Voxel Weight Matrix-Based Feature Extraction for Biomedical Applications

FAHAD ALBALAWI1,2,*, SULTAN ALSHEHRI1,*, ABDERRAZAK CHAHID1,*, (Member, IEEE), and TAOUS-MERIEM LALEG-KIRATI1, (Senior member, IEEE)

1Department of Electrical Engineering, Taif University, Taif, Saudi Arabia, 21955 (e-mail: f.albiloi@tu.edu.sa)
2Computer Electrical and Mathematical Science and Engineering (CEMSE) division, King Abdullah University of Sciences and Technology (KAUST), Thuwal, KSA (e-mails: fahad.albalawi@kaust.edu.sa, sultanmg.alshehri@gmail.com, abderrazak.chahid@kaust.edu.sa, TaousMeriem.Laleg@kaust.edu.sa)
*These authors are shared first co-authors having contributed equally.
Corresponding author: Taous-Meriem Laleg-Kirati (e-mail: TaousMeriem.Laleg@kaust.edu.sa).

This research project has been jointly funded by King Abdullah University of Science and Technology (KAUST) Base Research Fund (BAS/1/1627-01-01).

ABSTRACT Functional Magnetic Resonance Imaging (fMRI) is an emerging medical tool used to measure brain activities that were induced normally such as cognitive states (e.g., reading a sentence or viewing a picture) or abnormally (e.g., brain activity occurs after a stroke or brain injury). These measured data can be used to construct a model via machine learning techniques to predict the occurrence of a certain cognitive behavior or brain disease. The difficulty of this prediction problem can be summarized in two points: first, the size of the dataset is very small due to the small number of subjects (i.e., patients) who can contribute to these research-based experiments. Second, the size of the feature vector resulted from these medical tools is very large compared to the few number of samples that were collected. One possible way to overcome these obstacles is to develop a feature generation methodology that can produce a small-sized and descriptive feature vector that may improve the overall prediction performance. Motivated by these considerations, this paper proposes a novel feature generation methodology termed Voxel Weight Matrix (VWM) method. This feature generation technique can transform the original high-dimensional feature vector to a two-dimensional discriminative feature domain. The main contribution of this feature generation technique is its ability to represent the statistical measures of the original feature vector via a two-dimensional feature vector. After generating the VWM-based feature set, various classification tools such as logistic regression (LR) models and Support Vector Machine (SVM) are used for cognitive state prediction based on publicly available fMRI dataset called state/plus dataset. The classification models with the proposed VWM features outperformed the best two reported prediction models associated with the star/plus dataset with an average accuracy of 99.8%. To further illustrate the effectiveness of the proposed feature generation methodology, another publicly available Electroencephalography (EEG) dataset are used for Epileptic Seizure Prediction.

INDEX TERMS Feature extraction, fMRI, Voxel Weight Matrix (VWM), quantization, classification, EEG and cognitive state.

I. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) has emerged as a powerful brain imaging modality that measures indirectly the brain activity with good spatial and temporal resolution. It provides three-dimensional scans of the brain per unit of time which allows the detection of the activated regions as a response to a neural activity [1]. Since its discovery [2], fMRI has been used for the detection and monitoring of neurological diseases and disorders such as Schizophrenia [3] and Alzheimer disease [4]. fMRI captures the changes in blood oxygen levels that occur in the activated regions in the brain which provides a signal called Blood Oxygen Level Dependent (BOLD) signal. The BOLD signal is triggered by changes in the cerebral blood flow and reflects the increase in the deoxyhemoglobin content. fMRI modality has been used in many clinical and research conditions allowing the generation of a huge amount of data; as a result, several studies focused on the use of data-driven approaches
to analyze, interpret and extract relevant information from fMRI data [5]–[8]. In addition, fMRI data were used to estimate parameters that help model efficiently the relation between a neural stimulus and the observed BOLD signal [9]. To improve the analysis of brain diseases, a framework that was proposed in [10] utilizes fMRI data to estimate the states and parameters of the stochastic Metabolic Hemodynamic Model (sMHM). A recent research work [11] investigated the influence of hunger and satiety on rs-fMRI using three connectivity models namely; local connectivity, global connectivity and amplitude rs-fMRI signals.

A particular interest has been on the development of machine learning methods to classify the cognitive state of a human subject based on fMRI data. For instance, in [12] an fMRI study was conducted to investigate the brain activity when bilingual adults read English and Spanish words. Recently, a machine learning approach was developed in [13] to differentiate fMRI Data of Chronic Fatigue Syndrome (CFS) from a sedentary control where fMRI data of 69 patients were used to demonstrate the proposed scheme. In [14], the authors decoded cognitive states that correspond to distinct brain tasks such as viewing a sentence/picture and reading an ambiguous/non-ambiguous sentence. The main challenge that the authors have addressed when decoding human cognitive states is the big discrepancy between the number of available samples for a given cognitive state and the dimension of the feature vectors. Therefore, optimal search techniques can be developed to choose an optimal feature vector with a well-suited classifier [15].

