

Detection of EMG Signals by Neural Networks Using Autoregression and Wavelet Entropy for Bruxism Diagnosis

Temel Sonmezocak^{1,2,*}, Serkan Kurt²

¹*Department of Electrical and Energy, Istanbul Halic University,
34445 Istanbul, Turkey*

²*Department of Electronics and Communications Engineering, Yildiz Technical University,
34220 Istanbul, Turkey
temelsonmezocak@halic.edu.tr*

Abstract—Bruxism is known as the rhythmical clenching of the lower jaw (mandibular) by the contraction of the masticatory muscles and parafunctional grinding of the teeth. It affects patients' quality of life adversely due to tooth wear, pain, and fatigue in the jaw muscles. Recently, effective diagnosis methods that use electromyography, electrocardiography, and electroencephalography have been developed for bruxism. However, these methods are not economical since they require specialization and can be performed in clinical conditions. Although using surface electromyography signals alone is an economical solution, it is difficult to identify fatigue and parafunctional movements of the jaw muscles via electromyography signals due to peripheral effects. In this study, to achieve an accurate diagnosis of bruxism economically with only electromyography measurements, a new approach based on Autoregression and Shannon Entropies of Discrete Wavelet Transform Energy Spectra to identify jaw muscle activities and fatigue conditions is proposed. By using Artificial Neural Networks in the proposed model, bruxism activities can be detected most accurately.

Index Terms—Artificial intelligence; Autoregressive processes; Electromyography; Fatigue; Wavelet transform.

I. INTRODUCTION

Bruxism is a chronic dental problem seen in many people today. Its treatment is difficult and has a low success rate [1]. Bruxism is defined as the parafunctional grinding of the teeth and involuntary contraction of the lower jaw [2]–[4]. In bruxism patients, hypertrophy in masticatory muscles, tooth wear, implant fractures, tooth loss, and fatigue and pain in jaw muscles can be seen. These symptoms are due to the application of a force that is very high for the physiological structure by the involuntary contraction of jaw muscles [5]. Earliest studies performed with sensors for the diagnosis of bruxism aimed at detecting force. In these studies, healthy and afflicted individuals can be identified by monitoring the number of dental contacts in individuals with bruxism [6]. In bruxism events, the duration of contractions per hour during sleep is measured as 47.8 s–174.9 s. Forces that occur in the mouth during bruxing can be measured with strain gauge sensor systems placed in the lower and

upper jaws. In experiments on patients, it was found that 22.5 kg average force of duration 7.1 s can be generated 3.6 times in one hour [7]. Moreover, occlusal forces during bruxism can be detected by an artificial intelligence algorithm via strain gauge pressure sensors placed inside the mouth [8].

An examination of surface electromyography (sEMG) studies in the literature reveals that rhythmic parafunctional activity of jaw muscles can be observed in the population due to stress and anxiety [9], [10]. During these incidents, bursts of rhythmic masticatory muscle activity at a frequency of approximately 1 Hz and up to approximately 10 % of maximum voluntary clenching (MVC) can occur in EMG signals of masticatory muscles. In addition, these rhythmic contractions last more than 8 seconds during sleep [11], [12]. Detection of this activity, especially during sleep, is difficult and requires expert evaluation in the laboratory environment [13].

Recently, studies based on joint analysis of electrocardiography (ECG) and electroencephalography (EEG) signals, together with sEMG, have been carried out for the clinical diagnosis of bruxism. By classifying two features, namely, the mean absolute value (MAV) of the EMG signal, that is, its average amplitude over 1 s and heart rate (HR) obtained from ECG signals using multi-layer perceptrons (MLPs), it was demonstrated that changes in HR occur simultaneously with bruxism-based contractions of the jaw muscles. When MAV is considered together with HR, bruxism diagnosis results improve so that 80 % of healthy individuals and 60 % of low-frequency patients are predicted correctly, whereas 64 % classification accuracy can be achieved for the whole group [14]. When EMG signals are considered in combination with ECG and EEG by using the decision tree (DT) algorithm, 97.21 % \pm 0.836 % accuracy can be achieved in evaluating bruxism [15]. However, these methods for clinical diagnosis of bruxism are costly and time-consuming, and they require specialization.

Currently, a grinding sound reported by the partner of the patient or tooth wear is used in the assessment of bruxism. Additionally, the diagnosis is made through various methods that make use of EMG, as well as clinical findings, such as

pain in the temporomandibular joint and masticatory muscles. Assessment of bruxism through clinical practice is important for the diagnosis, but the accuracy of this assessment is not considered sufficient [16], [17]. For this reason, the golden standard for bruxism diagnosis, especially during sleep, is the evaluation of EMG of the masticatory muscles together with audio-video recordings through polysomnography in a sleep laboratory or hospital setting [18]–[20]. Currently, the complexity and high cost of this method is a disadvantage. Therefore, it is crucial to develop practical, efficient, and economical solutions based only on the use of EMG signals.

In this study, experimental work has been performed for accurately detecting jaw clenching, jaw relaxation, teeth grinding through rhythmical chewing movements of the mandibular, and pain and fatigue in the lower masseter muscles for the diagnosis and treatment of bruxism disorder by using sEMG signals. For an accurate classification, autoregression (AR) coefficients yielded by the autoregression analysis of each EMG signal obtained from the subjects were used together with artificial neural networks (ANN) and 95.2 % success rate was achieved. The same signals were decomposed by using discrete wavelet transform (DWT). The components of the signal were obtained by using the Daubechies wavelet function (db4 8-level) in this decomposition and the Shannon entropies of the discrete wavelet transform energy spectra (WShEn) of these components were calculated. All these calculated parameters were reclassified by ANN, and the highest success rate of 100% was achieved so that the physical effects seen in masseter muscles during bruxism could be classified most accurately. For this reason, the most accurate classification was obtained based on EMG measurements only. EMG measurements have been used in bruxism diagnosis for years. However, through the use of autoregression and wavelet transform energy spectra in bruxism diagnosis with EMG measurements, this study brings a novel approach to the literature. It also contributes to the literature as a practical, efficient, and economical solution by accurately identifying lower jaw movements that

occur in bruxism diagnosis with only an EMG sensor.

This article is organized as follows. Section II introduces the materials and methods used towards diagnosing the bruxism conditions. Section III presents the experimental and analytical results through graphics and tables, and Section IV discusses the impact of these results. Section V concludes the article.

II. MATERIALS AND METHODS

A. EMG Signal Acquisition and Pre-Processing

The experimental setup used for acquiring data from the subjects is shown in Fig. 1. A dual-channel data logger was designed for sEMG signals collected from the skin surface. EMG signals received from the skin surface were amplified with AD624 instrumentation amplifier (Gain: 1000, CMMR: 130 dB). The amplified signals should be freed from noise. Since the meaningful frequency of the EMG signals obtained from the muscles is within the 10 Hz–500 Hz range, a second-order Butterworth band-pass filter with corner frequencies of 10 Hz and 500 Hz was designed [21]. In the next stage, a 12-bit A/D converter was used for digitization. EMG signals consist of bipolar positive and negative signals. To sample the negative amplitudes of the signal in the positive region, DC voltage was added to the signals and they were passed through a shifting circuit. Analog EMG signals are digitized and transferred to PC environment by using a microcontroller with a 32-bit Atmel SAM3X8E ARM Cortex-M3 processor. Data received from the subjects are sampled to have 4500 samples per second (fs: 4500 Hz). The data are analysed by using MATLAB software.

In this study, data from ten subjects between the ages of 14 and 42 were taken repetitively at different times. The mean age of the subjects was 27.2 ± 9.57 . All subjects gave their informed consent in accordance with the Helsinki Declaration and understood that they were free to opt out of the experiment at any time. Moreover, the participants filled a questionnaire about bruxism awareness, anxiety, sleep habits, stress, irritability, fatigue, current facial pain intensity, fatigue of the jaw muscles, and jaw pain.

