j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 7, special issue 2024

# Using Mathematical Techniques to Analyse Biomedical Data: A K-complexes EEG Signal Classification Study

Hanan Al-Hadeethi[1*], Ikhlas Al-Hadeethi[2]

[1]*College of Health and Medical Technologies, Uruk University, Baghdad, Iraq.*

[2] *College of Media/ Iraqia University, Baghdad, Iraq.*

*hanan.ali.hammoodi@uruk.edu.iq*

**Abstract** This paper endeavored to characterize the design, elaboration, and investigation of the execution of the K-complexes classification method in Electroencephalogram (EEG) signals. To solve many-aims optimization issues for the high dimensionality of every database, a mechanism for feature extraction that depends on merging the Discrete Fourier Transform (Discrete-FT) with Covariance Matrix (Cov-matrix) has been suggested. An EEG signal was split into comparatively little intervals and segments as the first step of the model design. For every EEG segment, Discrete-FT was applied. The Cov-matrix were employed to figure out the most efficient input features to represent the EEG signal. As the input to diverse classifiers, for instance, K-means and the Naïve Bayes algorithm, the extracted features were used. The suggested procedure equips a high rate of accuracy, ~94% when the outcomes were compared with current studies. In conclusion, results exhibited that the submitted process can evolve the classification of K-complexes in EEG signals. Compared with other methods, the proposed method supplied the best outcomes. Furthermore, the presented method can have functional applications to assist physicians in classifying transient events in sleep stages more precisely than the current methods. The new procedure can be utilized for several medical data species, such as restless legs syndrome, epilepsy, Focal and Non-Focal, etc.

Crossref 10.36371/port.2024.special.19

**Keywords:** *Electroencephalogram, Classification, K-complexes, Features Extraction, K-means, Naïve Bayes Algorithm.*

## 1. INTRODUCTION

The sleep staging classification procedure is the most crucial function within the context of sleep studies. Over the previous slight decades, analyzing EEG recordings has evolved on a group of principles presented by the search undertaken by Rechtschaffen and Kales [37]. These principles authorize physicians to designate particular labels to particular time intervals applied to characterize diverse situations of sleep stages, for instance, wakefulness, stages 1-4, and rapid eye movements (REM), as elucidated in Fig. 1.

The American Academy of Sleep Medicine (AASM) [6] produced several amendments to the R&K guidelines (1968). In their commendation, the non-REM stages were diminished into three stages, namely, Stage 1, Stage 2 and Stage 3.

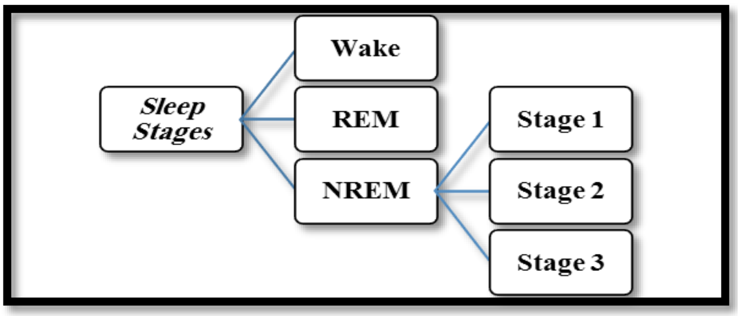


***Fig. 1.*** *Sleep Stages.*

Transient events, such as sleep spindles, K-complexes, micro-arousals, and various modalities demand an analysis to be implemented [6]. An EEG waveform that appears during Stage 2 of the non-REM stage is called a K-complex. It is deemed to be the most considerable incident in a healthy human EEG [10].

The K-complex was specified in 1937 in the particular laboratory presented by Alfred Lee Loomis [27].

The K-complexes have two major tasks: first, repress cortical arousal in response to an impulse that the sleeping brain considers is not due to signal danger; second, assisting sleep-



Hanan Al-Hadeethi, Ikhlas Al-Hadeethi, 2024. Using Mathematical Techniques to Analyse Biomedical Data: A K-complexes EEG Signal Classification Study. *Journal port Science Research*, 7(special), pp.208-218. https://doi.org/10.36371/port.2024.special.19

based memory consolidation [10]. Therefore, it has been proposed that K-complexes both protect sleep and evolve information processing. These tasks are a substantial part of the synchronization of the non-REM sleep stage. While they react to both the internal and the external stimuli in a reactive mode [8, 27], this would be consistent with a primal task in suppressing the cortical arousal in response to the stimuli that the brain chooses to initially procedure whether the stimulus is serious or not [10, 19]. Fig. 2 elucidates a K-complex waveform.

Furthermore, based on the current AASM definition [6], a K-complex is a well-determined negative severe wave, instantly pursued by a positive component standing out from the background EEGs, with an aggregate period overrunning 0.5s. Ordinarily, maximum amplitude when recording utilizing frontal derivations. Diverse studies have also presumed a maximum duration of ordinarily between one to three seconds [7, 13, 24, 25, 38]. Since they are among the most substantial aspects of Stage 2, the K-complexes are deemed to be one of the fundamental features that participate in the valuation of the several sleep stages. Nevertheless, the visual consistency of the K-complexes is time-consuming, relying on the information and experiences of clinicians due to they cannot be performed orderly [9]. This hardness is deemed to be a fundamental defiance of the automated K-complexes correspondence issues, for example, a shortage of convenient descriptions of the wave and its resemblance to various other EEG waves. Such example is the delta waves.

