reconstruction errors at a given CR [44]. Several different mother wavelets have been successfully used for EEG compression and de-noising. The most common ones include Cohen-Daubechies-Feauveau 9/7 (*CDF9/7*) [25], [26], Symlet 9 (*Sym9*) [45], and Daubechie 8 (*Db8*) [46]. Hence, the DWT-based compression was implemented with these three mother wavelets. To select the optimal wavelet function for compression on the test database, the compression performances with these wavelets are compared in Fig. 2. As shown in the figure, at a given decomposition level (L), the CRs are very similar at both low and high PRD values across the examined wavelet functions. Compression with *CDF9/7* and *Sym9* very slightly outperform other methods.

3) WAVELET DECOMPOSITION LEVELS

Selection of wavelet decomposition levels (L) is dependent on segment length (N), mother wavelet function and desired resolution at subbands. Since frequency resolution increases with L , L should be selected to provide sufficient frequency resolution required in the subbands of interests. A mathematical relationship can be established between the bandwidth of each frequency subband and each level l of the wavelet decomposition tree, given the F_s of the input EEG signal:

$$\begin{aligned} F_{A,l} &= [0, \frac{F_s}{2^{l+1}}] \\ F_{D,l} &= [\frac{F_s}{2^{l+1}}, \frac{F_s}{2^l}] \end{aligned} \quad (7)$$

where $F_{A,l}$ is the bandwidth covered by CA_l , and $F_{D,l}$ the bandwidth covered by CD_l at level l [47]. Given $F_s = 200\text{Hz}$

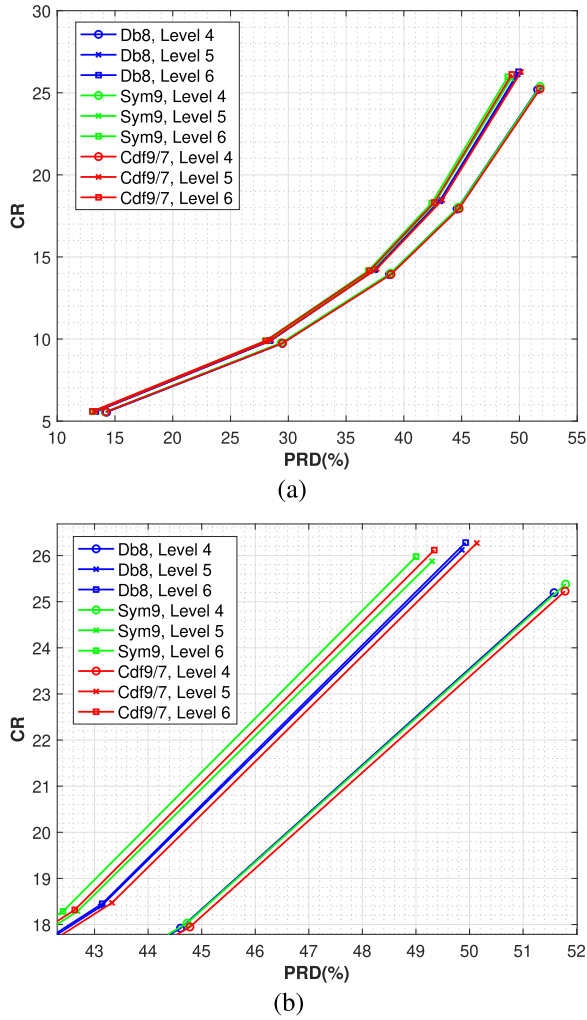


FIGURE 2. (a) CR versus PRD for DWT-based compression using different wavelet functions and number of decomposition levels; (b) Zoomed in figure at higher PRD. DWT-based compression was implemented with a hard thresholding method, 9 bit uniform quantizer and an arithmetic coder (AC).

of the DREAMS Subjects database, at $L = 6$ frequency subbands CA_6 and CD_6 have enough frequency resolution to capture low-frequency and low-bandwidth features used in [19]. Though frequency resolution increases with L , L is constrained by the boundary effects in DWT that arise from convolution on finite length signals. For each of the selected wavelet function, compression was carried out at $L = 4, 5, 6$ to study the effects of L on compression performances. The wavelet decomposition levels that result in optimal CR-PRD trade-offs can be identified in Fig. 2. The optimal CR-PRD trade-off here gives the maximum CR w.r.t a PRD. It should be noted that compression performances improve with L . At fixed PRDs above 30%, CRs at $L = 4$ are 10% lower compared to that of compression at $L = 5, 6$ for all wavelets tested.

4) THRESHOLDING TYPE

Thresholding is the main source of compression in wavelet-based compression. The energy compaction property of DWT

suggests that wavelet coefficients of high absolute magnitudes are most representative of signal features. These coefficients should be retained, whereas low amplitude coefficients are less relevant and can be discarded. The threshold T controls the CR-PRD trade-off. In this paper, T is defined as:

$$T = \min(\text{abs}(a_i)) \times P \tag{8}$$

where P is the parameter adaptively set across each iteration to control the threshold and ensure target CR is achieved.

Two common modes of thresholding are *hard* and *soft* thresholding. In *hard* thresholding, the i^{th} wavelet coefficient a_i is computed as follows:

$$a_i = \begin{cases} a_i, & \text{if } |a_i| \geq T \\ 0, & \text{if } |a_i| < T \end{cases} \tag{9}$$

Soft thresholding, on the other hand, zeros coefficients with absolute value below threshold, but additionally the remaining non-zero coefficients are shifted towards zero. The wavelet coefficients with this method are computed as follows.

$$a_i = \begin{cases} \text{sign}(a_i) \times (|a_i| - T), & \text{if } |a_i| \geq T \\ 0, & \text{if } |a_i| < T \end{cases} \tag{10}$$

Fig. 3 compares compression performances of *hard* and *soft* thresholding. It can be seen from the figure that the keep or kill wavelet shrinkage approach from *hard* thresholding generates significantly higher CRs at given PRDs.

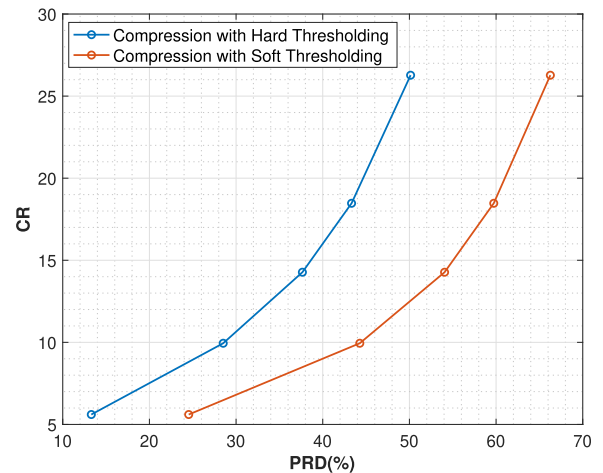


FIGURE 3. CR versus PRD for DWT-based compression using different thresholding types. Compression was implemented at 5 levels of decomposition with CDF9/7 wavelet, uniform quantizer at 9 bits, and an AC encoding.

5) ENCODER TYPE

Selection of the right type of encoder is essential to optimise the CR-PRD trade-off. Encoders should take advantage of the sparsity introduced in the wavelet domain from the preceding thresholding stages. Of the many different encoders cited in literature [23], [25], [28], [31], [34], [35], [48], three common ones have been selected to compare the performance of wavelet-based compression techniques with different encoders. These include (1) arithmetic coder (AC),

(2) huffman coder (HC), and (3) a modified run length coder (RLC).