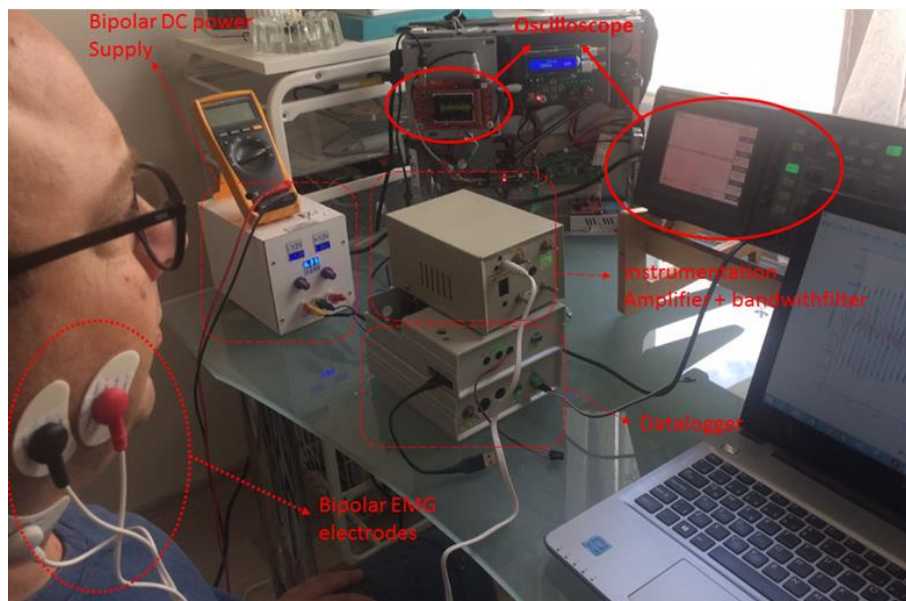


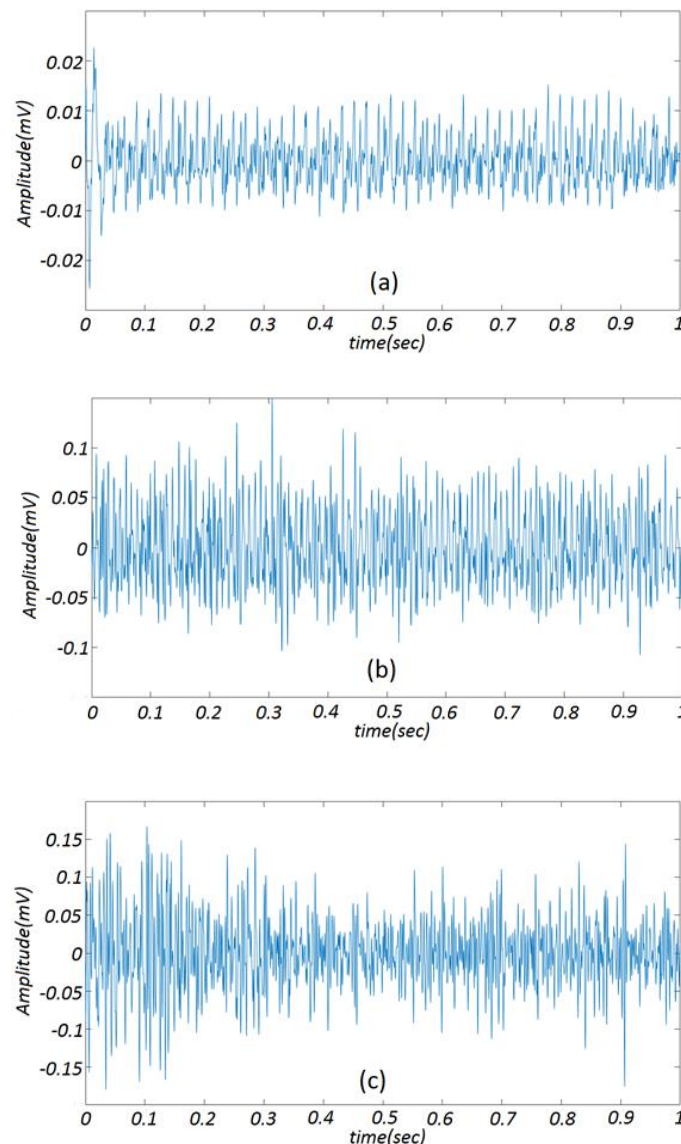
Fig. 1. Experimental setup of EMG measurement system.

For the recording of measurements, AgCl bipolar electrodes of 15 mm diameter were placed on the skin surface over the masseter muscle of the subjects with 1 cm intervals [22]. Calibration and normalization of the amplitude values of the EMG signals were performed and MVC value was determined for each subject [23]. To do this, the subjects were first made to wear an oral apparatus to protect their teeth, and then amplitude values were recorded for each subject by asking them to perform a biting (jaw squeezing) movements for 3 seconds with the maximum contraction of the lower jaw together with relaxing (jaw opening) movement. In the next stage, the bruxism states of the signals obtained from the subjects should be categorized. It is especially difficult to interpret muscle fatigue and pain states. The pain threshold for muscle fatigue can vary from person to person. Therefore, muscle fatigue/pain in case of muscle fatigue was evaluated on a numerical scale from 0 to 10 according to the research diagnostic criteria in [24]. By this means, the relationship between fatigue/pain conditions in the classification of EMG signals obtained from the subjects was determined. To do this, each individual was asked to squeeze his/her jaw muscles until pain and fatigue occurred while also monitoring the biting (jaw clenching) movement at 50 %

MVC through the amplitude values of real-time EMG signals. Even though there are variations from person to person in the onset of fatigue and pain, the average recording time was $60 \text{ s} \pm 20 \text{ s}$. This procedure was repeated at different times for each subject. Thus, EMG data for the fatigue and pain states of the subjects were obtained.

At another time, all subjects were asked to chew and grind their teeth randomly for 10 s repeatedly. At this stage, EMG data were recorded for each individual at 10 % MVC value of the jaw muscles. For analysing the recorded EMG signal data, windowing of 1 s periods was applied to each data group. Figure 2 depicts samples of real-time EMG signals obtained from the subjects in this study.

A total of 125 data records of EMG signals obtained from the subjects are used in the signal analyses to classify four different states. These states are: The jaw muscles are relaxed (0), The jaw muscles are closed (clenching) (1), Rhythmical teeth grinding (2), and Fatigue/pain (3). To identify these states, unlike other bruxism-related studies in the literature, DWT and AR models are used together to extract the most salient features from each EMG signal. From these two methods, 20 different parametric coefficients are obtained for each EMG signal of 1 s period (fs: 4500 Hz).



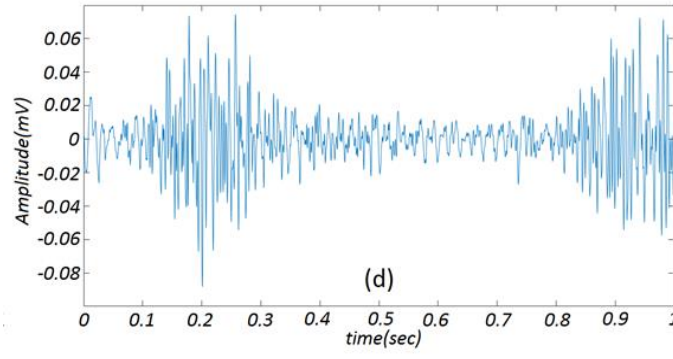


Fig. 2. Experimental results of the amplitude analysis of signals (EMG) recorded during (a) relaxation and (b) contraction of the masseter muscle as a bruxism episode. Muscle fatigue during contraction of the masseter is shown in (c). Rhythmic parafunctional activities (teeth grinding), which occur as a burst activity, are shown in (d).

Considering that there are 125 different EMG signals collected from each subject, a total of $20 \times 125 = 2500$ data are obtained. The block architecture of the proposed model that outlines the entire study is shown in Fig. 3.

In this model, the bruxism diagnosis is made by using the AR coefficients and WShEn values. The model proposed in this study will be explained below with mathematical equations.

