

Feature and Window Optimization for SNR Estimation in Power-Line-Contaminated Facial EMG Signals

Nurdeeyana Chemoh and Pornchai Phukpattaranont[†], Non-members

ABSTRACT

Electromyography (EMG) signals are widely used in biomedical and rehabilitation applications; however, they frequently suffer from power-line interference (PLI), which compromises signal quality and impacts subsequent analysis. Therefore, it is important to be able to accurately estimate a signal-to-noise ratio (SNR) to assess data quality and guide the method of noise removal. This study investigates the effects of feature number and window size on the accuracy of SNR estimation in the EMG signal contaminated by the PLI signal. The EMG signal is contaminated with synthetic 50 Hz interference at controlled noise levels. Eight features are derived from windows of differing lengths, and the SNR was estimated utilizing 23 regression-based models. Results indicate that waveform activity (WA) and kurtosis (KURT) were found to be the best for estimating SNR. Using window sizes between 250 ms and 2000 ms, these features produced RMSE values from 3.95 to 3.17, demonstrating that larger windows enhance estimation accuracy. In addition, the application of the proposed model is preliminarily validated with real-world facial EMG signals.

Keywords: Electromyography (EMG), Signal-to-Noise Ratio (SNR), Power-Line Interference, Feature Optimization, Window Size

1. INTRODUCTION

Electromyography (EMG) signals play an essential role in rehabilitation [1-2], diagnosis of neuromuscular diseases [3-4], human-machine interaction [5-6], and facial expression recognition [7-9]. The accuracy and reliability of EMG-based systems depend heavily on the signal quality. However, EMG recordings are corrupted by several types of noise, including motion artifacts, electrode impedance fluctuations, and power-line interference (PLI) originating from surrounding

electrical equipment operating at 50 or 60 Hz. The quality of the electrode-skin interface strongly influences the magnitude of 50/60 Hz PLI [10].

Common signal processing approaches for removing PLI from EMG signals include notch filters, comb filters, and adaptive filters. Strzecha et al. [11] proposed the removal of 50 Hz PLI and its harmonics using a 40th-order comb filter with a 3-dB stopband width of ± 0.5 Hz. Nagasirisha et al. [4] combined a sine-adapted whale optimization algorithm with least mean square (LMS) and recursive least squares (RLS) adaptive filters to eliminate PLI.

Several advanced signal processing techniques have been proposed to mitigate this problem. Chen et al. [12] introduced a synchrosqueezed wavelet transform-based approach to suppress PLI. The method demonstrates improved time-frequency localization with effective noise reduction in EMG recordings. In addition, Ladrova et al. [13] reports a comparison of PLI removal performance among notch filtering, adaptive noise cancellation, and wavelet-based methods.

Recent research has increasingly focused on optimization and decomposition techniques to enhance EMG signal reconstruction in noisy environments. Ma et al. [14] employed variational mode decomposition combined with sub-band thresholding, demonstrating superior filtering performance compared with conventional empirical mode decomposition methods. More recently, Koppolu and Chemmanagat [15] proposed an automated pipeline utilizing mode decomposition techniques to remove both PLI and motion artifacts, thereby improving pattern recognition performance for myoelectric prosthesis control.

This study aims to develop an efficient and accurate estimation of the signal-to-noise ratio (SNR) in EMG signals contaminated by PLI signals, enabling its use as a quantitative indicator of contamination level. Conventional filtering methods typically rely on fixed or model-based noise assumptions and may introduce signal distortion or remove physiologically relevant information when improperly applied. In contrast, the proposed SNR estimation framework provides an objective criterion for determining the necessity of filtering and for evaluating its effectiveness. Since the true SNR is generally unknown in practical recordings, direct estimation from observed EMG signals offers a significant advantage for adaptive preprocessing and signal quality assessment.

Existing SNR estimation approaches have been pro-

Manuscript received on November 29, 2025; revised on February 22, 2026; accepted on February 23, 2026. This paper was recommended by Associate Editor Chanon Warisarn.

The authors are with Department of Electrical and Biomedical Engineering, Faculty of Engineering, Prince of Songkla University, Songkhla, 90110, Thailand.

[†]Corresponding author: pornchai.p@psu.ac.th

©2026 Author(s). This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License. To view a copy of this license visit: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Digital Object Identifier: 10.37936/ecti-ec.2026242.262589

posed, ranging from model-based methods to statistical and feature-driven estimators. These methods often rely on extracting time-domain and frequency-domain features from the EMG signal to infer the underlying noise level. However, the performance of such estimators strongly depends on the selection of signal features and the size of the analysis window used. Inappropriate window size or excessive feature inclusion can degrade estimation accuracy, increase computational cost, and limit real-time applicability. Despite its importance, there has been limited investigation into how the number of features and the choice of window size jointly affect SNR estimation performance, particularly under PLI conditions.

To address this gap, this study investigates the optimization of feature number and window size for accurate and computationally efficient SNR estimation in EMG signals contaminated by PLI signals. The EMG signals are simulated and processed with varying window lengths and feature subsets to systematically analyze their impact on SNR estimation accuracy. The outcomes of this research provide insights into the trade-off between temporal resolution and estimation robustness, offering guidelines for designing real-time EMG monitoring and denoising systems.