The HC assigns variable codeword lengths so that more probable symbols are represented by shorter codewords. The AC encodes a sequence of symbols into a single floating-point number between zero and one, given the Probability Density Function (PDF) of the input sequence. Finally, the RLC replaces runs of the same symbol by the symbol followed by the length of the run. The RLC is less efficient in EEG compression compared to its use in applications involving ECG compression [23]. Compared to ECG signal in which signal energy is concentrated in low frequencies, the signal energy in EEG is distributed across different subbands. Hence, thresholding in the wavelet domain is less likely to introduce long strings of consecutive zeros, which reduces RLC's compression gains. Hence, in this paper, the conventional RLC was modified by adding an extra RLC stage to the sequence of repetition counts.

Fig. 4 shows that the best CR-PRD balance using the three encoders on the test dataset with DWT-based compression is obtained by using the AC encoder, closely followed by the RLC encoder. Further, at low CR of < 6 , compression with HC results in $\approx 30\%$ higher PRD compared to that of AC and RLC, increasing to $\approx 60\%$ at higher CR of > 10 .

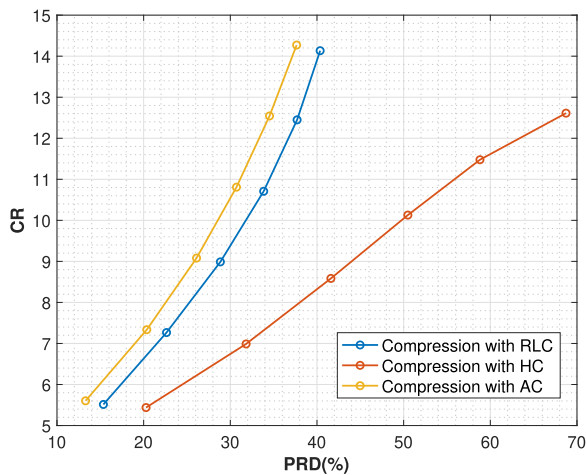


FIGURE 4. CR versus PRD for DWT-based compression on the test database using different encoder types. Compression was implemented at 5 levels of decomposition with CDF9/7 wavelet, hard thresholding, and uniform quantizer at 9 bits.

6) BEST BASIS APPROACH FOR WPT-BASED COMPRESSION

As discussed earlier, DWT and WPT are different in that WPT constructs a time-frequency representation of a signal through a complete subband tree decomposition. The rich variety of orthonormal bases offered in WPT provides flexibility in multi-resolution analysis, which may be an advantage in compression applications. Specifically, the wavelet transform basis may not provide the optimal time-frequency representation of signals for compression. To address this, an optimal wavelet packet or a best basis decomposition can be selected for each frame of EEG data w.r.t a cost function. This *best basis* representation may better capture time-frequency

characteristics of each input EEG frame, and hence improve compression performances [39].

In this paper, the best basis was found for each input frame w.r.t the Shannon entropy cost function commonly adopted in EEG signal processing [49], [50]. Fig. 5 examines the effects of the best basis approach on the CR-PRD balance of WPT-based EEG compression using data from Subject 1 of the DREAMS Subjects Database. It is shown that at set PRDs, WPT based compression with best basis search generates a slightly higher CR than a WPT based compression with a full binary tree decomposition. The theoretical advantages of WPT in compression is not substantial in empirical results. This may be because of the approximate nature of the Shannon entropy cost function to compute the best basis, which does not accurately predict the actual cost of encoding a particular decomposition tree [39].

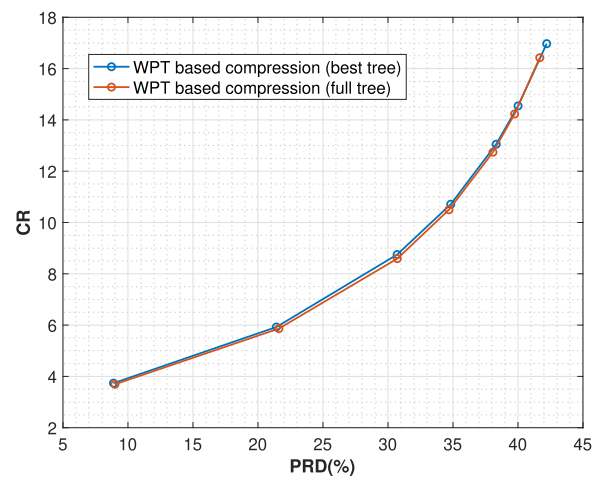


FIGURE 5. CR versus PRD for WPT-based compression with and without best basis search. Compression was implemented on the first 100 frames (102400 samples) of the test database, at 5 levels of decomposition with Db8 wavelet, hard thresholding, uniform quantizer at 9 bits and an AC encoding.

7) PREDICTOR ORDER FOR PREDICTOR-BASED COMPRESSION

For the predictor-based algorithm with a SLP model, the predictor order (p) is equivalent to the number of neurons in the input layer of the SLP. This model consisted of a linear activation function and was trained with a Levenberg-Marquart learning algorithm [31], [32]. The CR and PRD can be controlled by varying the number of quantization bits during the uniform quantization of the residues. Fig. 6 shows that the CR-PRD balances for the compression technique are very similar across the predictor orders $p = 2-5$. It also shows that compression at $p = 3, 5$ achieves the highest CRs at given PRDs.

8) CHANNEL COUNT FOR FILTER-BASED COMPRESSION

Channel count (M) of the N-PR CMFB governs the bandwidth of the subband outputs from the filter banks. Given $F_s = 200\text{Hz}$ of the DREAMS Subjects database, the CMFB with $M = 32, 16$ decompose input EEG into subbands

TABLE 1. Optimal compression parameters.

Segment length	Wavelet transform properties	Threshold type	Encoder	Predictor order	Filter bank channels
1024	<i>Db8</i> , <i>CDF9/7</i> or <i>Sym9</i> at levels 5 or 6	Hard	AC	3	8

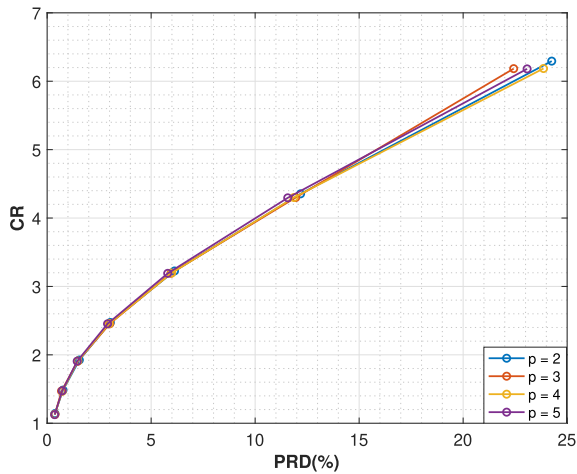


FIGURE 6. CR versus PRD for predictor-based compression using different predictor orders (p). Compression was implemented with a SLP model, uniform quantizer and AC encoding. The SLP model consisted of a linear output layer and was trained with the Levenberg-Marquart learning algorithm.

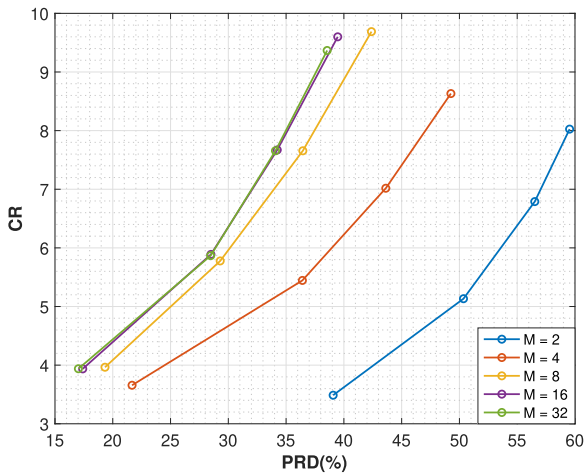


FIGURE 7. CR versus PRD for filter-based compression using a N-PR CMFB at different channel counts (M). Compression was implemented with hard thresholding, uniform quantizer at 10 bits and a RLC for encoding.

of $\approx 3\text{Hz}$ and $\approx 6\text{Hz}$ bandwidth respectively. Signal decomposition into these subbands may help retain the EEG’s characteristics after compression, because many meaningful EEG waveforms have disjoint spectral contents that lie in roughly 4Hz-wide bands [51]. As shown in Fig. 7, compression performances improve with M . Though the CR-PRD balance at $M = 32$ clearly outperforms that at $M = 2, 4$, the compression performances at $M = 8, 16, 32$ are similar. Further, an increase in M scales up the computational complexity of the compression algorithm as a result of the increase in filtering operations. The average runtime of compressing a signal segment increases by 300% when M is increased from 8 to 32,

despite little compression gains. Hence, in this paper, $M = 8$ is used as the optimal value for the filter-based compression technique.