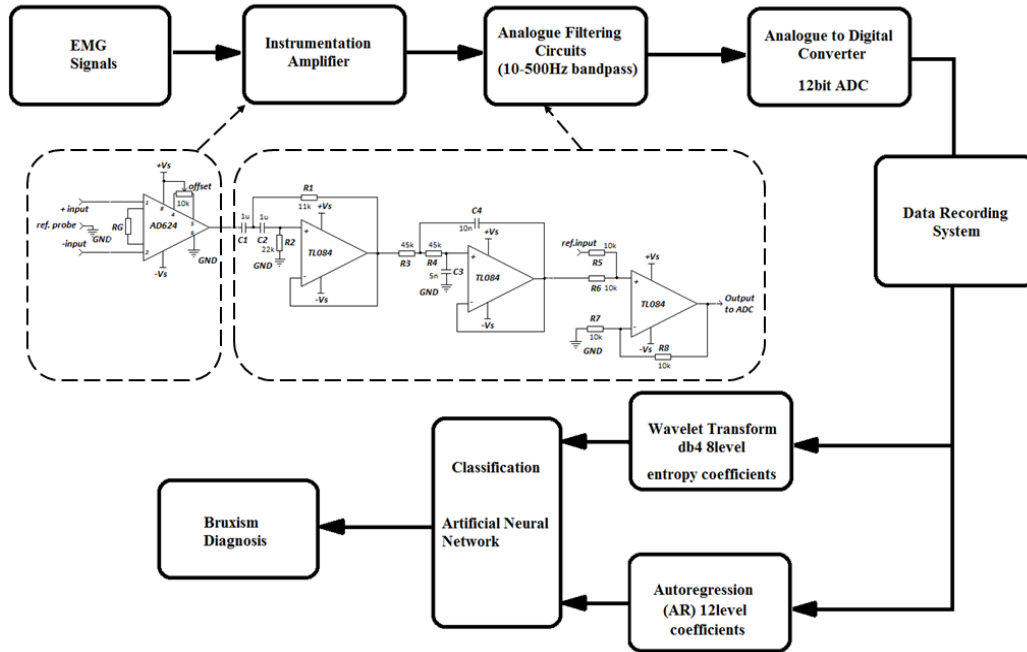


Fig. 3. Schematic diagram of the proposed method.

B. EMG Feature Extraction Using AR Modelling

AR contains important state parameters of a system and is of great importance for time series analysis since it is very sensitive to time-dependent changes in a system. This model is an important one in terms of capturing the key characteristics of any data set [25], [26]. A linear AR model can be expressed as in (1) [27]

$$X_m = \sum_{i=1}^p a_i X_{m-i} + \varepsilon_m. \quad (1)$$

In (1), X_m is the predicted value of the signal at a time m and a_i ($i = 1, 2, \dots, p$) are the weight coefficients. The parameter p represents the order of the AR model. ε_m is the error term that denotes the difference between the actual and predicted values. This value includes the white noise signal or nonstationary signals in the form of impulses [28], [29].

In this model, X_m is the time series and is expressed in terms of X_{m-i} values. Thus, autoregressive modelling of a signal results in a set of parameters. These parameters contain important information about the pattern of the EMG signal and a feature set is created during classification by means of this information. AR coefficients of order 12 are generated from the expansion of X_m the time series of each EMG signal. In the literature, AR coefficients may be created from the orders of 2, 3, 4, 6, and 8 [25]–[31]. However, in this study, AR coefficients of order 12 are created due to the complexity of the EMG signals. Employment of 12 orders in neural network classification provides high performance. In this context, we can express the AR coefficients of order 12 as a_i ($i = 1, 2, \dots, 12$) in (2) and (3):

$$X_m = a_1 X_{m-1} + a_2 X_{m-2} + \dots + a_{12} X_{m-12} + \varepsilon_m, \quad (2)$$

$$\underbrace{\begin{bmatrix} 1 & X_1 & X_2 & \cdots & X_{p-1} \\ X_1 & 1 & X_1 & \cdots & X_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ X_{p-1} & X_{p-2} & \cdots & X_1 & 1 \end{bmatrix}}_R \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_p \end{bmatrix}}_a = \underbrace{\begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_p \end{bmatrix}}_r. \quad (3)$$

In this study, $(a_1, a_2, \dots, a_{12})$ can be used in feature selection from EMG signals for bruxism diagnosis as coefficients of order 12 that represent X_m time series.

C. EMG Feature Extraction Using Discrete Wavelet Transform

Two important functions are involved in the application of DWT to an EMG signal. These are the scaling function $\phi(t)$ and the wavelet function $\psi(t)$. The scaling function is convolved with the input signal to form the approximation components. Mathematical expression of the scaling function is given in (4)

$$\phi(t) = 2^{-m/2} \phi(2^{-m}t - n). \quad (3)$$

In (4), m denotes the number of scaling levels and n denotes the amount of shifting. The mathematical expression of the approximation coefficients formed by the convolution of this function with the input signal is as shown in (5)

$$S_{m,n} = \int_{-\infty}^{+\infty} X(t) \phi_{m,n}(t) dt. \quad (4)$$

Approximation and detail components are calculated by reprocessing the approximation coefficients we obtained by the scaling function. Equation (6) expresses the approximation components and (7) expresses the detail components:

$$A(t) = \sum_{n=-\infty}^{\infty} S_{m,n} \phi_{m,n}(t), \quad (5)$$

$$D(t) = \sum_{n=-\infty}^{\infty} T_{m,n} \psi_{m,n}(t). \quad (6)$$

In (7), $T_{m,n}$ denotes the detail coefficients and $\psi_{m,n}(t)$ represents the wavelet function. Here, translation of the scaling function is important for calculating the approximation and detail components. This is shown in (8)

$$\phi(t) = \sum_k c_k \phi(2t - k). \quad (7)$$

In this equation, the function $\phi(t)$ is obtained by multiplying the function $\phi(2t - k)$, which is obtained through translation in the time axis, with the scaling coefficient c_k that corresponds to each k value. Another scaling function can be obtained from a previous scaling function through this equation. Here, there are two criteria that should be noted. To ensure orthogonality, the squared

sum of the scaling coefficients and the sum of the scaling coefficients c_k corresponding to each k value should be equal to two [32]–[35]. We can obtain the wavelet equation from (9) by using the scaling coefficients

$$\psi(t) = \sum_k (-1)^k c_{1-k} \phi(2t - k). \quad (8)$$

In (9), the mathematical expression $(-1)^k c_{1-k}$ is b_k , i.e., the wavelet coefficient. It is expressed as in (10)

$$\psi(t) = \sum_k b_k \phi(2t - k). \quad (9)$$

To decompose the signal into different bandwidths to calculate the coefficients c_k in the scaling function and b_k in the wavelet function, it should be passed through high-pass (HP) filter $h(k)$ and low-pass filter (LP) $g(k)$ as shown in Fig. 4. The functions obtained through this transformation are shown in (11) and (12) [32]–[35]:

$$\phi(t) = \sqrt{2} \sum_{k=0}^N g(k) \phi(2t - k), \quad (10)$$

$$\psi(t) = \sqrt{2} \sum_{k=0}^N h(k) \phi(2t - k). \quad (11)$$

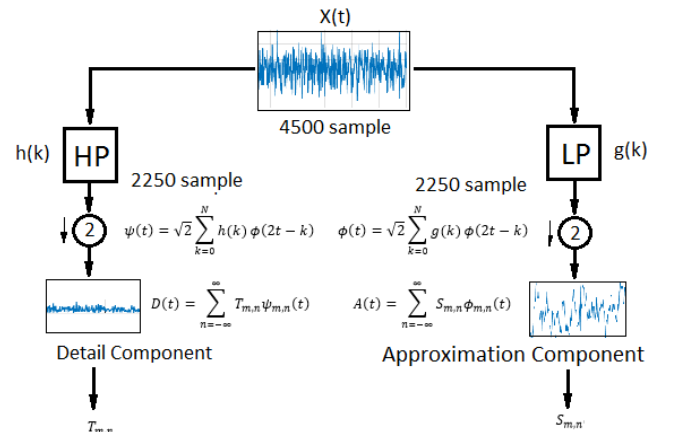


Fig. 4. One Level Wavelet Decomposition Example.