2. MATERIALS AND METHODS

This section presents the overall methodology used to develop and evaluate the proposed SNR estimation for EMG signals contaminated with PLI signals. The study consists of four main steps. First, synthetic EMG and PLI signals were generated under controlled conditions to simulate noisy environments. Next, a set of commonly used time-domain features was calculated from the EMG segments to represent their signal characteristics. Several regression models were then employed to estimate the SNR values based on the extracted features. Finally, an optimization strategy was designed to fine-tune the regression models and improve estimation accuracy across various SNR levels. The following subsections provide detailed descriptions of each step.

2.1 Signal Generation

This section describes the details on generating synthetic EMG ($x(i)$) and PLI ($v(i)$) signals. On the one hand, we created a synthetic EMG signal using white Gaussian noise that has a length of 2000 samples and used it as input to a band-pass filter with the frequency response [16]

$$H(f) = \frac{jf_U^2 f}{(f_L + jf)(f_U + jf)^2}, \quad (1)$$

where the lower bound, f_L , was sampled from a uniform distribution between 30 and 60 Hz. The upper bound, f_U , was subsequently defined as $f_U = f_L + c$, where c was independently drawn from a uniform distribution within the range of 30 to 100 Hz. Consequently, f_U spans a

range from 60 to 160 Hz. The duration of the synthetic EMG signal corresponds to 2 s of sampling at 1000 Hz.

On the other hand, a synthetic PLI signal was produced from a sinusoidal wave exhibiting a random phase and frequency ranging from 49.5 to 50.5 Hz, in increments of 0.25 Hz. The synthetic PLI signal was scaled to achieve predefined noise levels and was added to generate the contaminated signal $x(i) + v(i)$ to achieve the SNR in the range of $[-20, 20]$ dB. The SNR is defined in decibels (dB) as:

$$SNR = \log_{10} \left(\frac{P_s}{P_n} \right), \quad (2)$$

where the signal power P_s represents the average power of the clean EMG signal and the noise power P_n represents the average power of the added PLI signal. While P_s can be expressed as

$$P_s = \frac{1}{N} \sum_{i=1}^N x^2(i), \quad (3)$$

P_n is given by

$$P_n = \frac{1}{N} \sum_{i=1}^N v^2(i). \quad (4)$$

The controlling $v(i)$ allowed for generating EMG datasets with known ground-truth SNR values, which were then used as regression targets for model training and evaluation. This procedure ensures consistency between the theoretical SNR definition and the supervised learning framework. A total of 400 synthetic EMG signals contaminated with PLI signals were generated.

2.2 Feature Calculation

In the feature calculation step, eight frequently used features described in Table 1 are determined from the EMG signals contaminated with PLI signals at various SNRs to characterize the signal properties affected by noise. The eight features were selected for their sensitivity to PLI, ability to characterize EMG signal statistics, and suitability for implementation. Energy-related features (EMAV, IEMG, DOV, MAX, and MIN) capture amplitude and power variations associated with SNR changes, while variability-based features (WA and MEOP) reflect local fluctuations induced by sinusoidal interference. The higher-order statistical feature, KURT, was included to characterize distributional changes between quasi-periodic PLI and stochastic EMG activity. Together, these features provide complementary information with low computational complexity, supporting efficient signal quality assessment. The brief details of each feature are as follows.

The EMAV feature computes the weighted average of the absolute values of the EMG signal $x(i)$. The weight $p(i)$ depends on the position in the signal: $p = 0.75$ for the central portion ($0.2N \leq i \leq 0.8N$) and $p = 0.5$ otherwise. This weighting scheme emphasizes the importance of the middle segment of the signal while reducing the impact

Table 1: List of Features (FEs) Used in This Study.

Feature name	Equation
FE1: Enhanced Mean Absolute Value (EMAV)	$EMAV = \frac{1}{N} \sum_{i=1}^N p(i) x(i) $
FE2: Waveform Activity (WA)	$WA = \sum_{k=1}^{N-1} 1(x(k) - x(k+1) > threshold)$
FE3: Integrated EMG (IEMG)	$IEMG = \sum_{i=1}^N x(i) $
FE4: Myoelectric Onset Percentage (MEOP)	$MEOP = \frac{1}{N} \sum_{i=1}^N 1(x(i) \geq threshold)$
FE5: Difference of Variance (DOV)	$DOV = \frac{1}{N-2} \sum_{i=1}^{N-1} (x(i+1) - x(i))^2$
FE6: Kurtosis (KURT)	$KURT = \frac{\frac{1}{N} \sum_{i=1}^N (x(i) - \mu)^4}{(\frac{1}{N} \sum_{i=1}^N (x(i) - \mu)^2)^2}$
FE7: Maximum Amplitude (MAX)	$MAX = \max(x(i))$
FE8: Minimum Amplitude (MIN)	$MIN = \min(x(i))$

of its edges.

The WA calculates the number of times the absolute difference between consecutive signal values exceeds a specified threshold, reflecting the signal's rapid variations. The IEMG feature, on the other hand, is the total sum of absolute values of the signal, representing the overall energy.

The MEOP feature determines the proportion of signal values that exceed a specified threshold (default: 0.016). It is useful for identifying active regions of the signal. The DOV feature measures the variance of the differences between consecutive signal values. It highlights the variability in the signal's amplitude changes.

The KURT feature measures the peakedness of the signal's distribution. A high kurtosis value indicates the presence of sharp peaks, while a low kurtosis value suggests a flatter distribution. Finally, the MAX and MIN features represent the highest and lowest values of the signal within a segment, respectively. These features indicate the peak activation and resting levels of the muscle activity.