The pipeline of the algorithm also consists of *hard* thresholding and scalar uniform quantization after signal decomposition. Instead of coding the subband coefficients with arithmetic coder (AC), this paper follows [34], [35], which employs a run length encoder (RLC) to efficiently code the large number of zero-valued coefficients that remain after *hard* thresholding.

B. OPTIMAL COMPRESSION PARAMETERS

Based on empirical results that examine the effects of different compression parameters on EEG compression methods, several optimal compression parameters were selected in an effort to maximise compression performances (Table 1). As compression performances vary with the data to be compressed, these parameters were selected strictly for the DREAMS Subjects Database. Compression with the different mother wavelet functions used do not result in clear differences in compression performances. However, wavelet decomposition levels (L) can be set to 5 or 6 to improve CR-PRD balance, though larger decomposition levels increase computational complexity. It is also clear that *hard* thresholding method outperforms the *soft* method in compression. Further, among the different encoders tested, compression with arithmetic coder (AC) shows clear advantages over other encoders. In WPT-based compression, the *best basis* approach does not clearly improve the CR-PRD balance compared to a standard WPT (full binary tree) approach. In predictor-based compression, though CR-PRD balances are similar across different predictor orders (p), a lower p is desired to reduce computational complexity. In filter-based method, channel count $M = 8$ is selected, which has shown to optimize the tradeoff between compression performance and complexity. Note that the selection of thresholding and encoder type does not apply to Lossy SPIHT and QSPIHT compression methods, which encode coefficients with the SPIHT method and introduce loss through SPIHT’s thresholding and lossy quantization respectively.

Fig. 8 compares the CR versus PRD for compressing EEG signals using wavelet, SPIHT, filter and predictor-based methods on the test dataset. All of these compression methods used optimal parameters selected from Table 1. The class of SPIHT-based methods presents a clear advantage in compression performances compared to the other methods. At set CRs from 5 to 15, SPIHT-based methods consistently achieve at least 50% lower PRD than wavelet-based algorithms, 65% lower PRD than filter-based algorithms. The CR-PRD balance for Lossy SPIHT and QSPIHT compression are similar

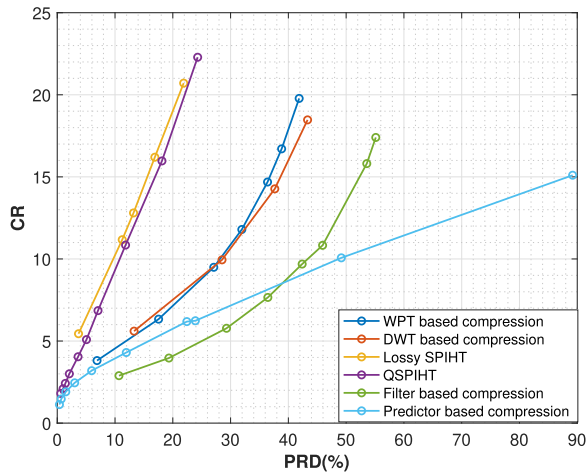


FIGURE 8. Comparison of CR versus PRD performances between wavelet-based, SPIHT-based, filter-based and predictor-based EEG compression methods.

with the Lossy SPIHT slightly outperforming the QSPIHT method. Also shown in Fig. 8, at fixed $CR \geq 10$, WPT-based compression with full binary tree generates slightly lower PRDs than DWT-based method. However, it is important to note that the time complexity of WPT-based compression is higher than that of DWT-based methods due to an increase in filtering operations from computing a complete subband tree decomposition. Predictor and filter-based techniques show the worst CR-PRD balances, with the predictor-based method outperforming the other at $CR < 8$ and underperforming at $CR \geq 10$.

TABLE 2. Runtime comparison of compression algorithms.

-	Runtime (s)			
	4	6	8	10
Compression Ratio (CR)				
DWT-based compression	1.51	2.23	2.15	2.31
WPT-based compression	2.33	3.67	3.85	3.68
Lossy SPIHT	0.35	0.29	0.32	0.32
QSPIHT	0.14	0.12	0.11	0.08
Predictor-based compression	0.15	0.14	0.14	0.13
Filter-based compression	0.45	0.52	0.58	0.61

C. RUNTIME COMPARISON

This section provides a runtime comparison between the EEG compression algorithms studied in this paper. The algorithms were implemented in MATLAB R2018b and run on a 2.3GHz Intel Core i7 processor. Table 2 compares the average runtime for compression and de-compression of one data segment (1024 samples) from the database. It can be seen that the predictor-based compression and QSPIHT algorithms require the least runtime, followed by Lossy SPIHT and filter-based algorithm. WPT and DWT-based algorithms are the slowest amongst examined compression techniques.

The compression algorithms with thresholding stages (DWT, WPT and filter-based algorithms and Lossy SPIHT) require longer runtimes, as these algorithms generally need several iterations of threshold adjustments to reach the

pre-defined CRs. For example, if an algorithm completes compression at a CR higher than the pre-defined CR, it will automatically repeat the compression again at a lower threshold and so on until the desired CR is reached, thereby increasing the algorithm runtime.

Further, the runtime of WPT-based compression is higher than that of DWT-based methods, due to the increase in filtering operations from computing a complete subband tree decomposition. The average time complexity of DWT is $O(N)$ compared to $O(N \log(N))$ for WPT [39]. In [52], Blanco-Velasco *et al.* suggested that signal decorrelation and reconstruction via CMFB carry lower computational costs than WPT, even after the WPT is pruned from best basis search. CMFB can be efficiently implemented through polyphase structures that further improve computational efficiencies [35]. The lower computational complexity of RLC compared to AC encoding also contributes to the lower runtime of filter-based compression technique compared to wavelet-based techniques.

D. COMPRESSION OF NOISY EEG SIGNALS

EEG signal artefacts often stem from physiological and non-physiological-related sources [53], [54]. Physiological activities such as blinking, eye movements, perspiration and respiration-related movements can produce low frequency EEG artefacts on EEG signals. Muscle activities (e.g. jaw clenching) and non-physiological activities such as abrupt body movements, cable movements and AC electrical interference can lead to high frequency artefacts. These artefacts are undesirable and often require pre-processing stages for their removal. If raw EEG is compressed without being pre-processed, these artefacts will be compressed as well, having an impact on the compression algorithm performance.

Wavelet-based EEG de-noising methods are popularly used to preserve the signal characteristics whilst discarding noise or artefacts [55]–[57]. The pipeline of this technique is similar to that of wavelet-based EEG compression: input signal is decomposed through wavelet transform, and low amplitude wavelet coefficients that identify noise affected frequency bands are discarded via thresholding. This means that compression algorithms with wavelet decomposition (DWT, WPT-based techniques, Lossy SPIHT and QSPIHT) may, to a certain extent, have a de-noising effect on the input EEG signal. Filter-based compression may also elicit this de-noising effect. EEG characteristics are often located at specific frequency subbands within 0–60Hz range and higher frequencies are often characterized as noise or artefacts [42]. Low magnitude subband coefficients in insignificant, high frequency subbands that contain artefacts may be thresholded during compression. In comparison to compression of EEG signals with less artefacts, PRDs at given CRs on noisy EEG signals will be higher as a result of poor reconstruction of the noisy parts of the signal.