The coefficients c_k and b_k should be known to obtain these functions that are passed through low-pass and high-pass filters. Although there are different wavelet models in the literature to determine these coefficients, Daubechies wavelet makes the transformation more practical. This wavelet is symbolized as dbN . This kind of wavelet has $2N$ scaling coefficients [36]. Therefore, in this study, the Daubechies $db4$ function with a scaling of 8 ($k = 0, 1, 2, \dots, 7$) as in Fig. 5 is applied. By using this function, c_k and b_k can be obtained in the time domain as in (13) and (14) [37], [38]:

$$\phi(t) = c_7(2t - 7) + c_6(2t - 6) + \dots + c_0(2t), \quad (12)$$

$$\psi(t) = b_7(2t - 7) + b_6(2t - 6) + \dots + b_0(2t). \quad (13)$$

Approximation and detail components $A(t)$ and $D(t)$ can be obtained from c_k and b_k .

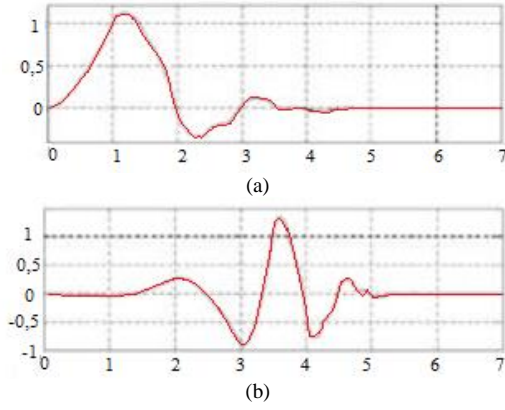


Fig. 5. Daubechies Function (db4 8-Level): (a) used as a scaling function $\phi(t)$; (b) used as a wavelet function $\psi(t)$.

In this study, Shannon entropies (E_1, E_2, \dots, E_8) of the detail components obtained through a scaling of 8 are

calculated. For this, the wavelet energy spectra (E_{jk}) of each component are considered. This is expressed in (15):

$$\left[\begin{array}{l} E_{jk} = |D_j(k)|^2, \\ \left\{ D_j(k) : \left[2^{-(j+1)} f_s, 2^{-j} f_s \right], \right. \\ \left. (j = 1, 2, \dots, 8). \right. \end{array} \right] \quad (14)$$

In (15), $D_j(k)$ expresses the detail components obtained through wavelet transformation in the frequency domain depending on the sampling frequency (f_s) at scale j and time k . Hence, WShEn of each detail component as shown in Fig. 6 is expressed in (16) as (E_j) [39]–[41]

$$E_j = -\sum_k E_{jk} \log E_{jk}. \quad (15)$$

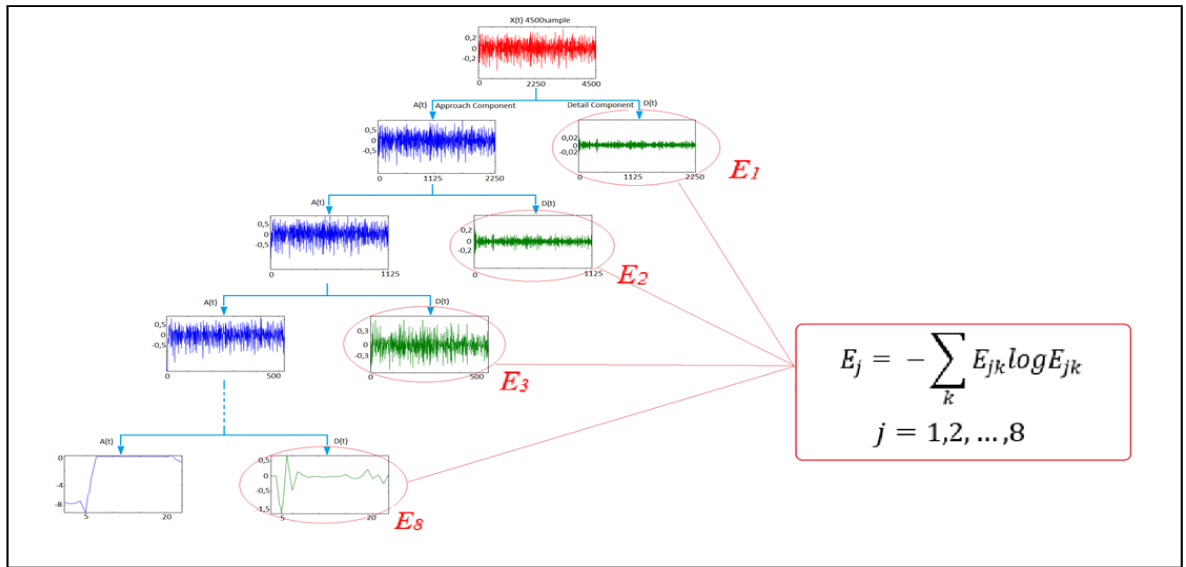


Fig. 6. Decomposition of 8-level WShEn values during contraction.

D. Diagnosis of Bruxism by Neural Network Classifier

It is essential that the system gives the correct responses in real-time measurements. In other words, it should discern the characteristics correctly even though the states that arise during jaw movement are different from the states previously encountered. For this reason, multilayer perceptron (MLP) based on feed-forward back-propagation algorithm, which is the most commonly used artificial intelligence algorithm, was used in classification. Thus, it was preferred due to its advantage of offering a higher number of variable parameters compared to other artificial intelligence algorithms. In this study, to perform a fast back propagation, network structures are trained by using Levenberg-Marquardt back propagation. The network structure is implemented by using MLP as a multilayer detector [42]–[44]. The main reason behind our preference of Levenberg-Marquardt algorithm for training the network is that this algorithm has a high learning speed due to its requiring a second degree derivative. Currently, it is the preferred method in network training [45], [46]. In these network structures, the data from the input layer produce a result at the output layer and the weights are updated to minimize the average mean square error (MSE) through

back propagation by comparing with the desired result. However, sometimes the network memorizes instead of learning. Therefore, the network should be trained properly. The data set is divided into three groups, such as training, validation, and test. MSE, bias, and weights in the training set are important for the performance of the created network. In other words, it is desired that the weights are updated correctly depending on the input data. For this reason, it should be ensured that the input data are meaningful and correlated, i.e., the regression coefficients should be high. Training operation is stopped when there is no more improvement in the validation set. The test set is for evaluating the performance of the network. Therefore, the interpretation of MSE in the test results is important for the success of the network. Here, it is expected that MSE is as close to zero as possible as a performance criterion. However, it is a major disadvantage for this network structure that there does not exist an effective method to minimize MSE. For this reason, in determining the weights of the input data, training, validation, and test ratios are found through trial and error by using different values of the neuron numbers of hidden layers with the MATLAB software. Having regression (R) close to 1 (slope

approximately 45 °) and MSE close to 0 are major criteria for evaluating the success of the network structure. If (R) is close to 0, but MSE is far from 0, there does not exist a linear relationship between the variables. Therefore, the network can be retrained to obtain an improved network structure by changing the initial weights and biases [47]. Once a successful network structure is formed, data obtained from different subjects at various times can be distinguished quickly and easily. This way, the disorder states of the data from the subjects can be detected accurately and quickly.

In this study, AR coefficients and WShEn values of the 125 different EMG signals obtained from the subjects for the cases of parafunctional teeth grinding, jaw squeezing, and the resulting pain and fatigue that occur during bruxism comprise the input to the MLP. For this reason, two different network structures with respect to the input characteristics of the EMG signals are formed. These structures are depicted in Fig. 7. The data are divided into three groups in different ratios, such as training, validation, and test with respect to the input characteristics of the created structures. Regression (R) coefficients of the network created according to these ratios are shown in Table I. The created network structures were trained by using Levenberg-Marquardt algorithm. The number of neurons for the intermediate layers was determined as 10. To investigate the performance of the method used, the classification accuracy was calculated according to (17)

$$Accuracy(\%) = \frac{1}{N} \sum_{k=1}^N \left[\frac{NCCS}{TNS} \right] \times 100. \quad (16)$$

Here, N is the number of classes, NCCS is the number of

correctly classified samples, TNS is the total number of samples, which was obtained from the confusion matrix [48] by using the MATLAB software.