2.3 Regression Models

This study utilized eight primary regression model categories, comprising twenty-three submodels, to examine the correlation between input features and the SNR target, as shown in Table 2. The models included Linear Regression (LR), Regression Tree (RT), Ensemble of Regression Trees (ERT), Support Vector Machines (SVM), Efficient Linear Regression (ELR), Gaussian Process Regression (GPR), Neural Networks (NN), and Kernel Approximation Regression (KAR) models. Each group exemplified a unique modeling philosophy, ranging from

Table 2: List of Regression Models Used in This Study.

Regression Models	Submodels
Linear Regression Models (LR)	Linear
	Interactions Linear
	Robust Linear
	Stepwise Linear
Regression Trees (RT)	Fine Tree
	Medium Tree
	Coarse Tree
Ensemble Regression Trees (ERT)	Boosted Trees
	Bagged Trees
Support Vector Machines (SVM)	Linear SVM
	Quadratic SVM
	Cubic SVM
	Fine Gaussian SVM
	Medium Gaussian SVM
	Coarse Gaussian SVM
Ensemble of Linear Regression Models (ELR)	Least Squares
	SVM Ensemble
Gaussian Process Regression (GPR)	Squared Exponential Kernel
	Matern 5/2 Kernel
	Exponential Kernel
	Rational Quadratic Kernel
Neural Network (NN)	Narrow Neural Network
	Medium Neural Network
	Wide Neural Network
	Bi-layered Neural Network
	Tri-layered Neural Network
Kernel Approximation Regression Models (KAR)	Support Vector Machine (SVM)
	Least Squares Regression

straightforward linear assumptions to intricate nonlinear and probabilistic frameworks, thereby facilitating a thorough assessment of predictive performance across various data characteristics.

The LR model uses interaction, robust, and stepwise selection methods to identify linear relationships. The RT and ERT models applied nonparametric and ensemble learning to handle nonlinear patterns and reduce overfitting through bagging and boosting. The SVM and ELR models aim to improve accuracy and computational efficiency. The SVM used linear, polynomial, and Gaussian kernels to model nonlinear data, while the ELR optimized training for large datasets. The GPR model applies probabilistic inference with different kernels to represent various data smoothness levels. The NN model, ranging from narrow to tri-layered architectures, captures complex and hierarchical nonlinear relationships. Finally, the KAR model efficiently handled nonlinear relationships by mapping data into higher-dimensional feature spaces. This transformation allows linear or SVM-based regression to perform effectively in the new domain.

2.4 Performance Evaluation

To evaluate how well the trained regression models generalized to new data, each model is tested using an

independent dataset that is not included in the training process. This assessment aimed to verify the model's ability to predict the SNR of EMG signals contaminated by PLI signals. After the regression model is trained, its predicted SNR is evaluated with two standard error metrics: root mean square error (RMSE) and coefficient of determination (R^2). The RMSE quantifies the average magnitude of prediction error by calculating the square root of the mean squared difference between the actual SNR $x(n)$ and the predicted SNR $\hat{x}(n)$ over N samples:

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (x(n) - \hat{x}(n))^2}. \quad (5)$$

R^2 measures how well a regression model explains the variability of the true SNR data. It indicates the proportion of variance in the actual SNR values captured by the model's predictions. An R^2 value of 1 means perfect prediction, while a value close to 0 indicates a weak ability of the model to explain the variation in the data. The R^2 equation is given by:

$$R^2 = 1 - \frac{\sum_{n=1}^N (x(n) - \hat{x}(n))^2}{\sum_{n=1}^N (x(n) - \bar{x})^2}. \quad (6)$$

where $\bar{x}$ is the mean of the actual SNR values.

These RMSE and R^2 together provide a clear picture of a model's predictive performance. RMSE shows how big the average SNR prediction errors are, with lower values meaning more accuracy. R^2 shows how well the model explains the variability in the target data, with higher values meaning stronger explanatory power. In general, a successful prediction model is characterized by a low RMSE and high R^2 , meaning that its predictions closely match the true values and effectively capture the underlying patterns in the data.

2.5 SNR Estimation and Optimization Strategy

The proposed method for estimating SNR consists of 8 main stages. Details on each stage are as follows.

- **Stage (1):** Generate the EMG signals contaminated with the PLI signals using uniformly random SNR in the range [-20, 20] dB, as detailed in Section 2.1. Three hundred pairs of synthetic EMG and PLI signals were randomly selected.
- **Stage (2):** Normalize each contaminated EMG segment x_j to have unit energy, as described by

$$x_i = \frac{x_j}{\sqrt{\sum_{i=1}^N x_j^2}}, \quad (7)$$

where N is the length of the signal.