Motivated by the above state of art, a cascade of classifiers that can improve the cognitive state prediction performance has been proposed [16]. However, the issue of over-fitting associated with the fMRI data has not been considered. Therefore, the authors in [17], [18] have developed a group of predictors named Generalized Sparse Classifiers (GSC) to address the issue of over-fitting caused by the high dimensional feature vector. The derived group of classifiers were applied to a benchmark dataset called star/plus [19], and achieved an average accuracy of 93.7%. In [20], [21], an algorithm called Support Vector DecomVoxel Machine (SVDM) that combines feature selection and classification learning into one single step was developed. Although the SVDM was able to project the dataset into 8 features, its prediction performance to the star/plus dataset was unsatisfactory with an average accuracy of 78%. In [22], [23], the authors have proposed a procedure to enable classification between two chosen cognitive tasks, using their respective fMRI image sequences. Different classification methods with two signal processing-based features were applied to the star/plus dataset where an average accuracy of 99% was obtained using the support vector machine algorithm. Even though the proposed method was able to achieve high prediction accuracy, the dimension of the extracted feature vector was extremely high. For classification purposes, it is highly desirable to reduce the feature vector size when the number of observations (i.e., samples) used to generate a prediction model is small [24]. Therefore, a new feature generation method with well-suited classifier that can both reduce the high-dimensional fMRI feature vector and improve the overall prediction performance is necessary.

Inspired by the above considerations, we propose a novel methodology that can generate a set of features termed Voxel weight matrix-based features. This set of features can represent the voxel activity in the human brain when performing cognitive tasks. The main advantage of this new feature set is its ability to project the high-dimensional voxels features vec-
tor into a two-dimensional feature domain. Star/plus dataset has been used to assess the performance of the proposed features when they are used to classify two cognitive tasks. The paper is organized as follows: in the second and third sections, we provide the materials and methods used to generate the proposed VWM-based feature vector, respectively. In the fourth section, the machine learning model and the training/testing technique utilized to assess the performance of the proposed work with respect to the best two performing algorithms are given. In addition, we further demonstrate the applicability of the VWM feature extraction when it is applied to a different type of dataset (EEG dataset for epileptic seizure prediction). Finally, a brief summary about the proposed work is provided in the last section.

II. MATERIALS

In this paper, we mainly focus on the classification of cognitive behavior. For this purpose, we use the publicly available fMRI dataset called star/plus dataset [19]. This dataset was used to demonstrate the effectiveness of the proposed feature methodology. In such experiment, fMRI snapshots were obtained every half second (repetition time) when six subjects were performing two distinct cognitive tasks. Particularly, every subject first sees a sentence (semantic stimulus) or a picture (symbol stimulus) for 4 seconds, then a blank screen for 4 seconds is shown to the subject. Every sample is a collection of fMRI 8-seconds period: 4-seconds period of sentence or picture stimulus followed by 4-seconds period of blank screen. Following this strategy, a total of 80 samples are generated from each subject (40 samples for sentence class and 40 samples for picture class). Every sample includes 16 fMRI snapshots, resulting in an input feature vector of size $16N$ where $N$ represents the number of active voxels in a particular Regions Of Interest (ROIs) when the subject sees either a picture or a sentence. Due to the variation of the brain morphology between subjects, the number of active voxels $N$ within the ROIs is different for each subject. The dataset consists of 25 anatomically defined ROIs. Based on the hint mentioned in [19], only seven of these regions were recommended to be used when one trains a classifier to distinguish whether the subject is viewing a picture or sentence. In this work, we use fMRI dataset collected from these seven ROIs and subsets of them to compare the performance of the proposed VWM-based feature generation technique with respect to the best reported prediction models.

III. METHODS

In this section, the proposed VWM-based features generation methodology for the aforementioned fMRI dataset will be presented. Due to the fact that the proposed methodology is partially inspired by a biological feature generation technique called Position Weight Matrix (PWM)-based features, a brief background about PWM is provided in the first subsection. Then, we extend this feature generation methodology to generate the proposed VWM-based features. To do so, the second subsection introduces a quantization scheme, a pre-processing step before generating the VWM-based features, for the voxel intensity signal that can convert the discrete signal $X(n)$ into a sequence of finite set of symbols. Figure 1 depicts the framework of the VWM-based features generation methodology beginning from the raw data (i.e., fMRI snapshots) and resulting in the final VWM-based feature vector.