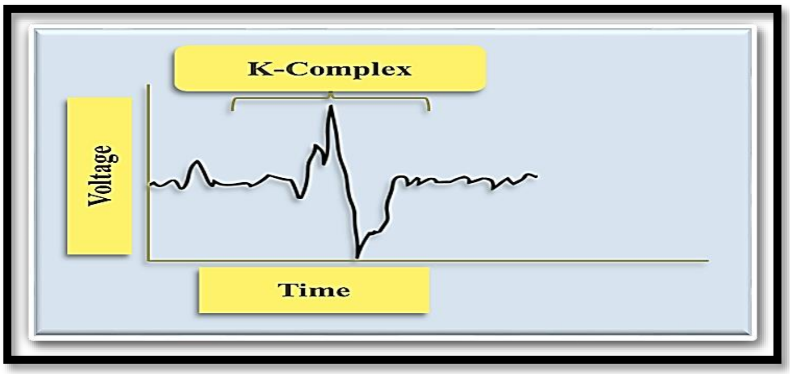


***Fig.2.*** *A K-complex waveform.*

## 2. LITERATURE REVIEW

Varied efforts have been made to boost the automatic identification of K-complexes in the literature where some studies have transacted with the detection of K-complexes [7, 14, 15, 24, 25, 40, 43, 46]. Some studies have utilized a complete night's recording, while others [4, 22, 23, 29, 30, 38, 40] while others have enforced the classification concept that utilizes EEG segments of a stable length. The first endeavor was to characterize the K-complex waveforms inclusive of the design of an electronic detection technique that was eligible to operate in real time [5]. However, the reliability proportion of the detection procedure could be argumentative for several reasons [7]. Later, Jansen et al. [23] presented a knowledge-based process for an automated sleep EEG analysis and the detection of sleep spindles and K-complexes. The results of Jansen et al. [23] were credible, on the other hand, the utilized datasets were not significant. In another model of exploratory research, Jansen BH [22] employed an artificial neural network (ANN) for the K-complexes detection. The ANN procedure was implemented employing both the filtered digitized data as well as the raw dataset, but it generated a non-conclusive outcome.

Detecting K-complexes and sleep spindles in sleep EEGs of humans applying a non-linear paradigm relies on two elements: transient, and oscillatory, which occurs in a low-frequency outcome. The results present that applying a non-linear model for detection is more efficient compared to using classic detection algorithms [34]. Da Rosa et al. [13] suggested a detector of K-complexes and vertex waves whose paradigms expose the transient events and neuronal feedback loops over a maximum-likelihood estimator. Therefore, the implementation of the K-complexes could offer a perfect comprehension of how the slow waves are activated by employing a sensory system as a portion of the continuum of reactive sleep slow waves. This is a major factor emphasizing how the brain may elaborate in response to a sensory input [18]. Furthermore, Bankman et al. [4] presented a feature-based detection process utilizing ANNs, which provided agreement with the recognition of visual K-complex. The mentioned study fulfilled a sensitivity and false positive (FP) proportion of 90% and 8%, respectively. As the aforementioned study stated, the information included in the features provided outcomes significantly more than the classification based on raw data. Subsequently, Jansen et al. [23] conducted a study aimed at evolving real and simulated EEG data with two basic ANN architectures [24]. Those ANNs received phase values and normalized dimensions over a Fourier transformation as inputs, gaining the preferable accuracy results with a low false positive proportion in the classification of K-complexes. The three


Hanan Al-Hadeethi, Ikhlas Al-Hadeethi, 2024. Using Mathematical Techniques to Analyse Biomedical Data: A K-complexes EEG Signal Classification Study. *Journal port Science Research*, 7(special), pp.208-2018. https://doi.org/10.36371/port.2024.special.19

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 7, special issue 2024

feature selection methods that were used were Wrappers, filters, and embedded methods. Also, other techniques were deemed, such as logistic regression and support vector machine [20].

The fundamental goal of the project was to improve a new technique with the most accurate outcomes for K-complexes classification. The first step was to split an EEG signal into segments with a period of about 1.0s, subsequently, each segment was conveyed into a feature extraction [12]. Statistical approaches, for example, sensitivity, specificity, F-measure, Matthews's correlation coefficient, Youden's J statistic, and accuracy were applied to assess the performance of the suggested technique, diverse and well-established. In the classification stage, two types of classifiers, including K-means and the Naïve Bayes were utilized to supply greater insights into the features determined from the EEG signal. The proposed methodology could provide results with high accuracy compared to other existing methods, thus it is substantial for practical applications of medical diagnostic areas.

## 3. EEG data

The database was acquired from the DREAMS project [14], and it is obtainable on the website [http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseKcomplexes/](http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseKcomplexes/). The recordings were obtained from six females and four males aged between 20 and 47. In this study, ten recordings were used, namely: excerpt 1 to excerpt 10. The recordings included two electrooculography (EOG) channels (P8-A1 and P18-A1), one submental electromyography (EMG) channel, and three EEG channels (CZ-A1 or C3-A1, FP1-A1, and O1-A1). At a frequency of 200 Hz, the data were sampled and for data storage, the standard European data format (EDF) was used. The mentioned data are characterized in Fig. 3.

**Subjects**

| Name | Age | Gender | Time of extract signal | No. of *K*-complexes by expert 1 |
|---|---|---|---|---|
| Excerpt 1 | 20 | Man | 02:00:00 to 02:30:00 | 34 |
| Excerpt 2 | 23 | Woman | 05:00:30 to 05:30:30 | 45 |
| Excerpt 3 | 47 | Woman | 06:06:00 to 06:36:00 | 12 |
| Excerpt 4 | 24 | Woman | 01:16:00 to 01:46:00 | 78 |
| Excerpt 5 | 23 | Woman | 07:15:00 to 07:45:00 | 39 |
| Excerpt 6 | 23 | Man | 00:10:00 to 00:40:00 | 28 |
| Excerpt 7 | 27 | Man | 02:20:00 to 02:50:00 | 11 |
| Excerpt 8 | 46 | Woman | 04:50:00 to 05:20:00 | 4 |
| Excerpt 9 | 27 | Man | 05:30:00 to 06:00:00 | 5 |
| Excerpt 10 | 21 | Woman | 02:30:00 to 03:00:00 | 16 |