- **Stage (3):** Produce 8 features from x_i as described by the equations in Section 2.2 (Table 1).
- **Stage (4):** Normalize each feature by

$$z = \frac{f - \mu}{\sigma}, \quad (8)$$

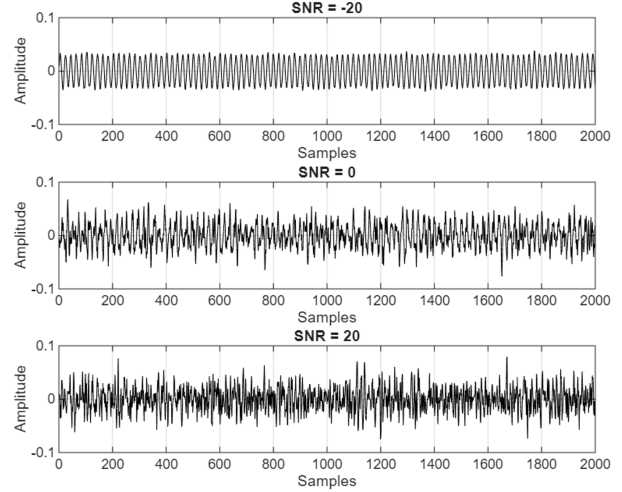


Fig. 1: Simulated signals under three SNR conditions: -20 dB, 0 dB, and 20 dB.

where f is a feature, μ , and σ are the mean and standard deviation of each feature, respectively.

- **Stage (5):** Train the regression models described in Section 2.3 (Table 2) with the normalized features from Stage (4). The input vectors and target vectors will be randomly divided into three sets as follows: 70% will be used for training; 15% will be used to validate that the model is generalizing and to stop training before overfitting; and the last 15% will be used as a completely independent test of generalization.
- **Stage (6):** Evaluate performance between the actual SNR target and the estimated SNR from the regression model using the RMSE and R^2 described in Section 2.4.
- **Stage (7):** Investigate the appropriate number of features and the appropriate size of window for feature calculation.
- **Stage (8):** Validate the proposed method with a real-world facial EMG signal.

3. RESULTS

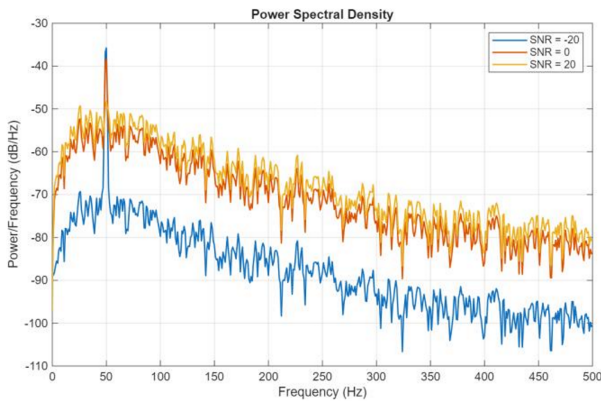
3.1 Estimation Accuracy

Figs. 1 and 2 illustrate the effects of different SNR conditions on simulated signals generated using the method given in Section 2.1 and their spectral characteristics. Fig. 1 presents the simulated signals at SNR levels of -20 dB, 0 dB, and 20 dB. The results indicate that higher SNR values produce clearer EMG signals with less PLI interference. Fig. 2 compares the corresponding power spectral densities (PSDs) of these signals. The results demonstrate that lower SNR conditions cause higher spectral peaks at 50 Hz PLI, whereas higher SNR levels preserve the dominant frequency components and overall spectral structure of EMG signals, which are broadband.

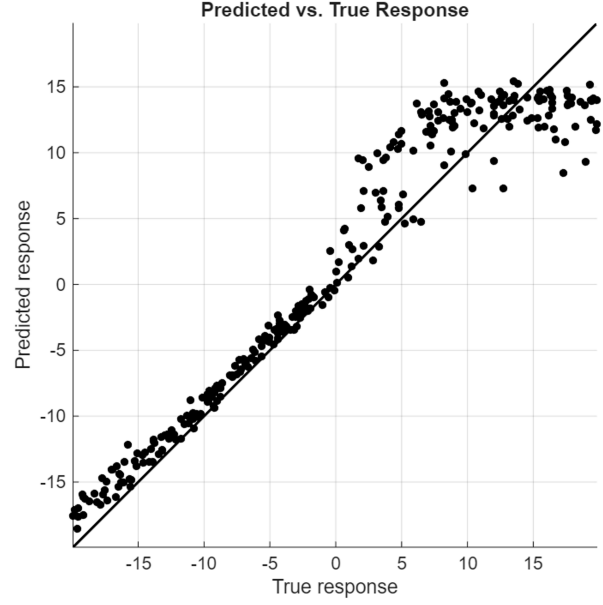
Table 3 shows a performance comparison of various regression models for SNR estimation. This table presents a comprehensive model selection analysis, comparing the

Table 3: Performance Comparison of Various Regression Models for SNR Estimation.

Model	Sub model	RMSE	R^2
LR	Linear	3.35	0.90
	Interaction	3.24	0.91
	Robust	3.48	0.89
	Stepwise	3.28	0.91
RT	Fine	4.01	0.86
	Medium	3.66	0.88
	Coarse	3.79	0.88
SVM	Linear	3.41	0.90
	Quadratic	3.29	0.90
	Cubic	3.30	0.90
	Fine Gaussian	3.48	0.89
	Medium Gaussian	3.27	0.91
	Coarse Gaussian	3.39	0.90
ELR	Least Squares	3.36	0.90
	SVM	3.41	0.90
ERT	Boosted Trees	3.34	0.90
	Bagged Trees	3.53	0.89
GPR	Squared Exponential	3.18	0.91
	Matern 5/2	3.16	0.91
	Exponential	3.34	0.90
	Rational Quadratic	3.17	0.91
NN	Narrow	3.31	0.90
	Medium	3.34	0.90
	Wide	4.77	0.80
	Bilayered	3.43	0.90
	Trilayered	3.60	0.89
KAR	SVM	3.87	0.87
	Least Squares Regression	3.34	0.90