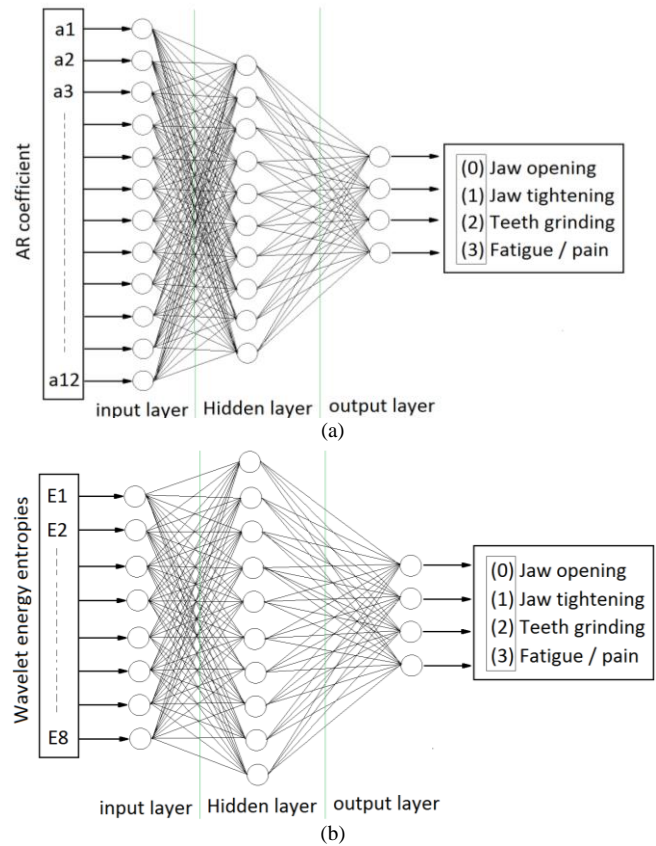


Fig. 7. Neural network topology created according to the input neuron structure, input features of the ANN with (a) AR coefficient and (b) WShEn

TABLE I. PERFORMANCE OF THE NEURAL NETWORK USING DIFFERENT PARAMETERS.

| IF | Number of Hidden Neurons | Dataset Rates (%) | R coef. | IF | Number of Hidden Neurons | Dataset Rates (%) | R coef. | IF | Number of Hidden Neurons | Dataset Rates (%) | R coef. |
|----|--------------------------|-------------------|---------|-------|--------------------------|-------------------|---------|------------|--------------------------|-------------------|---------|
| AR | 6 | 70 %* | 0.97 | WShEn | 6 | 70 %* | 0.96 | AR + WShEn | 6 | 70 %* | 0.97 |
| | | 15 %** | 0.97 | | | 15 %** | 0.93 | | | | |
| | | 15 %*** | 0.95 | | | 15 %*** | 0.92 | | | | |
| | 10 | 70 %* | 0.99 | | 10 | 70 %* | 0.98 | | 10 | 70 %* | 0.99 |
| | | 15 %** | 0.96 | | | 15 %** | 0.95 | | | | |
| | | 15 %*** | 0.95 | | | 15 %*** | 0.96 | | | | |
| | 6 | 60 %* | 0.93 | | 6 | 60 %* | 0.92 | | 6 | 60 %* | 0.97 |
| | | 30 %** | 0.91 | | | 30 %** | 0.86 | | | | |
| | | 10 %*** | 0.9 | | | 10 %*** | 0.9 | | | | |
| | 10 | 60 %* | 0.98 | | 10 | 60 %* | 0.96 | | 10 | 60 %* | 0.99 |
| | | 30 %** | 0.94 | | | 30 %** | 0.86 | | | | |
| | | 10 %*** | 0.97 | | | 10 %*** | 0.91 | | | | |
| | 6 | 50 %* | 0.95 | | 6 | 50 %* | 0.96 | | 6 | 50 %* | 0.96 |
| | | 15 %** | 0.9 | | | 15 %** | 0.9 | | | | |
| | | 35 %*** | 0.94 | | | 35 %*** | 0.81 | | | | |
| | 10 | 50 %* | 0.96 | | 10 | 50 %* | 0.97 | | 10 | 50 %* | 0.99 |
| | | 15 %** | 0.92 | | | 15 %** | 0.9 | | | | |
| | | 35 %*** | 0.95 | | | 35 %*** | 0.82 | | | | |

Note: Network parameters: (a) Dataset Rates(%): *training, **validation, ***test, (b) Training Algorithm: Levenberg-Marquardt, (c) Input Feature (IF) : AR, WShEn, (d) Output Activation Function: Sigmoid.

III. EXPERIMENTAL RESULTS

In this study, AR coefficients obtained from the autoregression analyses and WShEn values obtained from wavelet transformation of the pre-processed EMG signals received from the subjects are used by the created MLP structures for bruxism diagnosis to accurately classify the closing of the lower jaw (teeth clenching), muscle fatigue and pain in the lower jaw muscles (masseter), and especially parafunctional rhythmical jaw movements (teeth grinding). To do this, the signals shown in Fig. 1 are collected from the subjects and AR and WShEn feature parameters are obtained to classify these signals for diagnosis.

When AR and WShEn are used together for the most

accurate classification, the performance of the network is high. This can be understood from the accumulation of a significant portion of the MSE values at the zero error line in the histogram bar graphics in Fig. 8 and the best validation performance error values being lower than AR and WShEn in Fig. 9. Moreover, it can be seen from Fig. 10 that the slope is approximately 45° and the regression coefficient (R) is very close to 1 when the best correlation performance is obtained by using AR and WShEn together. In Table I, when the dataset is divided in ratios of 60 % training, 30 % validation, and 10 % test with 10 hidden layers and tested with the input feature parameters of both AR and WShEn, $R = 0.99$ and $R = 0.98$ are obtained (Fig. 10).

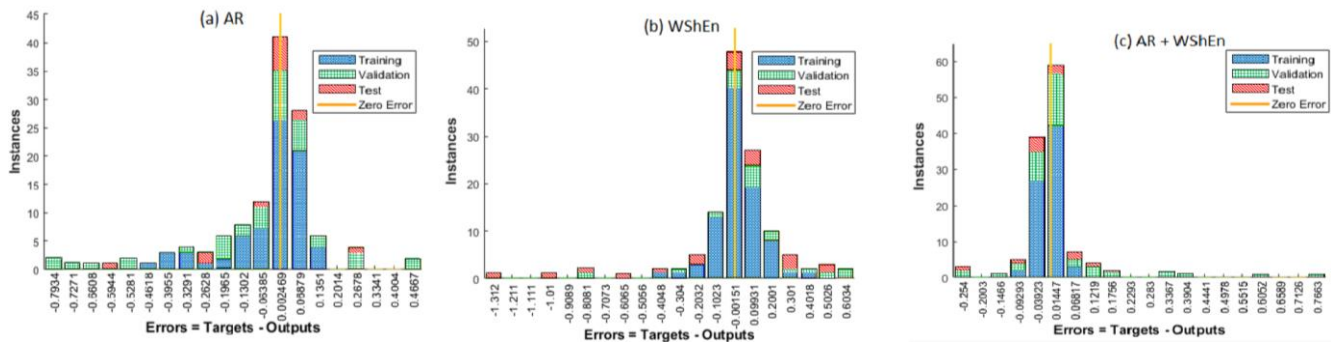


Fig. 8. MSE histogram graphics created according to the input neuron structure: a) AR; b) WShEn; c) AR + WShEn.

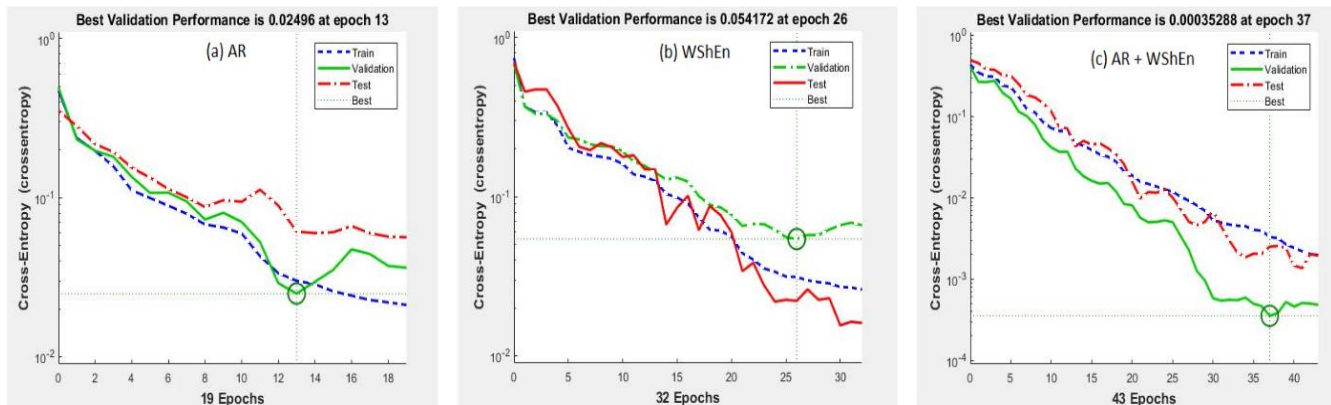


Fig. 9. Best Validation Performances of neural network structures.