A. BACKGROUND ON PWM

Position weight matrix based features extraction is a popular method used for motifs representation in DNA/RNA sequences [25], [26]. In this features extraction paradigm, two Position weight matrices are often extracted from a set of aligned DNA/RNA sequences that are believed to be functionally related. This feature generation technique has a significant role in many software tools for computational motif discovery [27]. Traditionally, two PWMs are usually derived from two sets of aligned DNA sequences that are thought to be functionally related. Then, these two matrices are utilized to generate number of features that may help improve the classification performance [28]. In order to construct the PWMs, the dataset or the samples have to belong to a finite set of integers or characters. Naturally, the DNA sequences are represented by a set of four nucleotides A, C, G and T. To generate the two PWM matrices, two groups of DNA sequences that belong to two distinct classes are first aligned, then the probability of occurrence of each of the four DNA nucleotides (A,C,G and T) is calculated at each Voxel of the DNA sequence. The probability of occurrence of each DNA nucleotide in a certain Voxel (A,C,G or T) is equal to the number of occurrences of such nucleotide divided by the total number of sequences.

Example 1: in this example, we illustrate the PWM construction by assuming that we have two classes of DNA sequences (e.g., positive class and negative class). In each class, we have 4 DNA samples where the length of each sample is 5. Then, the PWM for every class is of size $4 \times 5$. When the DNA sequences in one of these two classes take the following structure:

$$\text{DNA sequences} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}$$

Then the corresponding PWM for the above class of DNA samples is as follows:

$$\text{VWM} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}$$

In every column of the above PWM, we have four probabilities that sum up to 1 where every element in that column indicates the probability of occurrence of one of the four DNA nucleotides. For the star/plus fMRI dataset, the features are real-valued voxel intensities measured at different time
steps. Therefore, a quantization scheme that can transform the real-valued feature vector into a finite set of symbols or characters depending on the mapping criteria is required in order to construct the Voxel matrices. In the following subsection, we introduce a quantization scheme that we consider in this work.

### B. QUANTIZATION SCHEME

Quantization is the process of mapping signals from continuous set to output values in a (countable) smaller set [29]. In this work, we define the quantization mapping \( F: \mathbb{R} \rightarrow \Omega \) as follows:

\[
F(X) = \begin{cases} 
q_1 \quad & \text{if } X < \mu + \frac{(2 - M)}{2} r \\
q_{k+M/2} \quad & \text{if } \mu + (k - 1) r \leq X < \mu + k r \\
q_M \quad & \text{elsewhere}
\end{cases}
\]

such that:

\[
\Omega = \{q_1, q_2, \ldots, q_M\} ; \quad k = \frac{-M}{2} + \frac{2(M - 1)}{3} \frac{M}{2} - 1.
\]

where \( \mu \) and \( r \) are the centroid and the resolution of the quantization which will be defined later in this section.

The mapping \( F \) is utilized to convert the voxels intensity discrete sequence \( X(n) \) to a symbol sequence \( Q(n) = F(X(n)) \). In order to choose a suitable quantization scheme, we first need to analyze the probability distribution of the real-valued voxels intensity sequence for both classes (i.e., picture and sentence). To do so, we observed the histogram of the voxels intensity values of the six subjects for both classes. Figure 2 shows the fitted Gaussian Probability Distribution Function (PDF) to the histogram plots for both classes where we zoomed in at the region where there is a discrepancy between the two Gaussian distributions.

![Figure 2](image)

**FIGURE 2.** Voxels intensity histogram for the six subjects.

To quantify the voxels intensity values (shown in figure 2) for both classes, an appropriate quantization scheme that can transform \( X(n) \) into \( Q(n) \) is implemented. A schematic diagram that illustrates the implementation strategy is depicted in Figure 3. The main parameters that can adjust the quantization scheme are the number of levels \( M \) and the resolution \( r \) where the scheme shown in figure 3 sets these values to 8 and one standard deviation \( \sigma \), respectively. In this study, the \( k^{th} \) quantization interval denoted as \( l_k \) is defined as follows:

\[
l_k = [\mu + (k - 1) r, \mu + k r],
\]

\[
r = \alpha \sigma
\]

where \( k = \frac{-M}{2} + 1, \ldots, \frac{M}{2} \) and \( \alpha \) are positive scaling factors, \( \mu \) and \( \sigma \) are the average mean and standard deviation of the six subjects defined as:

\[
\mu = \frac{1}{6} \sum_{n=1}^{6} \mu_n \quad \text{and} \quad \sigma = \frac{1}{6} \sum_{n=1}^{6} \sigma_n,
\]

where \( \mu_n \) and \( \sigma_n \) are the mean and the standard deviation of the voxels intensity for the \( n^{th} \) subject. In probability theory and statistics, the probability that a Gaussian random variable \( H \) is greater than \( 3\sigma + \mu \) is almost zero. This observation is the well-known rule in statistics called 3-sigma rule [30] [31]. Therefore, the most significant variabilities and randomness that characterizes a random sequence can still be observed if

\[
|h - \mu| \leq 3\sigma
\]