**Fig. 2:** Power spectral density (PSD) comparison of simulated signals at three SNR levels: -20 dB, 0 dB, and 20 dB.

performance of numerous regression models and their various submodels using the RMSE and the R^2 metrics. The results indicate that the GPR is one of the optimal overall choices. The GPR with the Matern 5/2 submodel

**Fig. 3:** Predicted Versus True Response of the GPR with Rational Quadratic Submodel ($n = 300$).

achieves the best performance metrics: the lowest RMSE of 3.16 and a high R^2 of 0.91. Other effective submodels include LR Interaction and SVM Medium Gaussian, both of which also achieved an R^2 of 0.91 with slightly higher RMSE values.

Conversely, the analysis highlights models that are poorly suited for this estimation task. The NN using the Wide submodel delivered the worst fit across the entire selection, resulting in the highest RMSE of 4.77 and the lowest R^2 of 0.80. Similarly, the RT with the Fine submodel also performed poorly (RMSE=4.01, $R^2 = 0.86$).

Fig. 3 shows the predicted versus true responses for the GPR model using the Rational Quadratic submodel with $n = 300$. The plot demonstrates a nearly linear correlation between the predicted and true responses, particularly within the SNR range of -20 to 0 dB. The results indicate that the model accurately captures the underlying signal behavior in this range.

3.2 Effect of Feature Number

This section presents the results of the feature reduction analysis described in Stage (7) of Section 2.4. Table 4 summarizes the performance of the GPR model with the rational quadratic submodel across various feature subsets, labeled FE1 through FE8. The analysis aimed to determine a minimal feature set that maintains high predictive performance, using RMSE and R^2 as evaluation metrics. Most combinations of features did well, with R^2 values between 0.91 and 0.92. The best results came from using a subset that included FE1-FE2 and FE4-FE8, which had a minimum RMSE of 3.16 and an R^2 of 0.92. Simpler subsets such as FE2 combined with FE6 also performed well (RMSE = 3.17, $R^2 = 0.92$).

Table 4: Regression Model Performance Based on Various Feature Subsets (Feature Reduction Analysis).

Feature Reduction	RMSE	R^2
FE1, FE2, FE3, FE4, FE5, FE6, FE7, FE8	3.17	0.91
FE1, FE2, FE4, FE5, FE6, FE7, FE8	3.16	0.92
FE1, FE2, FE4, FE6, FE7, FE8	3.17	0.92
FE2, FE4, FE6, FE7, FE8	3.18	0.92
FE2, FE4, FE6, FE7	3.24	0.91
FE2, FE4, FE6	3.27	0.91
FE2, FE6	3.17	0.92
FE6	3.54	0.90

Table 5: Influence of Window Size on Regression Model Performance Metrics.

Time (ms)	RMSE	R^2
2000	3.17	0.92
1000	3.35	0.91
500	3.56	0.89
250	3.95	0.87

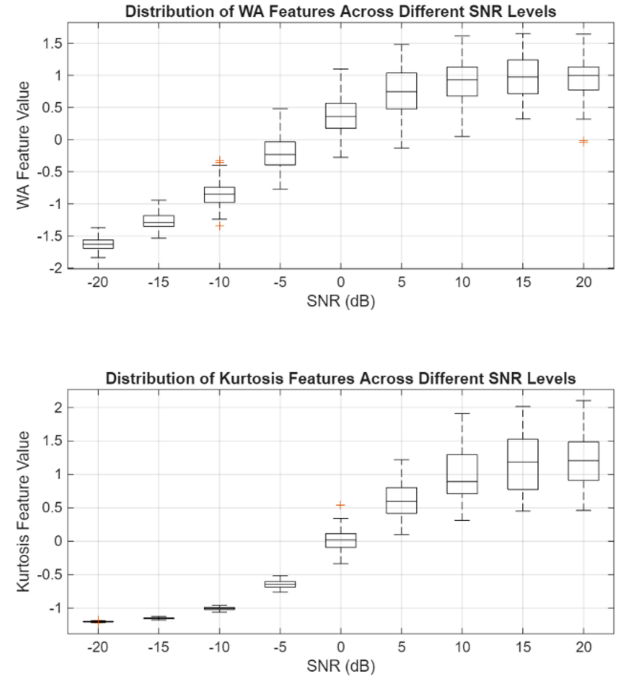
This result indicates that substantial feature reduction is possible without compromising accuracy. However, using only a single feature, such as FE6, led to the poorest performance, with the highest RMSE of 3.54 and the lowest R^2 of 0.90, highlighting the benefit of combining multiple features.

3.3 Effect of Window Size

This section presents the results of the window size analysis described in Stage (7) of Section 2.4. Table 5 summarizes the performance of the GPR model with the rational quadratic submodel. The table demonstrates the impact of window size on the performance metrics. Specifically, it tracks the RMSE and the R^2 value across four different window sizes ranging from 2000 ms down to 250 ms. The results reveal a clear and consistent trend: larger window sizes yield superior model performance. The best result is observed at a window size of 2000 ms, where the model achieves the lowest RMSE of 3.17 and the highest R^2 of 0.92. Conversely, reducing the window size progressively increases the error, resulting in the worst performance (RMSE of 3.95 and R^2 of 0.87) at the smallest size of 250 ms. This suggests that for this specific application, utilizing a broader data segment is critical for capturing robust features, which significantly enhances the overall accuracy and of the estimation model.