The compression performance of predictor-based algorithms relies heavily on the predictor's accuracy. Higher prediction accuracy translates to residues of lower magnitudes,

which can be encoded with shorter bit strings. This hence increases compression gains. When compressing noisy EEG data, predictor's accuracy may drop as a result of high frequency and spontaneous artefacts. This increases the magnitudes of residues and reduces the compression performances.

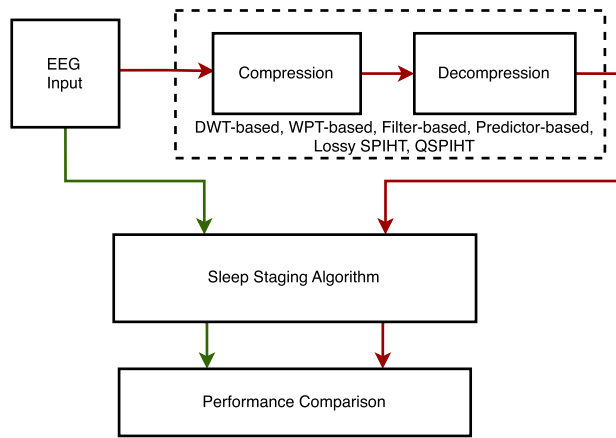


FIGURE 9. An overview of the approach used to study the impact of compression on the output of the sleep staging algorithm.

V. SLEEP STAGING RESULTS

The specific effects of lossiness in the reconstructed EEG signals, as a result of lossy compression at different compression ratios, were analyzed by performing the steps shown in Fig. 9. EEG signals were first compressed and decompressed using the six compression methods to obtain different compression ratios. The decompressed signals were passed on to the sleep staging algorithm and its performance compared to the case where no compression was performed. It should be noted that the staging algorithm performance is used strictly to assess the change due to compression and not to study the algorithm on its own. Hence, the sleep staging performance with uncompressed data with SSA = 74.75%, is used as the reference point.

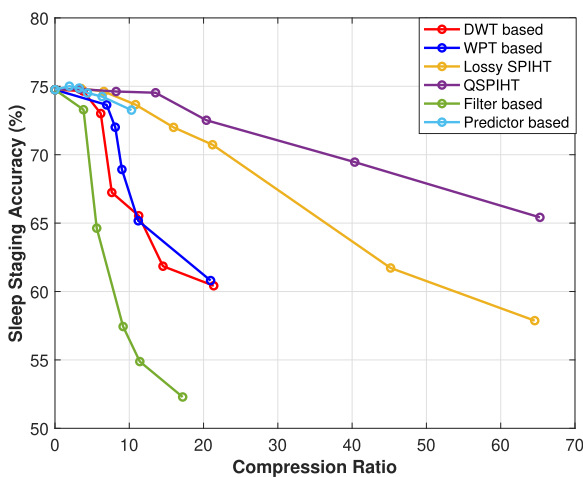


FIGURE 10. Change in sleep staging accuracy (SSA) with increasing compression ratio (CR) using different compression methods.

Fig. 10 shows the change in sleep staging accuracy (SSA) of the staging algorithm with the increase in compression

ratio (CR) using the different compression methods. It can be seen that for all the compression methods, a minimum CR ≈ 4 can be achieved, whilst limiting a decrease in SSA of less than 2% compared to that of the reference point. Applying the QSPIHT compression, a peak CR of 65.28 is achieved with ≈ 10% decrease in SSA.

Fig. 11 show the SSA against PRD for the six studied compression methods. As expected, distortions on discriminative features for sleep staging increase with PRD, which result in lower SSA. For all compression methods studied, SSA degrades by < 5% from reference point as PRD reaches 10% and by ≤ 10% as PRD reaches ≈ 20%.

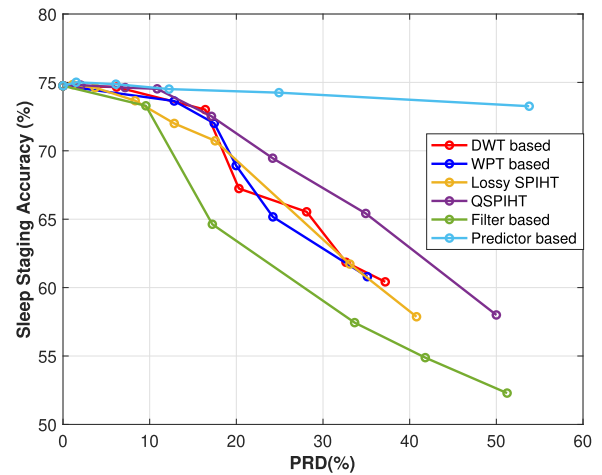


FIGURE 11. Change in sleep staging accuracy (SSA) vs average PRD using different compression methods.

It is important to compare the performances of different compression methods in the context of automatic sleep staging. As shown in Fig. 8, the class of SPIHT-based algorithms presents a clear advantage in CR versus PRD. This is reflected in Fig. 10, where it is observed that SPIHT-based algorithms outperform wavelet and filter-based methods, and higher CRs are achieved at set SSA values. This advantage becomes more apparent at lower SSA values < 70%. Given a cut-off limit in SSA at 65%, QSPIHT and Lossy SPIHT techniques achieve CR ≈ 65 and ≈ 32 respectively, compared to CR ≈ 11 for DWT, WPT-based techniques and CR ≈ 6 for filter-based technique. Among all compression techniques, the QSPIHT algorithm achieves the best trade-off in CR versus SSA. At set CR values > 40, QSPIHT compression can achieve 12% higher SSA compared to Lossy SPIHT. Surprisingly, sleep staging remains very robust against signal distortion from compression when using the predictor-based technique. Even when significant signal distortion is shown at PRD > 50%, SSA only degrades by < 2%. At similar PRDs, the predictor-based method achieves the highest SSA amongst all studied methods.

Compression gains in QSPIHT and predictor-based algorithms are controlled by the number of quantization bits rather than thresholding. This results in less precise and less direct control over CR and PRD. The predictor-based method is also constrained to achieve high CRs, reaching peak CR = 10.3

TABLE 3. Detailed sleep staging results at different CRs with DWT-based compression.