***Fig. 3.*** *Data description.*

## 4. THE SUGGESTED METHODOLOGY

The outline of the proposed technique demonstrated in Fig. 4. In the beginning, using a sliding window procedure to partition EEG signals into small sub-segments. After a comprehensive experiment in the training phase, a window size of 1.0 seconds with an overlap of 0.8 s is set. Then, in feature extraction, each 1.0 s segment is decomposed into five frequency bands: {Delta, Theta, Alfa, Beta and, Gamma} utilizing the Discrete-FT algorithm. Later, five sets of features are extracted from each band employing the Cov-matrix, then used as an input to K-means, and Naïve Bayes classifiers. The suggested methodology is assessed in terms of the level of accuracy, sensitivity, specificity, Youden's J statistic, Matthew's correlation coefficient, and F-measures for the capability of these techniques to correctly detect and classify the K-complexes in sleep EEG signals. To fully explore the method, the performance of the proposed approach is also compared with other existing techniques using the same database. The outcomes display that the suggested approach could detect K-complexes in the EEG signal, and the method outperforms other current studies. Employing Matlab, the experiments were carried out (Version: R2022b).


Hanan Al-Hadeethi, Ikhlas Al-Hadeethi, 2024. Using Mathematical Techniques to Analyse Biomedical Data: A K-complexes EEG Signal Classification Study. *Journal port Science Research*, 7(special), pp.208-218. https://doi.org/10.36371/port.2024.special.19

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 7, special issue 2024

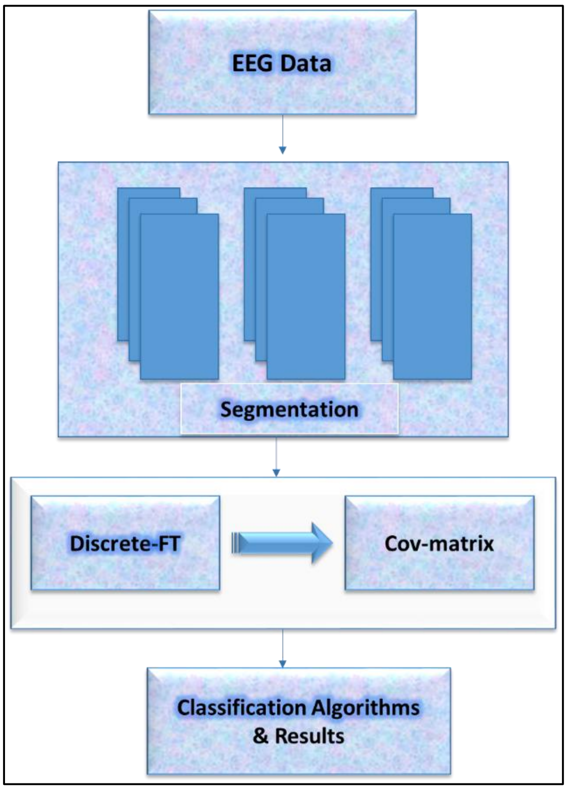


***Fig. 4***. *The suggested methodology.*

### 4.1. Segmentation

The detection of K-complexes in EEGs is the most crucial task in sleep staging. The frequency range of the *K*-complex waves is between 8 and 16 Hz. The optimal length of those waveforms is 0.5 s. To segment the EEG signals into their respective intervals, a sliding window technique was used. The length of window size is set to 1 second with an overlap of 0.8 s

### 4.2 Feature extraction

Feature extraction is employed to diminish the demanded resources to characterize an enormous set of high-dimensional data, this is one of the major issues stemming from the number of variables involved when implementing an analysis of complex data. Commonly, analysing a massive number of variables requires considerable computation capacity and a huge quantity of computer memory. Another possible problem is the classification algorithm could overfit the training samples and subsequently provide an inferior generalization of the new samples used as a test set [17]. The project combines Discrete-FT with discrete Cov-matrix in one model to design a procedure that captures relevant features of EEG signals.

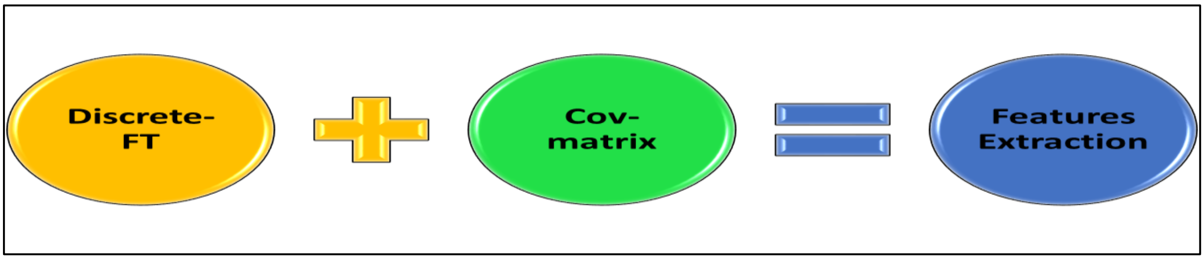


***Fig. 5***. *Feature Extraction Steps*

### 4.3 Discrete Fourier Transform (Discrete-FT)

The Discrete-FT is a data analysis technique applied to modify a specific species of sequence of a given function into other types of representation. Furthermore, this procedure is capable of transforming the structure of the cycle of a waveform into the sine components. The fundamental formula representing the Discrete-FT process is defined in Equation 1 [31, 11]:

$$T_k = \sum_{n=0}^{N-1} t_n e^{-i2\pi k\frac{n}{N}} \quad , \quad k = 0, 1, 2, \dots N-1 \; \dots \quad (1)$$

where $T_k$ is the Discrete-FT coefficients, $N$ is the total number of input EEG samples, and $n$ is the total number of points in Discrete-FT.