Since the 3-sigma rule indicates that most of the information of a Gaussian random variable is located within \( 3\sigma + \mu \), the quantization interval length of the proposed quantization scheme is chosen to be an integer multiple of the standard deviation \( \sigma \) as shown in Figure 3.

![Figure 3](image)

**FIGURE 3.** The quantization of the voxels sequence with a resolution \( r = \sigma \) and \( M = 8 \). The percentage values reflect the significance of each real-valued interval in the input signal.

The implementation strategy of the proposed quantization scheme is defined in Algorithm 1. The output of this quantization scheme is a symbol sequence that can only take finite set of symbols (i.e., \( \{q_1, \ldots, q_M\} \)). When this scheme is applied to the real-valued voxel intensity samples for both classes, two matrices of size \( 40 \times 16N \), namely a picture matrix \( P \) and a sentence matrix \( S \), will be generated. These two integer-valued matrices \( P \) and \( S \) will be utilized to generate the final VWM-based feature vector.

**Remark 1:** The proposed quantization scheme depicted in Algorithm 1 can be applied to different types of dataset that represent the voxels intensity when a subject is performing...
### Algorithm 1 Uniform Quantizer Algorithm

**Input:** $X$: Real-valued voxel sequence  
- $M$: Number of Quantization levels  
- $\mu$: Quantization centroid  
- $r$: Quantization resolution  

**Output:** $Y$: Quantized sequence

$$N_X \leftarrow \text{length of } X$$

for $n \leftarrow 1$ to $N_X$ do

for $k \leftarrow \frac{2 - M}{2}$ to $\frac{M}{2}$ do

if $X(n) < \mu + \frac{(2 - M) r}{2}$ then

$Q(1) = q_1$;

else if $\mu + (k - 1) r \leq X(n) < \mu + k r$ then

$Q(k) = q_{(k+M/2)}$;

else

$Q(M) = q_M$;

end

end

end

A specific cognitive task (not necessarily viewing a picture or a sentence). The quantization parameters $M$ and $r$ will be chosen based on the distribution of the real-valued voxels sequences that are related to such cognitive task.

### C. VWM-BASED FEATURES GENERATION

To extract the VWM-based features, we extend the existing VWM technique and propose a methodology that can generate a reduced feature vector of two values. These two values provide a probabilistic representation of the voxel pattern in response to two distinct cognitive tasks. This feature vector can both overcome the high dimensionality of the feature vector associated with cognitive state prediction, and represent the statistical distribution of the values of the original feature vector. To generate such a feature vector, the following two steps need to be implemented:

**Step 1:** When Algorithm 1 is applied to all the picture and sentence voxel intensity sequences (samples), two matrices (picture $P$ and sentence $S$ matrices) are constructed. Every entry of these matrices can take one of the $M$ symbols $(q_1, \ldots, q_M)$ generated by Algorithm 1. After that, two weighting matrices namely the Picture Voxel Weight Matrix (PVWM) and the Sentence Voxel Weight Matrix (SVWM) can be constructed as exactly the way that the Voxel Weight matrices (VWMs) are generated in Example 1. For instance, when the number of quantization levels $M$ of Algorithm 1 is set to 6 and for any resolution $r$, the size of the two Voxel Weight-based matrices PVWM and SVWM will be $6 \times 16N$. The matrix structure of the PVWM and the SVWM can take the following form:

$$\begin{pmatrix}
0.1 & 0.5 & 0.4 & 0.1 & \ldots \ W(1, 16N) \\
0.2 & 0.1 & 0.2 & 0.1 & \ldots \ W(2, 16N) \\
0.1 & 0.1 & 0.1 & 0.1 & \ldots \ W(3, 16N) \\
\vdots & \vdots & \vdots & \vdots & \ddots \ W(4, 16N) \\
0.3 & 0.1 & 0.1 & 0.2 & \ldots \ W(5, 16N) \\
0.2 & 0.1 & 0.1 & 0.2 & \ldots \ W(6, 16N)
\end{pmatrix}$$