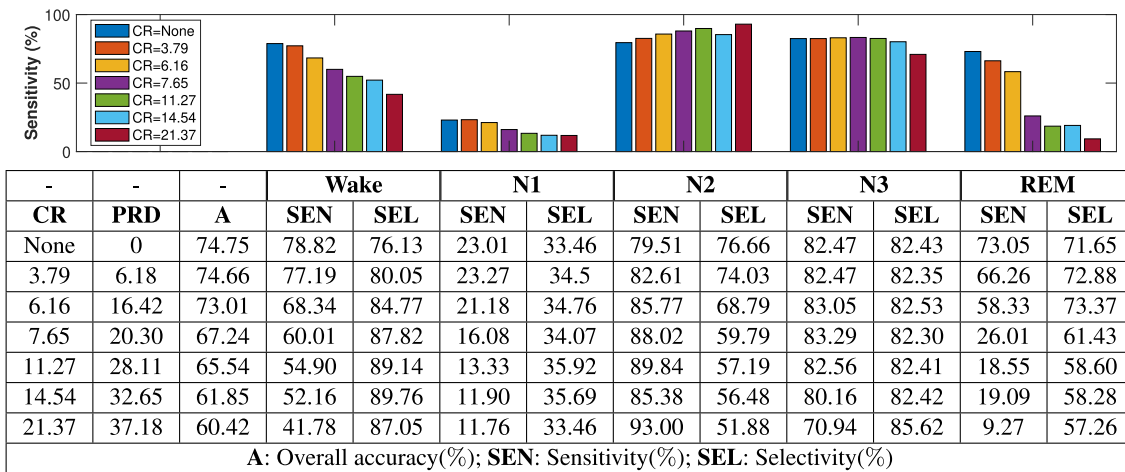
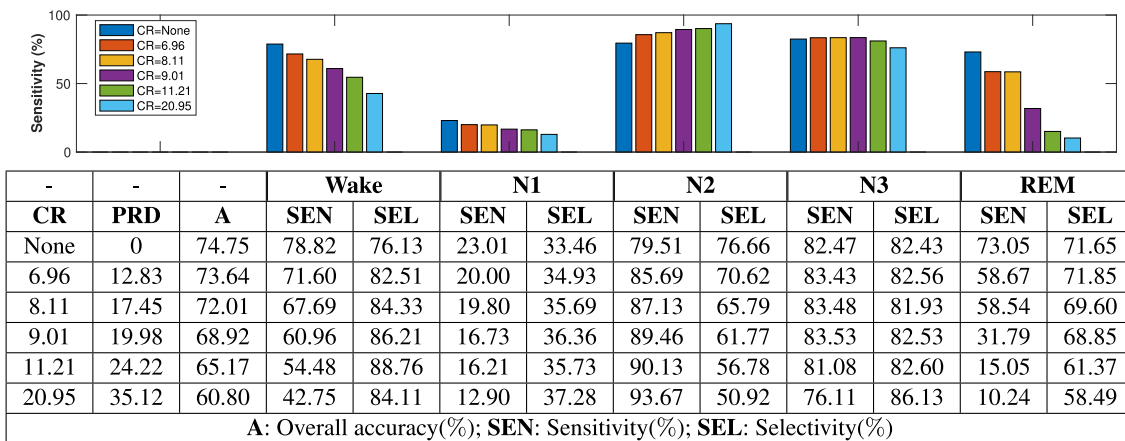


TABLE 4. Detailed sleep staging results at different CRs with WPT-based compression.



at 2 bit quantization. It is hence difficult to study sleep staging performances with this compression technique at higher CRs.

While the overall accuracy of the staging algorithm decreases at different rates with increasing compression across all compression methods, it is useful to study which sleep stages are significantly impacted due to the loss in signal introduced by compression. Table 3 shows the sensitivity (SEN) and selectivity (SEL) across different sleep stages using DWT-based compression. It can be seen that SEN in N1 and REM stages degrade by a larger percentage compared to that in Wake and N3. Especially in the REM stage, it reduces significantly by 63.78% as CR reaches 21.37. Interestingly, SEN rises with CR in the N2 stage, however the reduction in SEL suggests several false positives being detected. Further, both the SEN and SEL remain relatively stable in N3 and Wake stages with increasing compression. Since WPT and DWT-based compression essentially follow the same compression architecture, similar trends in SEN and SEL are observed across the different sleep stages as shown in Table 4. The degradation in performance across sleep stages

is also similar in the Lossy SPIHT compression method, as shown in Table 5, where SEN degrades in Wake, N1, N3 and REM with CR, and increases in the N2 stage. Finally, the results for QSPIHT compression is shown in Table 6, where it can be observed that the reduction in SEN and SEL is not as sharp as those observed in the other compression methods across all sleep stages.

Interestingly, filter-based compression has a more negative effect on the REM stage’s classification in comparison to other algorithms. Shown in Table 7, both SEN and SEL of the REM stage drop to 0% at $CR \geq 5.62$ and $PRD \geq 17.23\%$. This may be because many REM epochs are wrongfully identified as the N2 stage, hence explaining the sharp drop in SEL in N2 as a result of the rise in false positives of N2. In predictor-based compression, the SEN and SEL across CRs at different sleep stages in Table 8 also share some similar trends observed in Table 3. Unlike staging results of other algorithms, the Wake, N2 and N3 stages are not hugely affected across increasing PRDs and since they constitute for $\approx 70 - 80\%$ of the total data, the SSA remains stable too.

TABLE 5. Detailed sleep staging results at different CRs with Lossy SPIHT compression.

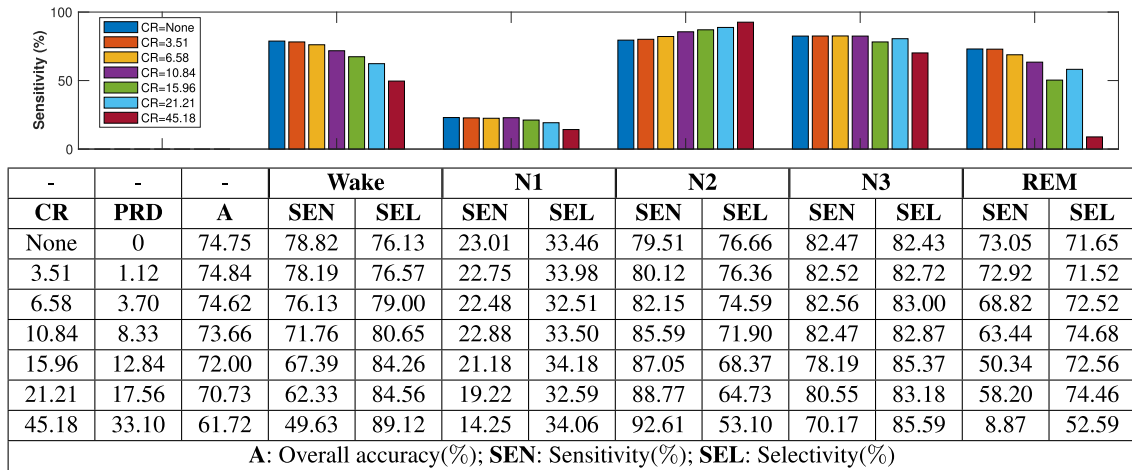


TABLE 6. Detailed sleep staging results at different CRs with QSPIHT compression.