Hanan Al-Hadeethi, Ikhlas Al-Hadeethi, 2024. Using Mathematical Techniques to Analyse Biomedical Data: A K-complexes EEG Signal Classification Study. *Journal port Science Research*, 7(special), pp.208-2018. https://doi.org/10.36371/port.2024.special.19

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS
Journal port Science Research
Available online www.jport.co
Volume 7, special issue 2024

### 4.4 Covariance Matrix (Cov-matrix)

The covariance matrix is an essential term applied in the area of statistics and probability theory. The covariance matrix elucidates that there is an asymmetric array of numbers, it has diverse substantial properties [14]:

Let the covariance matrix of a random vector $A \in R^n$ and mean vector $m_A$ is defined as:

$$H_A = E[(A-m)(A-m)^T].$$

The elements $(i,j)^{th}$ of the covariance matrix $H_A$ is given by

$$H_{ij} = E\left[(A_i - m_i)(A_j - m_j)\right] = \sigma_{ij}.$$

The diagonal entries of $H_A$ are the vars of the components of the $A$ such as

$$H_{ii} = E[(A_i - m_i)^2] = \sigma_i^2.$$

$$H_{jj} = E\left[\left(A_j - m_j\right)^2\right] = \sigma_j^2.$$

The trace (tr) of H_A is positive because all diagonal entries are positive such as

$$tr(H_A) = \sum_{i=1}^{n} H_{ii} > 0.$$

The $H_A$ is symmetric, $H_A = H_A^T$ because $H_{ij} = \sigma_{ij} = \sigma_{ji} = H_{ji}$.

The $H_A$ is positive semidefinite, for all $b \in R^n$

$$E\{[(A-m)^T b]^2\} = E\{[(A-m)^T b]^T [(A-m)^T b]\} \geq 0$$

$$E[b^T (A-m)(A-m)^T b] \geq 0, b \in R^n$$

$$b^T H_A b \geq 0, b \in R^n$$

The H_A is symmetric, this is mean self-adjoint with the usual inner output, its eigenvalues are real and positive and the eigenvectors that belong to distinct eigenvalues are orthogonal, $H_A = \vee\wedge\vee^T = \sum_{i=1}^{n} \gamma_i \vec{v}_i \vec{v}_i^T$.

## 5. CLASSIFICATION ALGORITHMS

### 5.1 Naïve Bayes Algorithm

Frequently, Naïve Bayes is applied for classification and pattern recognition, it is a functional approach that relies on the implementations of the posterior hypothesis and principles of the Bayes theorem. The Naïve Bayes algorithm is a classification algorithm based on a series of conditional independence theories and Bayes rule [21, 28, 39]. To the aim of a learning process, indicated as f(Y/T) where T = (t1 ..., tni), the Naïve Bayes algorithm allocates a theory that each ti is conditionally independent of each of the other Tks given Y, also it is independent of each subset of the other Tk's, given Y. The value of the theory is facilitated by the representation of f(T/Y) and the problem of measuring it from the training data. For instance, the case where $T = (t_1, t_2)$, we can state that [28, 39],

$$f(T/Y) = P(t_1, t_2 /Y)$$

$$= f(t_1/t_2, Y)P(t_2/Y) \quad ... \quad (4)$$

$$= f(t_1/Y)P(t_2/Y) \quad ... \quad (5)$$

where $f(t_1/t_2, Y)P(t_2/Y)$ follows a common property of probabilities and $f(t_1/Y)P(t_2/Y)$ follows directly from the above explanation of conditional independence. In general, when $T$ contains $n$ attributes which satisfy the conditional independence hypothesis, we have

$$f(t_1 \ldots t_n/Y) = \prod_{i=1}^{n} f(t_i/Y) \qquad (6)$$

where $Y$ and ti are Boolean variables.

It is relevant to state that this stage demands just $2n$ parameters to define $f\ (T_i = t_{ik}/Y = y_j)$. This is a dramatic diminution compared to the $2(2n-1)$ parameters required to describe $f\ (T/Y)$ if we do not make a conditional independence assumption. To derive the algorithm, let us presume that $Y$ is any discrete-valued variable, and the characters $t_1, \ldots, t_n$ are any discrete or real-valued characters. The major aim here is to train a classifier that will produce the probability distribution over potential values of $Y$, for each new instance $T$ be classified. The expression for the probability that $Y$ will use on its kth possible value, according to the Bayes rule, is [28, 39]:

$$f(Y = y_k/t_1 \ldots t_n) = \frac{f(Y = y_k) f(t_i/Y = y_k)}{\sum_j f(Y = y_j)\ f(t_i/Y = y_j)} \qquad (7)$$

This sum is taken over all potential values $y_j$ of $Y$. For example, let us presume that $t_i$ are conditionally independent given $Y$, we can employ Equation (7) to rewrite this as (McCallum and Nigam 1998; Rish 2001)

$$f(Y = y_k/t_1 \ldots t_n) = \frac{f(Y = y_k) \prod_i f(t_i/Y = y_k)}{\sum_j f(Y = y_j)\ \prod_i f(t_i/Y = y_j)} \qquad (8)$$


Hanan Al-Hadeethi, Ikhlas Al-Hadeethi, 2024. Using Mathematical Techniques to Analyse Biomedical Data: A K-complexes EEG Signal Classification Study. *Journal port Science Research*, 7(special), pp.208-218. https://doi.org/10.36371/port.2024.special.19

Equation (7) is the substantial equation for the Naïve Bayes classifier. Given a new character $T^{new}$ = $(t_1 ...tn)$, this equation illustrates how to compute the probability that $Y$ will take on any given value, given the observed attribute values of $T^{new}$ and given the distributions $P(Y)$ and $P(ti/Y)$ measured from the training data. If we are interested only in the most possible value of $Y$, then we have the rule of Naïve Bayes classification as:

$$Y \leftarrow argmax_{y_k} \frac{f(Y = y_k) \prod_i f(t_i/Y = y_k)}{\sum_j f(Y = y_j) \prod_i f(t_i/Y = y_j)} \qquad (9)$$

which simplifies to the following Equation (denominator does not depend on $y_k$).