As a result, it consists of work aimed at determining the performance of the network described so far. It can be clearly seen that an MLP structure with much higher performance can be created when AR and WShEn parameters are used together. This created structure can be trained quickly with high performance by the Levenberg-Marquardt algorithm.

In the next stage, EMG data, which are composed of the data obtained from the subjects and represent bruxism-related states, were classified with the MLP structure created using AR and WShEn parameters. Performance results obtained from these classifications were given in confusion matrices for 4 different output states (Table II). State 0 represents jaw opening and comprises 23 % of our dataset of 125 data obtained from the subjects. The remaining 77 % of the data are composed of data related to bruxism states. These data represent the states 1, 2, and 3 shown in Table II.

By using AR parameters, 95.2 % of the data were classified correctly, whereas 4.8 % were classified incorrectly. With WShEn parameters, classification was

made with a success rate of 94.4 %. When both parameters (AR + WShEn) were used for classification, the data were classified with 100 % accuracy so that the highest performance was achieved.

Since only EMG signals are used in this study, the number and weights of features are essential for an accurate classification in terms of the effect of the selected features. In this regard, before classification with the MLP structures, the performance of the network was determined in accordance with the features used. For an accurate classification, the decision is made by looking at R and MSE values with respect to the performance criteria of the network.

To obtain the maximum R and minimum MSE, data obtained from the subjects were simulated with MATLAB by using AR and WShEn features both separately and in combination in different ratios and the most appropriate network topology was determined as 60 % training, 30 % validation, and 10 % test. Having high training and validation ratios is an important factor for the performance

of the network. The network structure with the highest performance can be obtained by using AR and WShEn features together as shown in Table I and Figs. 8–10. For

this reason, the classification with these two feature groups has 100 % accuracy.

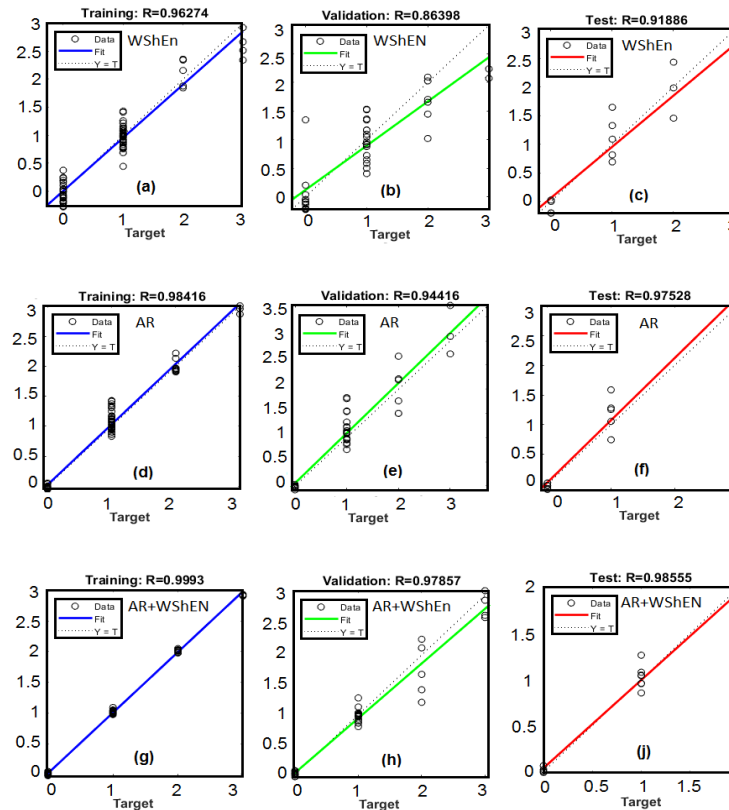


Fig. 10. Correlation performance of the network according to input characteristics (training 60 %, validation 30 %, test 10 % with hidden number 10). Correlation performance achieved by using only WShEn - (a), (b), and (c). Network performance achieved by using AR(12) order coefficients - (d), (e), and (f). Network performance achieved by using both features - (g), (h), and (j). When AR is considered together with WShEn correlations of training, validation and test data are much better than the other options in terms of network performance. The slope is approximately 45° and R is close to 1.

TABLE II. ACCURACY RATES CONFUSION MATRIX OBTAINED BY CLASSIFYING WITH AR AND WSHEN FEATURES.

| | | AR | | | | | | WShEn | | | | | | AR+WShEn | | | |
|----------------|---|----------------|---------|--------|--------|----------------|---|----------------|---------|--------|--------|----------------|---|----------------|---------|---------|--------|
| | | Target Classes | | | | | | Target Classes | | | | | | Target Classes | | | |
| | | 0 | 1 | 2 | 3 | | | 0 | 1 | 2 | 3 | | | 0 | 1 | 2 | 3 |
| Output Classes | 0 | 31.20 % | 0 % | 0 % | 0 % | Output Classes | 0 | 31.20 % | 0 % | 0 % | 0 % | Output Classes | 0 | 31.20 % | 0 % | 0 % | 0 % |
| | 1 | 0 % | 49.60 % | 2.40 % | 0 % | | 1 | 0 % | 51.20 % | 3.20 % | 0 % | | 1 | 0 % | 52.00 % | 0 % | 0 % |
| | 2 | 0 % | 2.40 % | 8.80 % | 0 % | | 2 | 0 % | 0.80 % | 8.00 % | 1.60 % | | 2 | 0 % | 0 % | 11.20 % | 0 % |
| | 3 | 0 % | 0 % | 0 % | 5.60 % | | 3 | 0 % | 0 % | 0 % | 4.00 % | | 3 | 0 % | 0 % | 0 % | 5.60 % |

0 / Jaw Open ; 1/Jaw Clenching ; 2/Teeth Grinding ; 3/Fatigue and Pain

IV. DISCUSSION

Recently, there are studies in the literature based on joint analysis of ECG and EEG signals together with sEMG for the clinical diagnosis of bruxism. If only EMG measurements are used, bruxism activities cannot be identified accurately and can be classified with 68 % accuracy. Results with higher accuracy can be obtained if ECG measurements are also used, and an 80 % success rate can be achieved in classification [14]; however, this is not sufficient.

Recently, there are studies that also take EEG measurements into consideration. In these studies, when EMG measurements are considered in combination with ECG and EEG measurements by DT algorithm, an accuracy

as high as $97.21 \% \pm 0.836 \%$ can be achieved in the clinical diagnosis of bruxism [15]. However, it is impractical, as well as costly to supplement EMG measurements with ECG and EEG.

Additionally, currently polysomnography using audio-video recordings is evaluated together with the EMG of the masticatory muscles for the diagnosis of bruxism [18]–[20].

As the application of these methods necessitates a sleep laboratory or clinical hospital setting, they are time-consuming considering the duration of hospital stay and require expert supervision. In the method proposed in this study, AR coefficients of order 12 of the EMG signals received from the subjects are obtained through autoregression and used together with 8-level WShEn values of the same signals through Wavelet transform. It is possible

to achieve 100 % accuracy by using only MLP with these selected features. When evaluated with respect to the outputs of this system, the electromyographic states of bruxism that occur in the masticatory muscles are determined by taking the measurements for the case of muscle fatigue into consideration in the classifications. Thus, in this study, all electromyographic activity that occurs during bruxism is supported by an accurate classification based on EMG measurements. By virtue of its identifying these signals with the highest accuracy, this study is important especially for developing therapeutic electronic-based stimulating biofeedback systems [49]–[51]. Furthermore, by being based only on EMG measurements, it can be implemented by using a portable EMG device that allows obtaining signals in a home setting without the need for a clinical setting. Therefore, it is practical, economical, and not time-consuming.