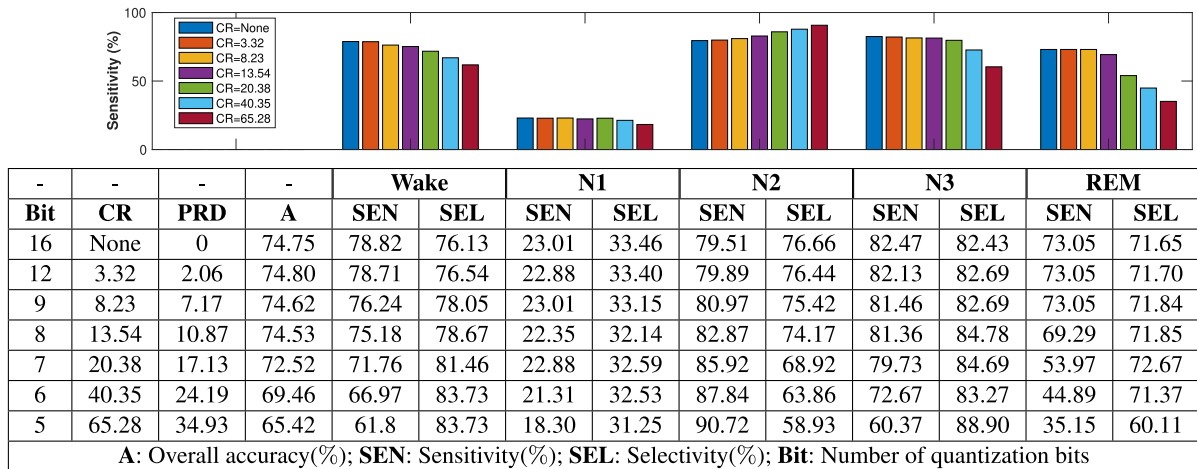
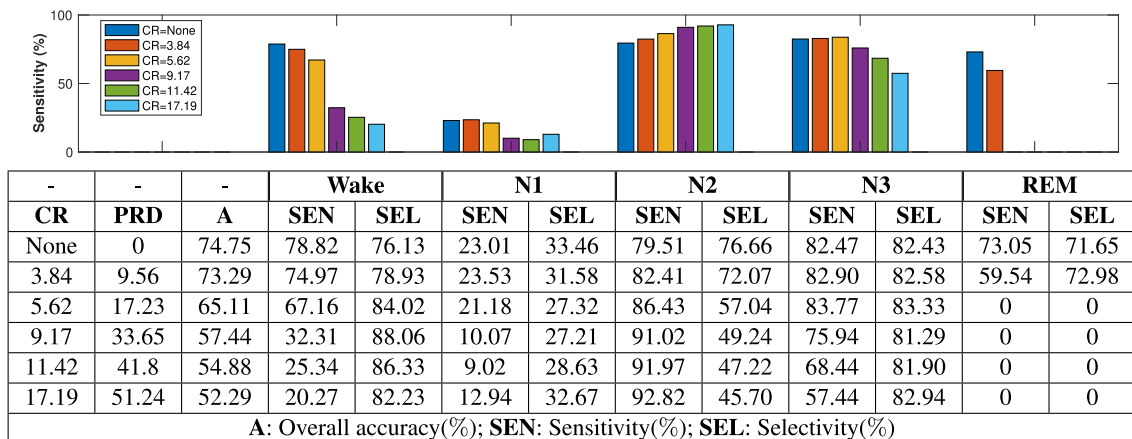
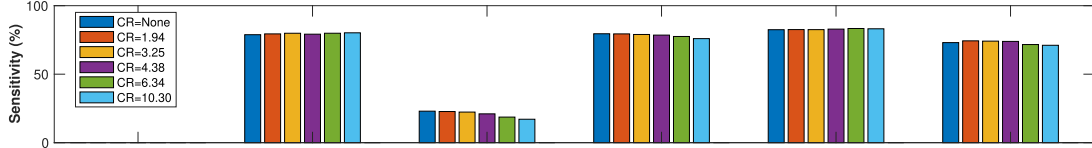


TABLE 7. Detailed sleep staging results at different CRs with Filter-based compression.



Across all studied compression methods apart from predictor-based technique, it can be summarized that at PRDs from 0% to 35%, the sharpest percentage decrease in SEN

with CR is observed in the REM stage, followed by Wake and N1 stages. In the predictor-based technique, the sharpest decrease in SEN is observed in N1, followed by N2 and REM.

TABLE 8. Detailed sleep staging results at different CRs with Predictor-based compression.


-	-	-	-	Wake		N1		N2		N3		REM	
Bit	CR	PRD	A	SEN	SEL	SEN	SEL	SEN	SEL	SEN	SEL	SEN	SEL
16	None	0	74.75	78.82	76.13	23.01	33.46	79.51	76.66	82.47	82.43	73.05	71.65
7	1.94	1.52	75.01	79.41	76.20	22.75	33.66	79.41	77.13	82.56	82.37	74.33	71.87
5	3.25	6.10	74.87	79.88	76.14	22.35	33.27	78.99	77.23	82.52	82.24	74.13	71.02
4	4.38	12.24	74.51	79.20	75.76	21.05	31.69	78.53	77.36	82.90	81.45	73.93	70.24
3	6.34	24.92	74.25	79.89	75.84	18.69	31.99	77.53	77.29	83.35	79.94	71.64	68.55
2	10.30	53.79	73.26	80.20	75.42	17.12	32.62	75.96	75.92	83.14	77.63	71.10	67.91

A: Overall accuracy(%); SEN: Sensitivity(%); SEL: Selectivity(%); Bit: Number of quantization bits

In DWT, WPT, filter-based compression methods and Lossy SPIHT, the N3 stage is comparatively more robust to compression, and experiences slowest percentage decrease in SEN. In these methods, compression gains and loss are introduced through some form of thresholding on wavelet coefficients. Loss of detail coefficients (CDs) from thresholding may lead to a smoothing effect that smoothes out high frequency components and noise. Compared to the N3 stage that mainly consists of low frequency spectral features, N1 and Wake stages are characterized by a mix of high and lower frequency components. As CR rises, the increase in distortions of high frequency features as a result of a loss of finer detail coefficients may increase the number of false negatives in N1 and Wake stages. This may explain the sharper percentage decrease in SEN in Wake and N1 stages compared to that of N3.

VI. DISCUSSION

In this paper, we examined and compared the effects of lossy EEG compression on automatic sleep stages classification. State-of-the-art lossy EEG compression techniques were first reviewed based on compression ratio, reconstruction accuracy and computational complexity. Six different EEG compression methods: DWT, WPT, filter-based, predictor-based, Lossy SPIHT and QSPIHT, were selected and implemented to compress and decompress full EEG recordings of 20 overnight sleep recordings. The decompressed recordings were passed through an automatic sleep staging algorithm and its accuracy studied for different compression methods at various compression ratios.

Among the six studied lossy EEG compression techniques, it was found that the QSPIHT algorithm, which uniformly quantizes wavelet coefficients before lossless SPIHT encoding, achieved the maximum CR at set sleep staging accuracies (SSA). Data could be compressed to $CR \approx 8$ with a degradation of 0.1% in SSA compared to the reference point and $CR > 65$ with a degradation of 10%. QSPIHT

also demonstrates highest computational efficiency, shown in lowest algorithm runtime at given CRs. Overall, SPIHT-based compression methods clearly outperformed wavelet and filter-based techniques in achieving better trade-offs between CR and SSA of automatic sleep staging. Amongst studied algorithms, the predictor-based technique sustains better SSA at higher rates of PRD. SSA only deteriorates by 0.5% at $PRD = 24.92\%$ and by $\approx 1.5\%$ at $PRD = 53.79\%$. The practical limitations of the different compression algorithms in the context of sleep staging are briefly summarized in Table 9. These limitations can be helpful in selecting the best algorithm for long-term sleep staging. For example, if Lossy SPIHT is to be used then the longer average runtime can be offset by using a faster processor or by using a faster algorithm such as the QSPIHT or predictor-based method.

To the best of the authors' knowledge, this study is the first work to examine and compare the effects of different single channel EEG compression methods in the context of automatic sleep staging. This is important particularly for long-term sleep monitoring systems that are required to capture EEG recordings across several hours for multiple nights. Further, while the results in this paper are presented using a specific sleep staging algorithm, the insight gained could be applicable to other sleep algorithms as well as other EEG-based systems. For example, brain monitoring systems for epilepsy monitoring can be used to collect data for sleep analysis as well if the right compression methods are being applied. However it should be noted that only single channel EEG compression methods are analyzed in this paper since the focus of this work is on their application in single channel EEG-based sleep monitoring. In multi-channel systems, the correlation between the different channels needs to be exploited in order to achieve better compression performance [58], [59]. Though requiring higher computational resources, multi-channel systems can result in higher sleep staging accuracy for given compression ratios, since more information will be available during sleep staging.

TABLE 9. Compression algorithms and their limitations.