Step 2: In step 1, the frequencies of any of the $M$ quantized levels along the picture and sentence sequences were used to derive the PVWM and the SVWM. In this step, we utilize these two matrices to compute two scores for every integer-valued sequence. These two scores indicate the likelihood of the sequence to be a picture or a sentence sequence. Following this strategy, two scores are computed for each sequence in the dataset and are calculated as follows:

Let $S_j$ be the $j^{th}$ row of the sentence matrix $S$, and similarly $P_j$ be the $j^{th}$ row vector of $P$. From step 1, $PVWM_{i,j}$ and $SVWM_{i,j}$ represent the probability of occurrence of the symbol $q_i$ where $i = 1, \ldots, M$ at a time instance $j$ along the picture and the sentence sequences, respectively. Therefore, the two scores for every given quantized sequence can be calculated as follows:

$$Score_1 = \sum_{j=1}^{16N} PVWM_{i,j}$$

$$Score_2 = \sum_{j=1}^{16N} SVWM_{i,j}$$

Originally, every picture or sentence sample is represented by a $16N \times 1$ feature vector. However, after applying the proposed feature generation methodology, this high dimensional feature vector will be mapped into a two-dimensional feature vector. The size of the full feature matrix is $80 \times 2$ where half of these samples represents the picture trials and the other half represents the sentence trials. Finally, a prediction model can be derived using this reduced feature vector.

Remark 2: The proposed VWM-based feature generation methodology can be generalized to multi-classification problems. These classification problems may arise when one needs to distinguish between three or more cognitive tasks based on the fMRI dataset associated with these cognitive tasks. The number of VWM-based matrices will be equal to the number of classes. The star/plus dataset used here provides only fMRI dataset for two cognitive tasks; therefore, we
TABLE 1. Comparison of the results obtained by the LR classifier trained on the proposed VWM-based features and the state-of-art methods [23], [32] with dataset collected from the ROI CALC

<table>
<thead>
<tr>
<th>Star/plus Subject</th>
<th>Method1 [23]</th>
<th>Method2 [32]</th>
<th>VWM method $[M = 6, r = \sigma]$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>classifier</td>
<td>feature size</td>
<td>Accuracy</td>
</tr>
<tr>
<td>4799</td>
<td>NB</td>
<td>4080</td>
<td>45</td>
</tr>
<tr>
<td>4820</td>
<td>NB</td>
<td>6528</td>
<td>90</td>
</tr>
<tr>
<td>4847</td>
<td>NB</td>
<td>5088</td>
<td>100</td>
</tr>
<tr>
<td>5675</td>
<td>NB</td>
<td>6384</td>
<td>95</td>
</tr>
<tr>
<td>5680</td>
<td>NB</td>
<td>5344</td>
<td>95</td>
</tr>
<tr>
<td>5710</td>
<td>NB</td>
<td>3504</td>
<td>100</td>
</tr>
<tr>
<td>Average</td>
<td>-</td>
<td>5155</td>
<td>87.50</td>
</tr>
</tbody>
</table>

TABLE 2. Comparison of the results obtained by the LR classifier trained on the proposed VWM-based features and the state-of-art methods [23], [32] with dataset collected from 4 ROIs [CALC', LIPL', LIPS' and 'LOPER']

<table>
<thead>
<tr>
<th>Star/plus Subject</th>
<th>Method1 [23]</th>
<th>Method2 [32]</th>
<th>VWM method $[M = 6, r = \sigma]$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>classifier</td>
<td>feature size</td>
<td>Accuracy</td>
</tr>
<tr>
<td>4820</td>
<td>NB</td>
<td>12240</td>
<td>82.5</td>
</tr>
<tr>
<td>4847</td>
<td>NB</td>
<td>26100</td>
<td>98.75</td>
</tr>
<tr>
<td>4799</td>
<td>NB</td>
<td>19400</td>
<td>98.75</td>
</tr>
<tr>
<td>5675</td>
<td>NB</td>
<td>24000</td>
<td>97.5</td>
</tr>
<tr>
<td>5680</td>
<td>NB</td>
<td>18222</td>
<td>96.25</td>
</tr>
<tr>
<td>5710</td>
<td>NB</td>
<td>12231</td>
<td>95</td>
</tr>
<tr>
<td>Average</td>
<td>-</td>
<td>18699</td>
<td>94.79</td>
</tr>
</tbody>
</table>

generated only two voxel weight matrices (namely PVWM and SVWM).

Remark 3: In this paper, we named the two weighting matrices as Picture and Sentence Voxel Weight Matrices PVWM and SVWM, respectively. However, the choice of the terminologies was just to clarify the relationship between the voxel intensities of the two classes and the output feature vector. One can use the same feature generation technique with these two matrices where one can name them as for example first weight matrix and second weight matrix for a different dataset.