V. CONCLUSIONS

In this study, unlike other studies in the literature that consider EMG signals for bruxism disorder, jaw locking and random teeth grinding movements are considered together with pain and fatigue in the jaw muscles as a whole and an effective system for accurate bruxism diagnosis is proposed. By virtue of its being based on the design of a system that records only EMG signals in a home setting without the need for a laboratory setting, this study is a practical and low-cost system. Activities that occur during bruxism should be recorded real-time through EMG signals and the most effective features should be extracted by way of time and frequency analyses. This way, the signals acquired from EMG measurements can be classified with 100 % accuracy by a neural network (MLP) algorithm by using AR(12) coefficients of order 12 and WShEn obtained from the Wavelet transform of the same signals. With the approach proposed in this study, when the cost and time spent for using only EMG data are considered, it is a practical method for wearable or portable EMG devices and a suitable approach for the effective use of therapeutic wearable electronic stimulation systems in future studies.

ACKNOWLEDGMENT

The authors thank Rasit Cevizci for his medical contributions.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- [1] F. Lobbezoo, R. Jacobs, A. De Laat, G. Aarab, P. Wetselaar, and D. Manfredini, "Chewing on bruxism. Diagnosis, imaging, epidemiology and aetiology", *Ned. Tijdschr. Tandheelkd.*, vol. 124, no. 6, pp. 309–316, 2017. DOI: 10.5177/NTVT.2017.06.16194.
- [2] F. Lobbezoo, J. Van Der Zaag, and M. Naeije, "Bruxism: its multiple causes and its effects on dental implants – An updated review", *Journal of Oral Rehabil.*, vol. 33, no. 4, pp. 293–300, 2006. DOI: 10.1111/j.1365-2842.2005.01609.x.
- [3] M. J. Sateia, "International classification of sleep disorders - third edition: Highlights and modifications", *Contemporary Reviews in Sleep Medicine*, vol. 146, no. 5, pp. 1387–1394, 2014. DOI: 10.1378/chest.14-0970.
- [4] F. Lobbezoo, J. E. I. G. Brouwers, M. S. Cune, and M. Naeije, "Dental implants in patients with bruxing habits", *Journal of Oral Rehabilitation*, vol. 33, no. 2, pp. 152–159, 2006. DOI: 10.1111/j.1365-2842.2006.01542.x.
- [5] G. Bader and G. Lavigne, "Sleep bruxism; An overview of an oromandibular sleep movement disorder", *Sleep Med. Rev.*, vol. 4, no. 1, pp. 27–43, 2000. DOI: 10.1053/smr.1999.0070.
- [6] M. J. Trenouth, "The relationship between bruxism and temporomandibular joint dysfunction as shown by computer analysis of nocturnal tooth contact patterns", *J. Oral Rehabil.*, vol. 6, no. 1, pp. 81–87, 1979. DOI: 10.1111/j.1365-2842.1979.tb00408.x.
- [7] Y. Amemori, S. Yamashita, M. AĪ, H. Shinoda, M. Sato, and J. Takahashi, "Influence of nocturnal bruxism on the stomatognathic system. Part 1: A new device for measuring mandibular movements during sleep", *J. Oral Rehabilitation*, vol. 28, no. 10, pp. 943–949, 2001. DOI: 10.1046/j.1365-2842.2001.00757.x.
- [8] J. Gao, L. Liu, P. Gao, Y. Zheng, W. Hou, and J. Wang, "Intelligent occlusion stabilization stress-sensor system for bruxism diagnosis and treatment", *Sensors (Basel)*, vol. 20, no. 1, p. 89, 2020. DOI: 10.3390/s20010089.
- [9] E. Winocur, N. Uziel, T. Lisha, C. Goldsmith, and I. Eli, "Self-reported bruxism-associations with perceived stress, motivation, dental anxiety and gagging", *J. Oral Rehabil.*, vol. 38, no. 1, pp. 3–11, 2011. DOI: 10.1111/j.1365-2842.2010.02118.x.
- [10] G. De Luca Canto, V. Singh, P. Conti, B. D. Dick, D. Gozal, P. W. Major, and C. Flores-Mir, "Association between sleep bruxism and psychosocial factors in children and adolescents: A systematic review", *Clin Pediatr (Phila)*, vol. 54, no. 5, pp. 469–478, 2015. DOI: 10.1177/0009922814555976.
- [11] G. J. Lavigne, S. Khoury, S. Abe, T. Yamaguchi, and K. Raphael, "Bruxism physiology and pathology: An overview for clinicians", *Journal of Oral Rehabilitation*, vol. 35, no. 7, pp. 476–494, 2008. DOI: 10.1111/j.1365-2842.2008.01881.x.
- [12] G. J. Lavigne, N. Huynh, T. Kato, K. Okura, K. Adachi, D. Yao, and B. Sessle, "Genesis of sleep bruxism: Motor and autonomic-cardiac interactions", *Arch. Oral Biol.*, vol. 52, no. 4, pp. 381–384, 2007. DOI: 10.1016/j.archoralbio.2006.11.017.
- [13] M. Farella, S. Palla, and L. M. Gallo, "Time-frequency analysis of rhythmic masticatory muscle activity", *Muscle Nerve*, vol. 39, no. 6, pp. 828–836, 2009. DOI: 10.1002/mus.21262.
- [14] T. Castroflorio, L. Mesin, G. M. Tartaglia, C. Sforza, and D. Farina, "Use of electromyographic and electrocardiographic signals to detect sleep bruxism episodes in a natural environment", *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 6, pp. 994–1001, 2013. DOI: 10.1109/JBHI.2013.2274532.
- [15] D. Lai, M. B. B. Heyat, F. I. Khan, and Y. Zhang, "Prognosis of sleep bruxism using power spectral density approach on EEG signal of both EMG1-EMG2 and ECG1-ECG2 channels", *IEEE Access*, vol. 7, pp. 82553–82562, 2019. DOI: 10.1109/ACCESS.2019.2924181.
- [16] C. Jonsgar, P.-A. Hordvik, M. E. Berge, A.-K. Johansson, P. Svensson, and A. Johansson, "Sleep bruxism in individuals with and without attrition-type tooth wear: An exploratory matched case-control electromyographic study", *J. Dent.*, vol. 43, no. 12, pp. 1504–1510. DOI: 10.1016/j.jdent.2015.10.002.
- [17] W. Yachida, T. Arima, E. E. Castrillon *et al.*, "Diagnostic validity of self-reported measures of sleep bruxism using an ambulatory single-channel EMG device", *J. Prosthodont. Res.*, vol. 60, no. 4, pp. 250–257, 2016. DOI: 10.1016/j.jpor.2016.01.001.
- [18] M. Wieckiewicz, J. Smardz, H. Martynowicz, A. Wojakowska, G. Mazur, and E. Winocur, "Distribution of temporomandibular disorders among sleep bruxers and non-bruxer - A polysomnographic study", *J. Oral Rehabil.*, vol. 47, no. 7, pp. 820–826, 2020. DOI: 10.1111/joor.12955.
- [19] F. Lobbezoo, J. Ahlberg, A. G. Glaros *et al.*, "Bruxism defined and graded: An international consensus", *J. Oral Rehabil.*, vol. 40, no. 1, pp. 2–4, 2013. DOI: 10.1111/joor.12011.
- [20] T. Yamaguchi, S. Mikami, M. Maeda, T. Saito *et al.*, "Portable and wearable electromyographic devices for the assessment of sleep bruxism: A literature review", *The Journal of Craniomandibular & Sleep Practice*, pp. 1–9, 2020. DOI: 10.1080/08869634.2020.1815392.
- [21] H. Tankisi, D. Burke, L. Cui *et al.*, "Standards of instrumentation of EMG", *Clinical Neurophysiology*, vol. 131, no. 1, pp. 243–258, 2020. DOI: 10.1016/j.clinph.2019.07.025.
- [22] B. Afsharipour, S. Soedirdjo, and R. Merletti, "Two-dimensional surface EMG: The effects of electrode size, interelectrode distance and image truncation", *Biomedical Signal Processing and Control*, vol. 49, pp. 298–307, 2019. DOI: 10.1016/j.bspc.2018.12.001.
- [23] L. Kahl and U. G. Hofmann, "Comparison of algorithms to quantify muscle fatigue in upper limb muscles based on sEMG signals", *Medical Engineering and Physics*, vol. 38, no. 11, pp. 1260–1269, 2019. DOI: 10.1016/j.medengphy.2016.09.009.