Algorithms	Limitations
DWT-based compression	<ul style="list-style-type: none"> • Lower SSA at set CRs compared to SPIHT-based methods. • Longer average algorithm runtime at given CRs than SPIHT, filter and predictor-based methods when tested on same datasets.
WPT-based compression	<ul style="list-style-type: none"> • Lower SSA at set CRs compared to SPIHT-based methods. • Longest average algorithm runtime at given CRs amongst studied algorithms, when tested on same datasets. • Higher computational complexity than DWT-based approach.
QSPIHT	<ul style="list-style-type: none"> • Less direct (precise) CR and PRD control compared to techniques with thresholding stages.
Lossy SPIHT	<ul style="list-style-type: none"> • Lower SSA at set CRs compared to QSPIHT. • Longer average algorithm runtime at given CRs than QSPIHT and predictor-based compression when tested on same datasets.
Predictor-based compression	<ul style="list-style-type: none"> • Less direct (precise) CR and PRD control compared to techniques with thresholding stages. • Constraints to achieve high CRs (e.g. CR > 10.30 on the test database).
Filter-based compression	<ul style="list-style-type: none"> • Lowest SSA at set CRs amongst studied algorithms. • More negative effect on REM stage's classification compared to other algorithms.

Overall, this paper demonstrates the huge potential of EEG compression in expanding the end-user acceptance of wearable EEG systems in long-term sleep monitoring. If compression techniques such as QSPIHT can be efficiently implemented for low-resource devices, the consequent gains due to EEG compression would improve battery performance, storage capacities and reduce battery size of these wearable EEG systems, with minimal impact on sleep staging performances. The subsequent improvement in end-user acceptance of wearable EEG systems would significantly improve the diagnosis and treatment of sleep disorders that require long-term monitoring.

REFERENCES

- [1] A. Anxiety. *Depression Association of America*. Accessed: Jul. 2020. [Online]. Available: [https://adaa.org/understanding-anxiety/related-illnesses/sleep-disorder%](https://adaa.org/understanding-anxiety/related-illnesses/sleep-disorder%5)
- [2] D. Looney, V. Goverdovsky, I. Rosenzweig, M. J. Morrell, and D. P. Mandic, "Wearable in-ear encephalography sensor for monitoring Sleep. Preliminary observations from nap studies," *Ann. Amer. Thoracic Soc.*, vol. 13, no. 12, pp. 2229–2233, Dec. 2016.
- [3] A. R. Wolfson and M. A. Carskadon, "Sleep schedules and daytime functioning in adolescents," *Child Develop.*, vol. 69, no. 4, pp. 875–887, Aug. 1998.
- [4] D. C. Baldwin and S. R. Daugherty, "Sleep deprivation and fatigue in residency training: Results of a national survey of First- and second-year residents," *Sleep*, vol. 27, no. 2, pp. 217–223, Mar. 2004.
- [5] B. M. Altevogt, *Sleep Disorders Sleep Deprivation: Unmet Public Health Problem*. New York, NY, USA: Academic, 2006.
- [6] C. Iber, *The AASM Manual For Scoring Sleep Associated Events: Rules, Terminology Technical Specifications*, vol. 1. Westchester, IL, USA: American Academy of Sleep Medicine, 2007.
- [7] A. V. Shelgikar, P. F. Anderson, and M. R. Stephens, "Sleep tracking, wearable technology, and opportunities for research and clinical care," *Chest*, vol. 150, no. 3, pp. 732–743, Sep. 2016.
- [8] K.-C. Lan, D.-W. Chang, C.-E. Kuo, M.-Z. Wei, Y.-H. Li, F.-Z. Shaw, and S.-F. Liang, "Using off-the-shelf lossy compression for wireless home sleep staging," *J. Neurosci. Methods*, vol. 246, pp. 142–152, May 2015.
- [9] A. Casson, D. Yates, S. Smith, J. Duncan, and E. Rodriguez-Villegas, "Wearable electroencephalography," *IEEE Eng. Med. Biol. Mag.*, vol. 29, no. 3, pp. 44–56, May 2010.
- [10] V. Goverdovsky, D. Looney, P. Kidmose, and D. P. Mandic, "In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring," *IEEE Sensors J.*, vol. 16, no. 1, pp. 271–277, Jan. 2016.
- [11] T. Torfs, V. Leonov, R. F. Yazicioglu, P. Merken, C. Van Hoof, R. J. M. Vullers, and B. Gyselinckx, "Wearable autonomous wireless electro-encephalography system fully powered by human body heat," in *Proc. IEEE Sensors*, Oct. 2008, pp. 1269–1272.
- [12] S. A. Imtiaz, Z. Jiang, and E. Rodriguez-Villegas, "An ultralow power system on chip for automatic sleep staging," *IEEE J. Solid-State Circuits*, vol. 52, no. 3, pp. 822–833, Mar. 2017.
- [13] A. Krakovská and K. Mezeiová, "Automatic sleep scoring: A search for an optimal combination of measures," *Artif. Intell. Med.*, vol. 53, no. 1, pp. 25–33, Sep. 2011.
- [14] K. Šuamáková and A. Krakovská, "Discrimination ability of individual measures used in sleep stages classification," *Artif. Intell. Med.*, vol. 44, no. 3, pp. 261–277, Nov. 2008.
- [15] S. Taran, P. C. Sharma, and V. Bajaj, "Automatic sleep stages classification using optimize flexible analytic wavelet transform," *Knowl.-Based Syst.*, vol. 192, Mar. 2020, Art. no. 105367.
- [16] D. Jiang, Y. Ma, and Y. Wang, "Sleep stage classification using covariance features of multi-channel physiological signals on Riemannian manifolds," *Comput. Methods Programs Biomed.*, vol. 178, pp. 19–30, Sep. 2019.
- [17] M. Dijkstra, Y. Li, and S. Abdulla, "EEG sleep stages identification based on weighted undirected complex networks," *Comput. Methods Programs Biomed.*, vol. 184, Feb. 2020, Art. no. 105116.
- [18] S. Abdulla, M. Dijkstra, R. L. Laft, K. Saleh, and R. C. Deo, "Sleep EEG signal analysis based on correlation graph similarity coupled with an ensemble extreme machine learning algorithm," *Expert Syst. Appl.*, vol. 138, Dec. 2019, Art. no. 112790.
- [19] S. A. Imtiaz and E. Rodriguez-Villegas, "Automatic sleep staging using state machine-controlled decision trees," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 378–381.
- [20] A. J. Casson and E. Rodriguez-Villegas, "Data reduction techniques to facilitate wireless and long term AEEG epilepsy monitoring," in *Proc. 3rd Int. IEEE/EMBS Conf. Neural Eng.*, May 2007, pp. 298–301.
- [21] K. Srinivasan, J. Dauwels, and M. R. Reddy, "A two-dimensional approach for lossless EEG compression," *Biomed. Signal Process. Control*, vol. 6, no. 4, pp. 387–394, Oct. 2011.