4. DISCUSSION

This study systematically analyzed the effects of feature number and window size on the accuracy of SNR estimation in EMG signals contaminated by synthetic

**Fig. 4:** Boxplot illustrating the distribution of extracted EMG features across various SNR levels.

PLI signals. The results highlight the strong dependency of estimation performance on these two parameters and offer useful suggestions for the design of efficient EMG quality monitoring systems. The following discussion provides further insight.

4.1 Estimation Accuracy

Table 3 presents a performance comparison of 23 regression models for SNR estimation using eight feature values. The RMSE values range from 4.77 to 3.16, while the R^2 values range from 0.80 to 0.91. Only two of the 23 models exhibit RMSE values greater than 4 and R^2 values below 0.87. These findings suggest that many regression models achieve comparable and satisfactory accuracy. Therefore, additional performance metrics, such as computation time and model size, should also be considered when selecting a suitable regression model, particularly for real-time or resource-constrained applications.

4.2 Effect of Number of Features

Table 4 presents a performance comparison of SNR estimation using different numbers of features with the GPR model based on the rational quadratic submodel. When the number of features varies from 1 to 8, the RMSE values range from 3.54 to 3.16, while the R^2 values range from 0.90 to 0.92. Even with just two features, FE2 (WA) and FE6 (KURT), the model still does well, with an RMSE of 3.17 and an R^2 of 0.92. These results indicate that we can reduce the features without significantly compromising accuracy due to

their redundancy. Therefore, two optimal features, WA and KURT, are sufficient for practical implementation.

To assess the effectiveness of the WA and KURT features for SNR estimation, boxplots of their values across different SNR levels are presented in Fig. 4. The top panel of Fig. 4 illustrates the WA feature, while the bottom panel depicts the KURT feature. A clear separation of feature values is observed within the SNR range of -20 to 0 dB. These results are consistent with the plot in Fig. 3, which demonstrates an approximately linear relationship between the predicted and true SNR values.

4.3 Effect of Window Size

Table 5 presents a performance comparison of four window sizes for SNR estimation using FE2 (WA) and FE6 (KURT) features when the regression method is the GPR model based on the rational quadratic submodel. As the window size decreases from 2000 to 250 ms, the RMSE increases from 3.17 to 3.95, while the R^2 decreases slightly from 0.92 to 0.87. These results indicate the increased accuracy for larger windows.

This performance improvement can be explained by the statistical properties of feature estimation. Larger window sizes provide a more reliable statistical representation of the EMG signal and the superimposed PLI. The investigated features, namely KURT, are moment-based descriptors whose estimation accuracy depends on the number of samples within the window. From a statistical perspective, increasing the window size increases the sample size, thereby reducing the variance of feature estimates according to the law of large numbers. This leads to more stable feature values and improves the regression model's ability to map features to the true SNR.

In addition to statistical stability, the spectral characteristics of the PLI signal also contribute to the observed trend. The PLI component is a narrowband, quasi-stationary sinusoidal signal at 50 Hz. A longer window captures a greater number of interference cycles, allowing its energy contribution to be more accurately reflected in time-domain statistics. In contrast, short windows may contain only a few cycles, causing fluctuations in estimated power and reducing the consistency of SNR-related features.

Finally, the stochastic and burst-like nature of EMG signals makes short windows more sensitive to transient amplitude variations unrelated to noise level. These local fluctuations introduce variability that weakens the relationship between extracted features and SNR. Larger windows mitigate this effect by averaging transient variations, thereby improving robustness while preserving overall signal characteristics.

For wearable applications, a window length of approximately 500–750 ms offers a practical compromise between accuracy and latency. While longer windows improve statistical stability and estimation reliability, they also introduce latency that may affect real-time

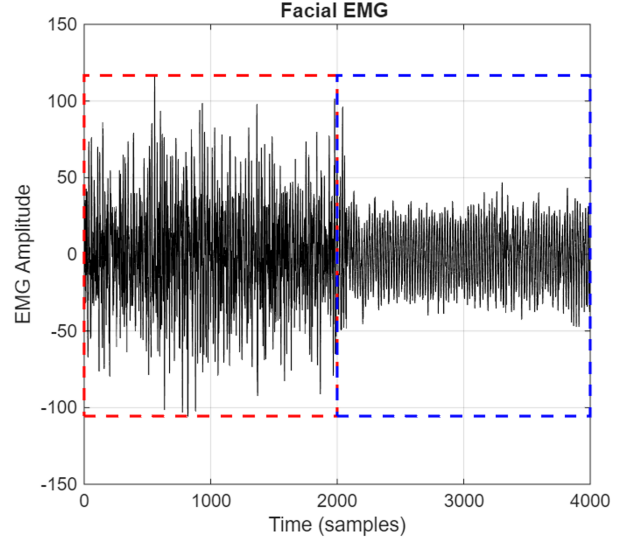


Fig. 5: Facial EMG signal partitioned into EMG-dominant (red dashed box) and PLI-dominant (blue dashed box) regions.

feedback. This limitation can be mitigated by employing overlapping sliding windows, such as 50% overlap, which allow frequent updates without sacrificing estimation performance. Nevertheless, smaller window sizes provide faster response times and may be preferable in applications where rapid adaptation is prioritized over maximum estimation accuracy.