$$Y \leftarrow argmax_{y_k} f(Y = y_k) \prod_i f(t_i/Y = y_k) \qquad (10)$$

### 5.2 *K*-means

*K*-means is a procedure based on the vector quantization technique, originally derived from the signal processing area, which is usually used in data mining for cluster analysis. The purpose of the *K*-means method is to split $n$ observations into $k$ clusters in which each observation belongs to the cluster with the closest mean serving as a prototype of the cluster. A group of observations $(t_1, t_2, \ldots, t_n)$, could use where each observation is a $d$-dimensional real vector, and *K*-means clustering aims to split the $n$ observations into $k$ $(\leq n)$ sets $S = \{S_1, S_2, \ldots, S_k\}$ to diminish the within-cluster sum of squares (WCSS). The goal is to find [16]:

$$\underset{s}{argmin} \sum_{i=1}^{k} \sum_{t \in S_i} \|t - \mu_i\|^2 = \underset{s}{argmin} \sum_{i=1}^{k} |S_i| Var\, S_i \qquad (11)$$

where $\mu_i$ is the mean of points in $S_i$. This is equivalent to diminishing the pairwise squared deviations of points in the same cluster:

$$\underset{s}{argmin} \sum_{i=1}^{k} \frac{1}{2|S_i|} \sum_{t,y \in S_i} \|t - y\|^2 \qquad (12)$$

The equivalence could deduced by identifying: $\sum_{t \in S_i} \|t - \mu_i\|^2 = \sum_{t \neq y \in S_i} (t - \mu_i)(\mu_i - y)$. Because the total variance is constant, also, this is equivalent to diminishing the sum of squared deviations between points in different clusters [26].

## 6. EVALUATION OF MODEL PERFORMANCES

The assessment of the suggested techniques allows to conduct of a topical valuation of the outcomes of the methodology. Evaluation methods are applied as following:

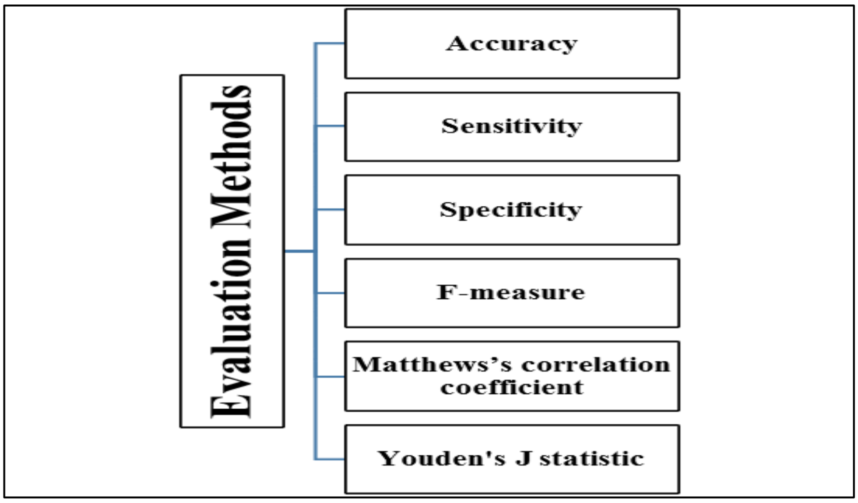


***Fig. 6**. The suggested Evaluation Methods*

### 6.1. Accuracy

Accuracy (Acc.) is the degree of vicinity of a calculated quantity to its actual (true) value. The accuracy defined in Equation (15) [3, 35]:

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS
Journal port Science Research
Available online www.jport.co
Volume 7, special issue 2024

Acc. = $(TP+TN)/(TP+TN+FP+FN)$ (15)

where true negative (TN) comprises the practical non-K-complexes that are correctly classified employing the proposed method as non-K-complexes. True positive (TP) denotes the practical K-complex waves that are correctly detected using the suggested procedure. False negative (FN) demonstrates the practical K-complex that is incorrectly marked as non-K-complexes. False positive (FP) refers to the number of K-complexes that are incorrectly defined by the proposed process.

### 6.2. Sensitivity (Sen.)

Sensitivity (Sen) is a statistical approach to the performance of a binary classification test applied to measure the rate of the actual positive predication, it is determined as follows [1, 41]:

$$\text{Sen.} = \frac{\text{The number of true positives}}{\text{The number of true positives+ the number of false negatives}} \quad \ldots. \quad (16)$$

$$\text{Sen} = TP/TP+FN \quad \ldots \quad (17)$$

### 6.3. Specificity (Spe)

Specificity (Spe) is employed to measure the proportion of the genuine negative predication, it is defined as follows [1, 41]:

$$\text{Spe.} = \frac{\text{The number of true negatives}}{\text{number of true negatives+number of false positives}} \quad .. \quad (18)$$

### 6.4. F-measure

F-measure is a substantial measurement that is applied to display the overlap between the collections of the true K-complexes and the extracted K-complexes utilizing the suggested procedure. It is defined in Equation (24) Sokolova [41]:

$$\text{F-measure} = 2 \times \frac{PPV \times recall}{PPV+recall} \quad (24)$$

where $ppv = TP/TP+FP$ ; (25)

recall = $TP/TP+FN$ (26)

### 6.5 Matthews correlation coefficient

To test the quality of binary classifications in machine learning, the Matthews correlation coefficient (MCC) is applied. The MCC is the correlation coefficient between predicted binary classifications and the observed. The value of the MCC is between -1 and +1. MCC = +1 denotes an optimal prediction and MCC =-1 represents the total disagreement between prediction and observation. The MCC is defined in Equation (27) Baldi [41]:

$$\text{MCC} = TP \times TN - FP \times FN/\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} \quad (27)$$

### 6.7 Youden's J statistic

Furthermore, the study applied Youden's J statistic as a single statistic that epitomizes the performance of a binary diagnostic test [50].