- [22] G. Higgins, B. McGinley, N. Walsh, M. Glavin, and E. Jones, "Lossy compression of eeg signals using spht," *Electron. Lett.*, vol. 47, no. 18, pp. 1017–1018, 9 2011.
- [23] B. Nguyen, D. Nguyen, W. Ma, and D. Tran, "Wavelet transform and adaptive arithmetic coding techniques for EEG lossy compression," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 3153–3160.
- [24] G. Xu, J. Han, Y. Zou, and X. Zeng, "A 1.5-D multi-channel EEG compression algorithm based on NLSPIHT," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1118–1122, Aug. 2015.
- [25] G. Higgins, B. McGinley, S. Faul, R. P. McEvoy, M. Glavin, W. P. Marnane, and E. Jones, "The effects of lossy compression on diagnostically relevant seizure information in EEG signals," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 1, pp. 121–127, Jan. 2013.
- [26] H. Garry, B. McGinley, E. Jones, and M. Glavin, "An evaluation of the effects of wavelet coefficient quantisation in transform based EEG compression," *Comput. Biol. Med.*, vol. 43, no. 6, pp. 661–669, Jul. 2013.
- [27] H. Daou and F. Labeau, "Dynamic dictionary for combined EEG compression and seizure detection," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 1, pp. 247–256, Jan. 2014.
- [28] J. L. Cárdenas-Barrera, J. V. Lorenzo-Ginori, and E. Rodríguez-Valdivia, "A wavelet-packets based algorithm for EEG signal compression," *Med. Informat. Internet Med.*, vol. 29, no. 1, pp. 15–27, Mar. 2004.
- [29] D. Birvinskas, V. Jusas, I. Martišius, and R. Damasevicius, "Fast DCT algorithms for EEG data compression in embedded systems," *Comput. Sci. Inf. Syst.*, vol. 12, no. 1, pp. 49–62, 2015.
- [30] G. Antoniol and P. Tonella, "EEG data compression techniques," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 2, pp. 105–114, Dec. 1997.
- [31] N. Sriraam and C. Eswaran, "Performance evaluation of neural network and linear predictors for near-lossless compression of EEG signals," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 1, pp. 87–93, Jan. 2008.
- [32] N. Sriraam, "Context-based near-lossless compression of EEG signals using neural network predictors," *AEU—Int. J. Electron. Commun.*, vol. 63, no. 4, pp. 311–320, Apr. 2009.
- [33] N. Sriraam and C. Eswaran, "Context based error modeling for lossless compression of EEG signals using neural networks," *J. Med. Syst.*, vol. 30, no. 6, pp. 439–448, Nov. 2006.
- [34] C. Bazan-Prieto, J. Cardenas-Barrera, M. Blanco-Velasco, and F. Cruz-Roldán, "Electroencephalographic compression based on modulated filter banks and wavelet transform," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 70–7067.
- [35] C. Bazán-Prieto, M. Blanco-Velasco, J. Cárdenas-Barrera, and F. Cruz-Roldán, "Retained energy-based coding for EEG signals," *Med. Eng. Phys.*, vol. 34, no. 7, pp. 892–899, Sep. 2012.
- [36] S. Devuyt. (2004). *The Dreams Subjects Database*. [Online]. Available: <http://www.tcts.fpms.ac.be/devuyt/Databases/DatabaseSubjects/>
- [37] R. Polikar. (1996). *The Wavelet Tutorial*. [Online]. Available: <http://users.rowan.edu/polikar/WTutorial.html>
- [38] M. A. Cody. (1994). *The Wavelet Packet Transform*. [Online]. Available: <http://collaboration.cmc.ec.gc.ca/science/rpn/biblio/ddj/Website/artic %e/DDJ/1994/9404/9404e/9404e.htm>
- [39] M. L. Hilton, "Wavelet and wavelet packet compression of electrocardiograms," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 5, pp. 394–402, May 1997.
- [40] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
- [41] Z. Lu, D. Youn Kim, and W. A. Pearlman, "Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 849–856, Jul. 2000.
- [42] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, pp. 205–211, Feb. 2007.
- [43] S. A. Imtiaz and E. Rodríguez-Villegas, "Recommendations for performance assessment of automatic sleep staging algorithms," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 5044–5047.
- [44] M. Nielsen, E. N. Kamavuako, M. M. Andersen, M.-F. Lucas, and D. Farina, "Optimal wavelets for biomedical signal compression," *Med. Biol. Eng. Comput.*, vol. 44, no. 7, pp. 561–568, Jul. 2006.
- [45] M. Mamun, M. Al-Kadi, and M. Marufuzzaman, "Effectiveness of wavelet denoising on electroencephalogram signals," *J. Appl. Res. Technol.*, vol. 11, no. 1, pp. 156–160, Feb. 2013.
- [46] N. Al-Qazzaz, S. H. B. M. Ali, S. Ahmad, M. Islam, and J. Escudero, "Selection of mother wavelet functions for multi-channel eeg signal analysis during a working memory task," *Sensors*, vol. 15, no. 11, pp. 29015–29035, 2015.
- [47] D. Chen, S. Wan, J. Xiang, and F. S. Bao, "A high-performance seizure detection algorithm based on discrete wavelet transform (DWT) and EEG," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0173138.
- [48] N. Sriraam, "A high-performance lossless compression scheme for EEG signals using wavelet transform and neural network predictors," *Int. J. Telemedicine Appl.*, vol. 2012, pp. 1–8, 2012.
- [49] D. Wang, D. Miao, and C. Xie, "Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection," *Expert Syst. Appl.*, p. 14, Jun. 2011.
- [50] O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schärmann, and E. Baáar, "Wavelet entropy: A new tool for analysis of short duration brain electrical signals," *J. Neurosci. Methods*, vol. 105, no. 1, pp. 65–75, Jan. 2001.
- [51] S. Sanei and J. A. Chambers, *EEG signal Processing*. Hoboken, NJ, USA: Wiley, 2013.
- [52] M. Blanco-Velasco, F. Cruz-Roldán, F. López-Ferreras, Á. Bravo-Santos, and D. Martínez-Muñoz, "A low computational complexity algorithm for EEG signal compression," *Med. Eng. Phys.*, vol. 26, no. 7, pp. 553–568, Sep. 2004.
- [53] A. Tandle, N. Jog, P. D'cunha, and M. Chheta, "Classification of artefacts in EEG signal recordings and EOG artefact removal using EOG subtraction," *Commun. Appl. Electron.*, vol. 4, no. 1, pp. 12–19, Jan. 2016.
- [54] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from EEG signals: A review," *Sensors*, vol. 19, no. 5, p. 987, Feb. 2019.
- [55] E. Estrada, N. Nazeran, E. Gustavo, S. Farideh, and S. Kamaledin, "Wavelet-based EEG denoising for automatic sleep stage classification," in *Proc. Int. Conf. Elect. Commun. Comput.*, 2011, pp. 295–298.
- [56] K. Asaduzzaman, F. Mohd-Yasin, and K. Hussain, "A study on discrete wavelet-based noise removal from EEG signals," in *Advances Computer Biology*. Cham, Switzerland: Springer, 2010, pp. 593–599.
- [57] V. Krishnaveni, S. Aravind, V. Hariharasudhan, and K. Ramadoss, "Automatic identification and removal of ocular artifacts from EEG using wavelet transform," *Meas. Sci. Rev.*, vol. 6, no. 4, pp. 45–57, 2006.
- [58] L. Lin, Y. Meng, J. Chen, and Z. Li, "Multichannel EEG compression based on ICA and SPIHT," *Biomed. Signal Process. Control*, vol. 20, pp. 45–51, Jul. 2015.
- [59] L. Shaw, D. Rahman, and A. Routray, "Highly efficient compression algorithms for multichannel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 5, pp. 957–968, May 2018.



DELAND HU LIU (Student Member, IEEE) received the M.Eng. degree in electrical and electronic engineering from Imperial College London, London, U.K., in 2019. He is currently pursuing the Ph.D. degree with The University of Texas at Austin, and focuses on creating brain-machine interfaces and neuroprosthetics to assist people with motor and cognitive disabilities. From 2019 to 2020, he was an Electronics Engineer at Dyson Technologies, U.K. His current research interests include the development of signal processing algorithms for use in wearable systems, biomedical signal analysis, neuroengineering, and neuroprosthetics.



SYED ANAS IMTIAZ (Senior Member, IEEE) received the B.Eng. degree from the National University of Sciences and Technology, Pakistan, in 2008, and the M.Sc. and Ph.D. degrees from the Imperial College London, London, U.K., in 2009 and 2015, respectively. From 2009 to 2010, he was a Digital Design Engineer with the Imagination Technologies, U.K. He is currently a Postdoctoral Research Fellow and focuses on creating novel wearable technologies to aid long-term monitoring and the diagnosis of different medical conditions. His current research interests include developing low complexity signal processing algorithms and their low-power mixed signal circuit design, particularly for use in sleep and respiratory medicine and epilepsy monitoring.

...