4.4 Validation with a Real-World Facial EMG Signal

The proposed method was validated using a real-world facial EMG signal recorded from the anterior belly of the digastric during a smiling facial expression performed by a 22-year-old male volunteer. The EMG signal was sampled at 1024 Hz. The study protocol was approved by the Human Research Ethics Committee of the Faculty of Medicine, Prince of Songkla University (REC.67-516-38-2). Written informed consent was obtained from the participant prior to data collection.

Fig. 5 shows an example of a real-world facial EMG signal. The red dashed box indicates the EMG segment corresponding to the smiling expression, while the blue dashed box represents the PLI segment during the rest period. The SNR, calculated as the ratio of the signal power in the red segment to that in the blue segment, was 2.38 dB.

When the EMG segment in the blue box was used as input to the GPR model employing the rational quadratic submodel (FE2+FE6), the resulting estimated SNR was 4.01 dB. Compared with the manually calculated SNR of 2.38 dB, the estimated value is greater by 1.63 dB. This difference may be due to the estimation model being trained using Gaussian noise, whereas the manual SNR calculation was based on data containing both Gaussian noise and PLI contamination.

For comparison, a synthetic EMG signal contaminated

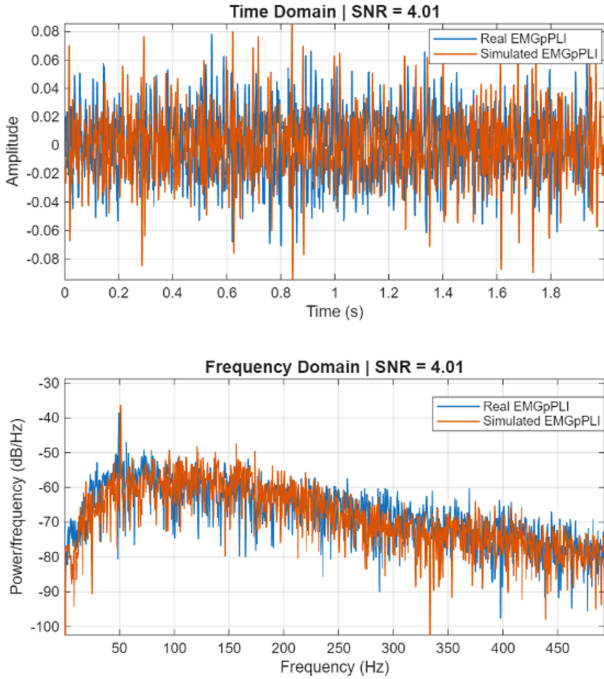


Fig. 6: Comparison Between Real-World and Simulated EMG Signals (SNR = 4.01 dB).

with a synthetic PLI component at an SNR of 4.01 dB was generated. Its waveform and corresponding PSD are shown as red lines in Fig. 6 (top) and (bottom), respectively. The real-world facial EMG segment from the red box in Fig. 5 is plotted in Fig. 6 (top) with a blue line, and its corresponding PSD is depicted with a blue line in Fig. 6 (bottom). The results indicate a similarity between the synthetic and real-world signals, demonstrating the feasibility of applying the proposed method to real-world EMG data.

4.5 Limitations and Future Work

In practical EMG systems, such as myoelectric prosthetic control, fatigue monitoring, and rehabilitation robotics, real-time signal quality assessment is essential to ensure reliable performance. The optimized configuration identified in this study (250 ms window with two features) can be readily implemented in embedded processors for adaptive filtering or data quality feedback. Moreover, since the features used are computationally simple, the proposed approach can operate continuously with minimal processing delay. However, applications requiring higher estimation accuracy than that achieved by the proposed framework may necessitate the development of more advanced algorithms or enhanced modelling approaches.

Although synthetic power-line interference allows precise control over noise levels and repeatability, real-world EMG signals may contain more complex interference patterns, such as harmonics or nonstationary disturbances. Future work should extend the analysis to experimentally recorded PLI-contaminated EMG data

and evaluate model generalization under variable electrode placements and muscle types.

5. CONCLUSION

This study presented a systematic optimization of feature numbers and window sizes for SNR estimation in EMG signals contaminated by PLI signals. Using synthetic PLI and a regression-based estimation model, the results demonstrated that both parameters strongly affect estimation accuracy. Of the eight features tested, waveform activity (WA) and kurtosis (KURT) were found to be the best for estimating SNR. When combined with window sizes ranging from 250 ms to 2000 ms, these features achieved RMSE values between 3.95 and 3.17, indicating improved estimation accuracy with larger windows. In addition, the application of the proposed model is preliminarily validated with real-world facial EMG signals. These findings provide practical guidelines for the design of real-time EMG quality monitoring and adaptive denoising systems, contributing to more reliable signal processing in biomedical and rehabilitation applications.

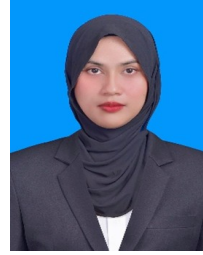
ACKNOWLEDGEMENT

This work was funded by the Faculty of Engineering, Academic Year 2025, Prince of Songkla University. In addition, this work was supported by Prince of Songkla University (Ref. No. ENG6901113S), Thailand Science Research and Innovation (TSRI) [Grant numbers FRB690047/0230], and National Science, Research and Innovation Fund (NSRF).