$$J = \frac{the\ number\ of\ true\ positives}{the\ number\ of\ true\ positives +\ the\ number\ of\ false\ negatives}$$

$$+\frac{the\ number\ of\ true\ negatives}{the\ number\ of\ true\ negatives + the\ number\ of\ false\ positives} - 1 \quad (29)$$

$$J = sensitivity+specificity-1 \quad (30)$$

## 7. EXPERIMENTAL RESULTS

The efficiency of the proposed method in detecting K-complexes is assessed. To determine all information from data without missing any part of EEG signals, the segmentation technique is employed. The EEG signal is divided into little segments. At the beginning, the interval of each segment is set at 1s. Then, the EEG signal is transformed into time-frequency data applying discrete FT with five bands (Alfa, Gama, Beta, Delta, and Theta). After that, the Covariance Matrix (Cov-matrix) is used to diminish dimensionality based on feature extraction. The outcomes demonstrated that the extracted features through the Covariance Matrix (Cov-matrix) procedure were efficient for precise results. For the classification, various kinds of classifiers (K-means and Naïve Bayes) were applied. The experimental results illustrated that


Hanan Al-Hadeethi, Ikhlas Al-Hadeethi, 2024. Using Mathematical Techniques to Analyse Biomedical Data: A K-complexes EEG Signal Classification Study. *Journal port Science Research*, 7(special), pp.208-218. https://doi.org/10.36371/port.2024.special.19

the suggested procedure fulfilled high classification outcomes compared to the Naïve Bayes classifier. Many tests have been done and the best 10-folds were chosen then the average of the folds was calculated. The proposed technique was assessed employing diverse assessment procedures. The results mentioned that the suggested process could evolve the classification of K-complexes in sleep stages. Furthermore, the proposed process presented an optimal prediction based on the Matthews correlation coefficient. Fig. 4 displays the average outcomes of the 10-folds. After many trials, several proportions were applied between training data and testing data. The results demonstrated that the best rate of the training data and the testing data is 1:1.

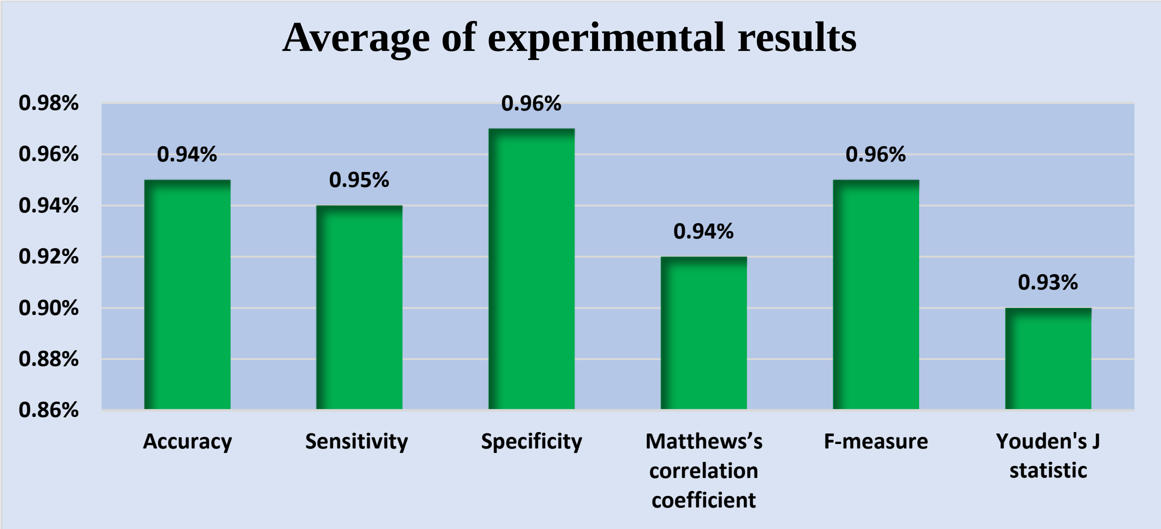


***Fig. 4*** *Average outcomes of the 10-folds*

### 7.1 Discussion and Comparison

Three types of comparison were built. First, the results of Discrete-FT combined with Covariance Matrix (Cov-matrix) and without were compared. Second, the performance of the presented technique was compared with current studies [14] applying diverse statistical evaluation processes of accuracy, sensitivity, and specificity with the two classifiers, including K-means and Naïve Bayes. The applied method achieved a sensitivity was 94%, while the sensitivity of recent researchers [4, 40] was 90%, and 86.47% respectively. Finally, the outcomes of Naïve Bayes and K-means have been compared. The suggested procedure with the Naïve Bayes classifier gained the highest F-measure value of 95%. Fig. 5 illustrates the outcomes of the comparisons and demonstrates that the applied method with Naïve Bayes acquired results better than other classifiers. The best accuracy was 95.35% by the Naïve Bayes. Moreover, the sensitivity and specificity with the same classifier were 94% and 97%, respectively. The second-highest results were recorded with the K-means classifier, the accuracy, sensitivity, and specificity were 82.5%, 70%, and 75%, respectively.