Remark 4: Principle Component Analysis (PCA) is a well-known statistical feature reduction technique that utilizes an orthogonal transformation to transform the set of correlated values of the original feature vector into a set of linearly uncorrelated values. The size of the final feature vector of the PCA could be of any size which depends heavily on the degree of correlation between the values of the original feature vector. However, in the VWM, the size of output feature vector is always two and these two values represent the projection of each class on both its own weight matrix and the other class weight matrix.

IV. RESULTS
A. MACHINE LEARNING MODEL

Due to the high dimensionality of the feature vector compared with the small number of samples, the Leave-One-Out (LOO) cross-validation scheme was used to avoid a biased measure of test accuracy. The generated features were trained using several classifiers: Logistic Regression (LR), Support Vector Machine (SVM), Naive Bayes classifier (NB) and multilayer perceptron (MLP) with two hidden layers, and Sparse Multinomial Logistic Regression (SMLR). Unlike most of the features used for classification, the VWM-based features cannot be generated independently. This is because the VWM-based features are correlated in the sense that the features cannot be extracted without the knowledge of the other training dataset (i.e., the whole training dataset needs to be processed together in order to generate the PVWM and SVWM). Therefore, one should be careful when implementing such features. To do so, the PVWM and SVWM are reconstructed for every training dataset (79 samples) that correspond to every fold of the 80 leave-one-out cross validation folds. The reason of doing this is that in any classification problem the testing set should not be included in the learning stage. Hence, the PVWM and the SVWM have to be re-calculated for every leave-one-out fold.

B. COMPARISON WITH EXISTING METHODS

For every subject of the star/plus dataset, the voxel intensity sequence $X(n)$ was first quantized using 6 levels and a resolution of one standard deviation (i.e., $M = 6$ and $r = \sigma$), and then the VWM-based features were generated as shown in section 3. To illustrate the effectiveness of the proposed VWM-based features, we compare the performance of the three classifiers (LR, SVM and NB) trained on the proposed VWM-based features with the best performing classifiers reported in literature ([23], [32]). For a fair comparison, the performance of the proposed VWM-based method was compared with prediction models that utilize feature vectors derived from 7 ROIs (CALC', LDLPFC, LIPL, LIPS, LOPER, LT and LTRIA), 4 ROIs (CALC, LIPL, LIPS and LOPER) and the "CALC" ROI. In [23], [32], the authors applied the Naive Bayesian (NB) and the Support Vector...
Machine (SVM) classifiers to feature vectors that were derived from all the 7 ROI, respectively. Table 1 reports the results obtained by using the feature vectors derived from the "CALC" ROI under method 1 ([23]), method 2 ([32]) and the proposed VWM-based method. In addition, Table 2 and 3 depict the results of the classifiers when they utilize feature vectors derived from the 4 ROIs and the 7 ROIs, respectively.

The two metrics used to compare these classifiers are the size (i.e., dimension) of the feature vector and the prediction accuracy. To illustrate the overall performance of each method, we compute also the average feature size and the average accuracy of the prediction model. From Table 1, the average accuracy of the cognitive state prediction problem obtained by the VWM-based prediction method was improved by about 2.3% and 14% compared to that obtained by method 2 and method 1, respectively. These prediction performances were achieved when only the 'CALC' ROI was used for feature generation. Tables 2 and 3 show that the VWM-based method improves the average accuracy of method 1 by 4.2% and 4.7% when 4 ROIs and 7 ROIs were used for feature generation, respectively. Similarly, the VWM-based method outperformed the average accuracy of method 2 by 1.43% and 2% when 4 ROIs and 7 ROIs were used for feature generation, respectively. In all the cases, the VWM-based method reduced significantly (two-dimensional) the size of the feature vector used for the other two methods specially when all the 7 ROIs were used for feature generation.

C. SENSITIVITY ANALYSIS

Table 4 depicts the sensitivity analysis that we performed in which the average accuracy across the six subjects was computed for every scenario. The purpose of this analysis is to investigate the impact of the relationship between the statistical measures (µ and σ) of the voxel intensity sequence X(n) and the quantization parameters (M and r) on the average accuracy of the prediction model. From Table 4, one can achieve a consistent (i.e., stable) prediction model with 100% average accuracy when the number of quantization level M is greater than or equal to 6 and the resolution r is between 0.8σ and 1.2σ.

Generally, when the resolution r is smaller than or equal to 0.6σ, the average accuracy for every subject fluctuates and goes down for most of the scenarios. One possible reason for this phenomena is the overfitting issue that may happen when the resolution gets very small and the number of levels increases. On the other hand, when the resolution r gets bigger (i.e., r ≥ 1.4σ which means α ≥ 1.4), the two VWM-based matrices namely PVWM and SVWM will be sparse. As a result, the generated features, under this choice of quantization parameters for both classes, will not be significantly different from each other. Consequently, the average accuracy of the classifier decreases as the resolution increases regardless of the number of intervals M as shown in Table 4.