REFERENCES

- [1] T. Haab, P. Leinen, and P. Burkey, "Role and Effectiveness of Surface EMG Feedback in Sports and Orthopedic Rehabilitation: A Systematic Review," *Exploration of Musculoskeletal Diseases*, vol. 2, pp. 391–407, 2024.
- [2] M. Al-Ayyad, H. A. Owida, R. De Fazio, B. Al-Naami, and P. Visconti, "Electromyography Monitoring Systems in Rehabilitation: A Review of Clinical Applications, Wearable Devices and Signal Acquisition Methodologies," *Electronics*, vol. 12, no. 7, Art. no. 1520, 2023.
- [3] E. Shandiz et al., "Applying High-Density Surface EMG to the Study of Neuromuscular Disorders: A Systematic Review," *Clinical Neurophysiology*, vol. 179, Art. no. 2110983, 2025.
- [4] R. K. Mani and B. Nagaraj, "Diagnosis and Classification of Neuromuscular Disorders using Bi-LSTM Optimized with Grey Wolf Optimizer for EMG Signals," *Scientific Reports*, vol. 15, Art. no. 19274, 2025.
- [5] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Intelligent Human-Computer Interaction Based on Surface EMG Gesture Recognition," *IEEE Access*, vol. 7, pp. 61378–61387, 2019.

- [6] J. Eby, M. Beutel, D. Koivisto, *et al.*, “Electromyographic Typing Gesture Classification Dataset for Neurotechnological Human–Machine Interfaces,” *Scientific Data*, vol. 12, Art. no. 440, 2025.
- [7] M. Kołodziej, A. Majkowski, and M. Jurczak, “Acquisition and Analysis of Facial Electromyographic Signals for Emotion Recognition,” *Sensors*, vol. 24, no. 15, Art. no. 4785, 2024.
- [8] L. Shu, V. R. Barradas, Z. Qin, and Y. Koike, “Facial Expression Recognition Through Muscle Synergies and Estimation of Facial Keypoint Displacements Through a Skin-Musculoskeletal Model Using Facial sEMG Signals,” *Frontiers in Bioengineering and Biotechnology*, vol. 13, Art. no. 1490919, 2025.
- [9] J. M. Rutkowska, T. Ghilardi, S. V. Vacaru, *et al.*, “Optimal Processing of Surface Facial EMG to Identify Emotional Expressions: A Data-Driven Approach,” *Behavior Research Methods*, vol. 56, pp. 7331–7344, 2024.
- [10] K. Strzecha, M. Duchnowski, T. Plechawska-Wójcik, and P. Forczmański, “Processing of EMG signals with High Impact of Power Line and Cardiac Interferences,” *Applied Sciences*, vol. 11, no. 10, Art. no. 4625, 2021.
- [11] B. Nagasirisha, M. Srikanth, V. Sailaja, and V. Prasad, “Reconstruction of EMG Signals from Noisy Environment using Sine Adapted Whale Optimization Algorithm,” *ECTI Transactions on Electrical Engineering, Electronics, and Communications*, vol. 23, no. 2, 2025.
- [12] J. Chen, Y. Sun, S. Sun, and Z. Yao, “Reducing Power Line Interference from sEMG Signals Based on Synchrosqueezed Wavelet Transform,” *Sensors*, vol. 23, no. 11, Art. no. 5182, 2023.
- [13] M. Ladrova, R. Martinek, J. Nedoma, and M. Fajkus, “Methods of Power Line Interference Elimination in EMG Signal,” *Journal of Biomimetics, Biomaterials and Biomedical Engineering*, vol. 40, pp. 64–70, 2019.
- [14] S. Ma, B. Lv, C. Lin, X. Sheng, and X. Zhu, “EMG Signal Filtering Based on Variational Mode Decomposition and Sub-band Thresholding,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 47–58, 2020.
- [15] P. Kumar Koppolu and K. Chemmanagat, “A Novel Procedure to Automate the Removal of PLI and Motion Artifacts using Mode Decomposition to Enhance Pattern Recognition of sEMG Signals for Myoelectric Control of Prosthesis,” *Biomedical Physics & Engineering Express*, vol. 10, no. 6, 2024.
- [16] P. McCool, G. D. Fraser, A. D. C. Chan, L. P. Petropoulakis, and J. J. Soraghan, “Identification of Contaminant Type in Surface Electromyography (EMG) Signals,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 774–783, 2014.



Nurdeeyana Chemoh is currently a master's degree student in the Department of Electrical and Biomedical Engineering, Faculty of Engineering, Prince of Songkla University (PSU), Songkhla, Thailand. She graduated with a bachelor's degree in biomedical engineering from PSU. Her main topic of research is closely related to electromyogram (EMG) signal processing.



Pornchai Phukpattaranont received the B.Eng. (Hons.) and M.Eng. degrees in electrical engineering from the Prince of Songkla University, Songkhla, Thailand, in 1993 and 1997, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, MN, USA, in 2004. He is currently a professor of electrical and biomedical engineering at the Prince of Songkla University. Examples of his ongoing research include the pattern

recognition system based on electromyographic signals, electrocardiographic signals, and microscopic images of breast cancer cells. His current research interests include signal and image analysis for medical applications and artificial intelligence for medical applications (E-mail: pornchai.p@psu.ac.th).