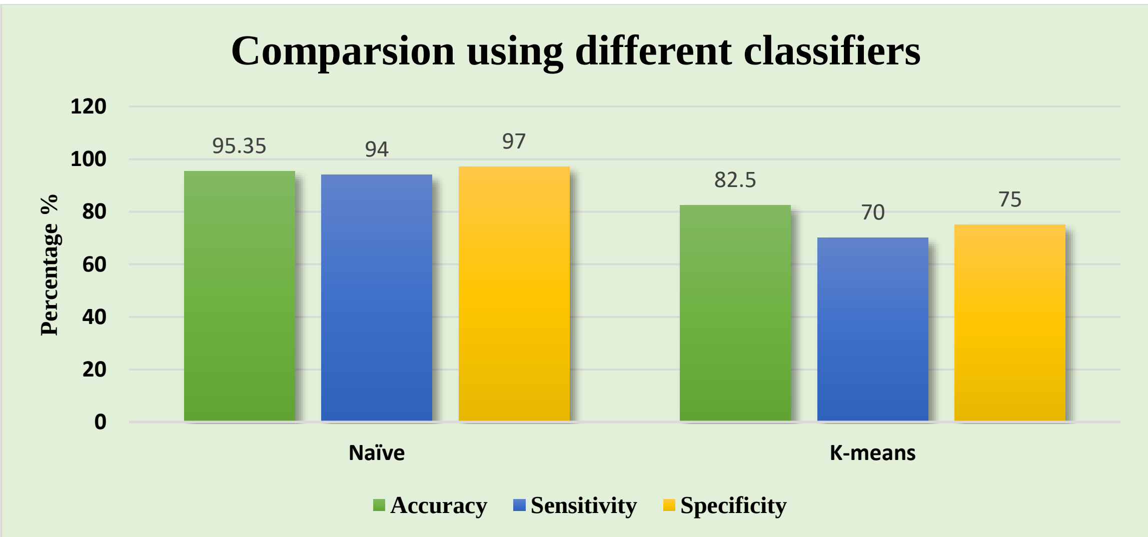


***Fig.5.*** *The outcomes of the comparisons of different classifiers, including the K-means, and Naïve Bayes*



Hanan Al-Hadeethi, Ikhlas Al-Hadeethi, 2024. Using Mathematical Techniques to Analyse Biomedical Data: A K-complexes EEG Signal Classification Study. *Journal port Science Research*, 7(special), pp.208-2018. https://doi.org/10.36371/port.2024.special.19

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS
Journal port Science Research
Available online www.jport.co
Volume 7, special issue 2024

## 8. CONCLUSION

A sufficient method of sleep stages classification was suggested. The fundamental objectives were to define a classifier that achieved high-accuracy outcomes. First, an EEG signal was divided into segments of 1s utilizing a sliding window. Then, each segment was transformed into time-frequency data using a Discrete-FT. After that, a Covariance Matrix (Cov-matrix) difference plot was employed to extract features. Later, several types of classifiers (K-means and Naïve Bayes) were applied. The Naïve Bayes classifier presented higher classification outcomes. The results of accuracy, sensitivity, specificity, Youden's J statistic, Matthew's correlation coefficient, and F-measure were supplied. The outcomes displayed that the proposed procedure could evolve the classification of K-complexes in EEG signals. The comparison demonstrated that the suggested procedure gained the best performance in terms of classification accuracy. The proposed technique could assist physicians in accurately detecting K-complexes in sleep EEGs, furthermore, it can be employed for diverse medical data types and several application areas.

**Informed consent:**

Informed consent was obtained from all individual participants included in the study.

## REFERENCES

[1] Altman DG, Bland JM (1994) Diagnostic tests. 1: Sensitivity and specificity. BMJ: British Medical Journal 308 (6943):1552

[2] Altman DG, Bland JM (1994) Statistics Notes: Diagnostic tests 2: predictive values. Bmj 309 (6947):102

[3] Baldi P, Brunak S, Chauvin Y, Andersen CA, Nielsen H (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16 (5):412-424

[4] Bankman IN, Sigillito VG, Wise RA, Smith PL (1992) Feature-based detection of the K-complex wave in the human electroencephalogram using neural networks. IEEE transactions on biomedical engineering 39 (12):1305-1310

[5] Bastien C, Campbell K (1992) The evoked K-complex: all-or-none phenomenon? Sleep 15 (3):236-245

[6] Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus C, Vaughn B (2012) The AASM manual for the scoring of sleep and associated events. Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine

[7] Bremer G, Smith JR, Karacan I (1970) Automatic detection of the K-complex in sleep electroencephalograms. IEEE Transactions on Biomedical Engineering (4):314-323

[8] Cabrero-Canosa M, Hernandez-Pereira E, Moret-Bonillo V (2004) Intelligent diagnosis of sleep apnea syndrome. IEEE Engineering in medicine and biology Magazine 23 (2):72-81

[9] Carskadon MA, Dement WC (2005) Normal human sleep: an overview. Principles and practice of sleep medicine 4:13-23

[10] Cash SS, Halgren E, Dehghani N, Rossetti AO, Thesen T, Wang C, Devinsky O, Kuzniecky R, Doyle W, Madsen JR (2009) The human K-complex represents an isolated cortical down-state. Science 324 (5930):1084-1087

[11] Cochran WT, Cooley JW, Favin DL, Helms HD, Kaenel RA, Lang WW, Maling GC, Nelson DE, Rader CM, Welch PD (1967) What is the fast Fourier transform? Proceedings of the IEEE 55 (10):1664-1674

[12] Cohen ME, Hudson DL, Deedwania PC (1996) Applying continuous chaotic modeling to cardiac signal analysis. IEEE Engineering in Medicine and Biology Magazine 15 (5):97-102

[13] Da Rosa A, Kemp B, Paiva T, da Silva FL, Kamphuisen H (1991) A model-based detector of vertex waves and K complexes in sleep electroencephalogram. Electroencephalography and clinical Neurophysiology 78 (1):71-79

[14] Devuyst S, Dutoit T, Stenuit P, Kerkhofs M Automatic K-complexes detection in sleep EEG recordings using likelihood thresholds. In: Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, 2010. IEEE, pp 4658-4661


Hanan Al-Hadeethi, Ikhlas Al-Hadeethi, 2024. Using Mathematical Techniques to Analyse Biomedical Data: A K-complexes EEG Signal Classification Study. *Journal port Science Research*, 7(special), pp.208-218. https://doi.org/10.36371/port.2024.special.19

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 7, special issue 2024

[15] Erdamar A, Duman F, Yetkin S (2012) A wavelet and teager energy operator based method for automatic detection of K-Complex in sleep EEG. Expert Systems with Applications 39 (1):1284-1290