V. FURTHER ANALYSIS OF THE VWM-BASED FEATURES FOR EPILEPTIC SEIZURE PREDICTION

In order to illustrate the robustness of the proposed quantization scheme of Algorithm 1 along with the proposed VWM-based features, we consider in this section another type of classification problem (i.e., different size and orientation of the dataset). Specifically, we utilize the publicly available data published in [33] to predict the occurrence of epileptic seizure. There are five sets (A,B,C,D and E), each containing 100 single-channel EEG segments, in the complete data set. Sets A and B contain segments extracted from surface EEG recordings of five healthy volunteers in an awake state with eyes open (A) and eyes closed (B). The remaining three sets (sets C, D, and E) are taken from EEG archive of presurgical diagnosis from five different patients. Sets C and D contained only seizure-free activity, while set E was the only one that contained seizure activity. In this paper, the full set of the datasets (i.e., A,B,C, D and E) were used, like in [11, 12]. Unlike the start/plus fMRI dataset, the dataset for the epilepsy prediction problem has large number of samples (the actual size of the full dataset is 11,501) to train and test for deriving the best prediction problem. Therefore, 10-fold cross-validation was used to extract the best prediction model using three classification techniques (LR, SMLR and NB). Since the given dataset is unbalanced where we have 2,300 positive samples (patients who have epileptic seizure) and 9,201 negative samples, we randomly choose 2,300 negative samples out of the 9,201 samples to balance the size of the dataset provided to the classifier. The number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) are utilized to assess the performance.

### Table 3: Comparison of the results obtained by the LR classifier trained on the proposed VWM-based features and the state-of-art methods [23], [32] with dataset collected from 7 ROIs [CALC, "DLPC", "LIPL", "LIPS", "LOPER", "LT" and "LTRIA"]

<table>
<thead>
<tr>
<th>Subject</th>
<th>Method1 (23)</th>
<th>Method2 (32)</th>
<th>VWM method [M = 6, r = σ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature</td>
<td>Classifier</td>
<td>Feature size</td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td>size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4820</td>
<td>NB</td>
<td>15630</td>
<td>88.75</td>
</tr>
<tr>
<td>4847</td>
<td>NB</td>
<td>30878</td>
<td>100</td>
</tr>
<tr>
<td>4799</td>
<td>NB</td>
<td>25322</td>
<td>93.75</td>
</tr>
<tr>
<td>5675</td>
<td>NB</td>
<td>32121</td>
<td>93.75</td>
</tr>
<tr>
<td>5680</td>
<td>NB</td>
<td>23456</td>
<td>91.25</td>
</tr>
<tr>
<td>5710</td>
<td>NB</td>
<td>16544</td>
<td>93.75</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>23992</td>
<td>93.54</td>
</tr>
</tbody>
</table>
of a classifier. Sensitivity is the percentage of people with disease, who have a positive test result and can be defined as follows:

\[ Sensitivity = \frac{TP}{TP + FN} \times 100\% \]  

(8)

where specificity is the percentage of people who do not have the epileptic seizure who have a negative test result and defined as:

\[ Specificity = \frac{TN}{TN + FP} \times 100\% \]  

(9)

Table 5 shows the performance results (sensitivity, specificity and precision) using the three classifiers for a resolution factor of \( \sigma \) and 6 quantization levels. The highest accuracy achieved using the VWM-based features along with the LR model is 95\%. However, the best performing prediction model reported with regard to this epileptic seizure dataset achieved an accuracy of 97.8\% [34], but the dimension of the feature vector was large with respect to the two-dimensional feature vector of the VWM. Unlike the fMRI dataset, the proposed VWM-based feature with the three classifiers did not achieve high accuracy with the respect to the best performing prediction model for the epileptic seizure dataset. Nevertheless, achieving accuracy of 95\% demonstrates that the proposed features are well-suited for various types of classification problems based on different size and orientation of data.

### VI. CONCLUSIONS

In this work, we developed a novel feature generation methodology that can improve the performance of cognitive state prediction models. We also present a quantification scheme required for the implementation strategy of the proposed feature generation. The novelty of this feature generation method is its ability to reduce the high dimensionality associated with the original feature vector into a two-dimensional feature vector while attaining significant prediction accuracy. To demonstrate that, we compare the performance of the best resulted classifier out of (LR, SVM and NB) under the VWM-based feature with the two optimal algorithms reported in the literature. Finally, a sensitivity analysis of the quantization parameters was performed to help choose the optimal quantization parameters for similar classification problems.