- [24] S. F. Dworkin, "Research diagnostic criteria for temporomandibular disorders: Current status & future relevance", *J. Oral Rehabil.*, vol. 37, no. 10, pp. 734–743. DOI: 10.1111/j.1365-2842.2010.02090.x.
- [25] A. Datteo, F. Luca, and G. Busca, "Statistical pattern recognition approach for longtime monitoring of the G. Meazza stadium by means of AR models and PCA", *Eng. Struct.*, vol. 153, pp. 317–333, 2017. DOI: 10.1016/j.engstruct.2017.10.022.
- [26] P. Saha, S. Ghorai, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "Tea quality prediction by autoregressive modeling of electronic tongue signals", *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4470–4477, 2016. DOI: 10.1109/JSEN.2016.2544979.
- [27] A. Datteo, G. Busca, G. Quattromani, and A. Cigada, "On the use of AR models for SHM: A global sensitivity and uncertainty analysis framework", *Reliab. Eng. Syst. Safe.*, vol. 170, pp. 99–115, 2018. DOI: 10.1016/j.ress.2017.10.017.
- [28] M. Shannon, H. Zen, and W. Byrne, "Autoregressive models for statistical parametric speech synthesis", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 3, pp. 587–597, 2013. DOI: 10.1109/TASL.2012.2227740.
- [29] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification", *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, 2012. DOI: 10.1016/j.eswa.2012.01.102.
- [30] L. F. Sommer, C. Barreira, C. Noriega, F. Camargo-Junior, R. T. Moura, and A. Forner-Cordero, "Elbow joint angle estimation with surface electromyography using autoregressive models", in *Proc. of 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, USA, 2018, pp. 1472–1475. DOI: 10.1109/EMBC.2018.8512512.
- [31] A. M. Khan, Y. K. Lee, and T. S. Kim, "Accelerometer signal-based human activity recognition using augmented autoregressive model coefficients and Artificial Neural Nets", in *Proc. of 30th Annual International IEEE EMBS Conference*, Vancouver, British Colombia, Canada, 2008, pp. 5172–5175. DOI: 10.1109/IEMBS.2008.4650379.
- [32] M. Mansouri, A. Al-khazraji, M. Hajji, M. F. Harkat, H. Nounou, and M. Nounou, "Wavelet optimized EWMA for fault detection and application to photovoltaic systems", *Solar Energy*, vol. 167, pp. 125–136, 2018. DOI: 10.1016/j.solener.2018.03.073.
- [33] U. Libal and Z. Hasiewicz, "Risk upper bound for a NM-type multiresolution classification scheme of random signals by Daubechies wavelets", *Eng. App. of Artificial Intelligence*, vol. 62, pp. 109–123, 2017. DOI: 10.1016/j.engappai.2017.04.005.
- [34] Y. Qin, Y. Mao, and B. Tang, "Multicomponent decomposition by wavelet modulus maxima and synchronous detection", *Mech. Systems and Signal Processing*, vol. 91, pp. 57–80, 2017. DOI: 10.1016/j.ymsp.2017.01.002.
- [35] R. V. Fonseca and A. Pinheiro, "Wavelet estimation of the dimensionality of curve time series", *Annals of the Institute of Statistical Mathematics*, vol. 72, no. 5, pp. 1175–1204, 2020. DOI: 10.1007/s10463-019-00724-4.
- [36] I. Daubechies, *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992. DOI: 10.1137/1.9781611970104.
- [37] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation. Pattern analysis and machine intelligence", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989. DOI: 10.1109/34.192463.
- [38] C. Busch, "Wavelet based texture segmentation of multi modal tomographic images", *Computers & Graphics*, vol. 21, no. 3, pp. 347–358, 1997. DOI: 10.1016/S0097-8493(97)00012-5.
- [39] Z. Gao, S. Li, Q. Cai, W. Dang, Y. Yang, C. Mu, and P. Hui, "Relative wavelet entropy complex network for improving EEG-based fatigue driving classification", *IEEE Trans. on Inst. and Meas.*, vol. 68, no. 7, pp. 2491–2497, 2019. DOI: 10.1109/TIM.2018.2865842.
- [40] J. Chen and G. Li, "Tsallis wavelet entropy and its application in power signal analysis", *Entropy*, vol. 16, no. 6, pp. 3009–3025, 2014. DOI: 10.3390/e16063009.
- [41] J. Chen, Y. Dou, Y. Li, and J. Li, "Application of Shannon Wavelet Entropy and Shannon Wavelet Packet Entropy in analysis of power system transient signals", *Entropy*, vol. 18, p. 437, 2016. DOI: 10.3390/e18120437.
- [42] N. Nazmi, M. A. Abdul Rahman, S. I. Yamamoto, and S. A. Ahmad, "Walking gait event detection based on electromyography signals using artificial neural network", *Biomedical Signal Processing and Control*, vol. 47, pp. 334–343, 2019. DOI: 10.1016/j.bspc.2018.08.030.
- [43] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 855–868, 2009. DOI: 10.1109/TPAMI.2008.137.
- [44] N. Wang, E. Ambikairajah, N. H. Lovell, and B. G. Celler, "Accelerometry based classification of walking patterns using time-frequency analysis", in *Proc. of 2007 29th Annual IEEE International Conference on Engineering in Medicine and Biology Society*, 2007, pp. 4899–4902. DOI: 10.1109/IEMBS.2007.4353438.
- [45] B. M. Wilamowski and H. Yu, "Improved computation for Levenberg–Marquardt training", *IEEE Transactions on Neural Networks*, vol. 21, no. 6, pp. 930–937, 2010. DOI: 10.1109/TNN.2010.2045657.
- [46] M. A. Cavuslu and S. Sahin, "FPGA implementation of ANN training using Levenberg and Marquardt algorithms", *Neural Network World*, vol. 2, pp. 161–178, 2018. DOI: 10.14311/NNW.2018.28.010.
- [47] D. Kaya and M. Turk, "Interpretation of disorder diagnosis in bioelectrical signals", *Science & Eng. Journal of Firat Univ.*, vol. 29, no. 1, pp. 263–269, 2017.
- [48] J. Too, A. R. Abdullah, N. M. Saad, and W. Tee, "EMG feature selection and classification using a Pbest-guide Binary Particle Swarm Optimization", *Computation*, vol. 7, no. 1, p. 12, 2019. DOI: 10.3390/computation7010012.
- [49] F. Jadidi, E. Castrillon, and P. Svensson, "Effect of conditioning electrical stimuli on temporalis electromyographic activity during sleep", *Journal of Oral Rehabilitation*, vol. 35, no. 3, pp. 171–183, 2008. DOI: 10.1111/j.1365-2842.2007.01781.x.
- [50] F. Jadidi, O. Norregaard, L. Baad-Hansen, L. Arendt-Nielsen, and P. Svensson, "Assessment of sleep parameters during contingent electrical stimulation in subjects with jaw muscle activity during sleep: A polysomnographic study", *Eur. J. Oral Sci.*, vol. 119, no. 3, pp. 211–218, 2011. DOI: 10.1111/j.1600-0722.2011.00822.x.
- [51] F. Jadidi, K. Wang, L. Arendt-Nielsen, and P. Svensson, "Effects of different stimulus locations on inhibitory responses in human jaw-closing muscles", *Journal of Oral Rehabilitation*, vol. 38, no. 7, pp. 487–500, 2011. DOI: 10.1111/j.1365-2842.2010.02177.x.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) license (<http://creativecommons.org/licenses/by/4.0/>).