[16] Forgy EW (1965) Cluster analysis of multivariate data: efficiency versus interpretability of classifications. biometrics 21:768-769

[17] Friedman JH (1997) On bias, variance, 0/1—loss, and the curse-of-dimensionality. Data mining and knowledge discovery 1 (1):55-77

[18] Halász P (2016) The K-complex as a special reactive sleep slow wave–A theoretical update. Sleep medicine reviews 29:34-40

[19] Halász Pt (2005) K-complex, a reactive EEG graphoelement of NREM sleep: an old chap in a new garment. Sleep medicine reviews 9 (5):391-412

[20] Hernández-Pereira E, Bolón-Canedo V, Sánchez-Maroño N, Álvarez-Estévez D, Moret-Bonillo V, Alonso-Betanzos A (2016) A comparison of performance of K-complex classification methods using feature selection. Information Sciences 328:1-14

[21] Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70 (1-3):489-501

[22] Jansen BH (1990) Artificial neural nets for K-complex detection. IEEE Engineering in medicine and Biology Magazine 9 (3):50-52

[23] Jansen BH, Dawant BM (1989) Knowledge-based approach to sleep EEG analysis-a feasibility study. IEEE Transactions on Biomedical Engineering 36 (5):510-518

[24] Jansen BH, Desai PR (1994) K-complex detection using multi-layer perceptrons and recurrent networks. International Journal of Bio-medical computing 37 (3):249-257

[25] Kam A, Cohen A, Geva A, Tarasiuk A Detection of K-complexes in sleep EEG using CD-HMM. In: Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, 2004. IEEE, pp 33-36

[26] Kriegel H-P, Schubert E, Zimek A (2017) The (black) art of runtime evaluation: Are we comparing algorithms or implementations? Knowledge and Information Systems 52 (2):341-378

[27] Loomis AL, Harvey EN, Hobart GA (1937) Cerebral states during sleep, as studied by human brain potentials. Journal of experimental psychology 21 (2):127

[28] McCallum A, Nigam K A comparison of event models for naive bayes text classification. In: AAAI-98 workshop on learning for text categorization, 1998. vol 1. Citeseer, pp 41-48

[29] Moloney D, Sukhorukova N, Vamplew P, Ugon J, Li G, Beliakov G, Philippe C, Amiel He, Ugon A (2011) Detecting K-complexes for sleep stage identification using nonsmooth optimization. The ANZIAM Journal 52 (4):319-332

[30] Noori SMR, Hekmatmanesh A, Mikaeili M, Sadeghniiat-Haghighi K K-complex identification in sleep EEG using MELM-GRBF classifier. In: Biomedical Engineering (ICBME), 2014 21th Iranian Conference on, 2014. IEEE, pp 119-123

[31] Nussbaumer HJ (2012) Fast Fourier transform and convolution algorithms, vol 2. Springer Science & Business Media,

[32] Pachori RB, Hewson D, Snoussi H, Duchêne J Postural time-series analysis using empirical mode decomposition and second-order difference plots. In: Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, 2009. IEEE, pp 537-540

[33] Pachori RB, Patidar S (2014) Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions. Computer methods and programs in biomedicine 113 (2):494-502

[34] Parekh A, Selesnick IW, Rapoport DM, Ayappa I (2015) Detection of K-complexes and sleep spindles (DETOKS) using sparse optimization. Journal of neuroscience methods 251:37-46

[35] Polat K, Güneş S (2007) Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Applied Mathematics and Computation 187 (2):1017-1026

[36] Poulos M, Rangoussi M, Alexandris N, Evangelou A (2002) Person identification from the EEG using nonlinear signal classification. Methods of information in Medicine 41 (1):64-75

[37] 37. Rechtschaffen A (1968) A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects. Brain information service

[38] Richard C, Lengelle R (1998) Joint time and time-frequency optimal detection of K-complexes in sleep EEG. Computers and biomedical research 31 (3):209-229

[39] Rish I An empirical study of the naive Bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence, 2001. vol 22. IBM New York, pp 41-46

[40] Shete V, Elgandelwar S, Sonar S, Cliarantimath A, Mytri V (2012) Detection of K-complex in sleep EEG signal using support vector machine. International Journal of Scientific & Engineering Research 3 (1-6)

[41] Sokolova M, Japkowicz N, Szpakowicz S Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In: Australasian joint conference on artificial intelligence, 2006. Springer, pp 1015-1021

[42] Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Information Processing & Management 45 (4):427-437

[43] Strungaru C, Popescu M (1998) Neural network for sleep EEG K-complex detection. Biomedizinische Technik/Biomedical Engineering 43 (s3):113-116

[44] Suykens JA, De Brabanter J, Lukas L, Vandewalle J (2002) Weighted least squares support vector machines: robustness and sparse approximation. Neurocomputing 48 (1-4):85-105

[45] Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. Neural processing letters 9 (3):293-300

[46] Tang Z, Ishii N (1995) Detection of the K-complex using a new method of recognizing waveform based on the discrete wavelet transform. IEICE TRANSACTIONS on Information and Systems 78 (1):77-85

[47] Thuraisingham R (2009) A classification system to detect congestive heart failure using second-order difference plot of RR intervals. Cardiology research and practice 2009

[48] Thuraisingham RA, Tran Y, Boord P, Craig A (2007) Analysis of eyes open, eye closed EEG signals using second-order difference plot. Medical & biological engineering & computing 45 (12):1243-1249

[49] Vapnik V (1963) Pattern recognition using generalized portrait method. Automation and remote control 24:774-780

[50] Youden WJ (1950) Index for rating diagnostic tests. Cancer 3 (1):32-35

[51] Hussein, M. J. M., & Hamodi, H. A. (2017). Comparison count regression models for the number of infected of pneumonia. *Global Journal of Pure and Applied Mathematics*, *13*(9), 5359-5366.