### REFERENCES


**TABLE 4.** The sensitivity analysis of the average accuracy \( (\text{Acc}_{\text{avg}}) \) w.r.t the quantization Levels \( M \) of and the resolution factor \( r = \alpha \sigma \)

<table>
<thead>
<tr>
<th>( M )</th>
<th>( 0.2\sigma )</th>
<th>( 0.4\sigma )</th>
<th>( 0.6\sigma )</th>
<th>( 0.8\sigma )</th>
<th>( \sigma )</th>
<th>( 1.2\sigma )</th>
<th>( 1.4\sigma )</th>
<th>( 1.6\sigma )</th>
<th>( 1.8\sigma )</th>
<th>( 2\sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.91</td>
<td>0.92</td>
<td>0.84</td>
<td>0.84</td>
<td>1</td>
<td>0.98</td>
<td>0.92</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>6</td>
<td>0.90</td>
<td>0.6625</td>
<td>0.67</td>
<td>1</td>
<td>0.99</td>
<td>1</td>
<td>0.92</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>8</td>
<td>0.75</td>
<td>0.59</td>
<td>0.69</td>
<td>0.99</td>
<td>1</td>
<td>1</td>
<td>0.92</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>10</td>
<td>0.78</td>
<td>0.42</td>
<td>0.70</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.92</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>12</td>
<td>0.56</td>
<td>0.29</td>
<td>0.76</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.92</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>14</td>
<td>0.66</td>
<td>0.35</td>
<td>0.75</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.92</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
</tr>
</tbody>
</table>

**TABLE 5.** The 10-folds cross-validation results using the data published in [33] \( [M = 6, r = \sigma] \)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
<th>Gmean</th>
<th>F1score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>95</td>
<td>93.57</td>
<td>96.35</td>
<td>96.35</td>
<td>94.93</td>
<td>94.90</td>
</tr>
<tr>
<td>MLR</td>
<td>92.24</td>
<td>85.22</td>
<td>99.26</td>
<td>99.16</td>
<td>91.95</td>
<td>91.62</td>
</tr>
<tr>
<td>NB</td>
<td>93.04</td>
<td>96.52</td>
<td>89.57</td>
<td>90.28</td>
<td>92.97</td>
<td>93.29</td>
</tr>
</tbody>
</table>


[35] SULTAN ALSHEHRI received the B.Sc. with honors in Electrical Engineering from Sultan Moulay Slimane University. He then got an M.Sc. degree in Electrical Engineering from the National School of Applied Sciences in Morocco before obtaining another M.Sc. degree in Embedded Systems from Lorraine University in France. Actually, he is a Ph.D. candidate in Electrical engineering department within CEMSE division in King Abdullah University of Science and Technology (KAUST). His current research interest is signal processing-based features generation for biomedical signal classification and prediction and the challenges of real implementation of artificial intelligence model.

[36] FAHAD ALBALAWI Fahad Albalawi received his B.S. in 2008 in Electrical Engineering from Umm Alqura University, Makka. He received his M.S. degree in 2013 in Electrical Engineering from George Washington University, Washington DC. In June 2017, Fahad obtained his Ph.D. degree in Electrical Engineering from University of California, Los Angeles. Right after his PhD graduation, Fahad obtained an associate professor position of Electrical Engineering at Taif University, Saudi Arabia. His research interests include model predictive control, nonlinear systems, dynamic process optimization, process operational safety and machine learning.

[37] ABDERRAZAK CHAHID received the B.Sc. in Electrical Engineering and Power Electronics from Sultan Moulay Slimane University. He then got an M.Sc. degree in Electrical Engineering from the National School of Applied Sciences in Morocco before obtaining another M.Sc. degree in Embedded Systems from Lorraine University in France. Actually, he is a Ph.D. candidate in Electrical engineering department within CEMSE division in King Abdullah University of Science and Technology (KAUST). His current research interest is signal processing-based features generation for biomedical signal classification and prediction and the challenges of real implementation of artificial intelligence model.

[38] TAOUS-MERIEM LALEG-KIRATI is an Associate Professor in the division of Computer, Electrical and Mathematical Sciences and Engineering at KAUST. She joined KAUST in December 2010. From 2009 to 2010, she was working as a permanent research scientist at the French Institute for research in Computer Sciences and Control Systems (INRIA) in Bordeaux. She earned her Ph.D. degree in applied mathematics from INRIA Paris, in 2008. She holds a master in control systems and signal processing from University Paris 11 in France. Her research interest includes modeling, estimation, and control of complex systems and signal/image analysis.