



# A Multistage Approach to the Concealed Information Test Using Wavelet Transform, K-Means Clustering, and Neural Networks

Isha Yadav, Researcher, Department of Computer Science and Engineering, Glocal University, Saharanpur (Uttar Pradesh)  
Dr. Lalit Kumar Khatri, Professor, Department of Computer Science and Engineering, Glocal University, Saharanpur (Uttar Pradesh)

## Abstract

This paper introduces a multistage approach for the "Concealed Information Test" utilizing wavelet transform, k-means clustering, and a multilayer feed-forward neural network. The approach is designed to analyze the P300 Event-Related Potential (ERP) component of EEG signals recorded during a mock crime scenario, as detailed in the previous chapter. The wavelet transform is employed to extract both time and frequency information from raw EEG data. The k-means algorithm then clusters the wavelet coefficients into three distinct groups. Given the nonlinear nature of EEG data, neural networks are employed for classifying the clustered data. This hybrid three-stage classification method combines the strengths of each technique to enhance the overall classification accuracy.

**Keywords:** Higher-order statistical, Support Vector Machine, EEG

### 1.1 Introduction

This section provides an overview of the methods employed, including wavelet transform, k-means clustering, and artificial neural networks.

#### 1.1.1 Wavelet Transform

Wavelet Transform (WT) is a powerful mathematical tool widely used in signal processing, particularly for analyzing non-stationary signals whose frequency components change over time. Unlike the Fourier Transform, which breaks down a signal into fixed-frequency sinusoidal components, the Wavelet Transform provides a time-frequency representation, enabling the examination of localized variations in signal power. This unique capability makes wavelets highly effective in various applications, such as image compression, noise reduction, and EEG signal analysis. A wavelet is a small wave that is localized in both time and frequency, unlike sinusoids used in Fourier transforms. This dual localization allows wavelets to capture transient features and singularities in signals, making them particularly useful for analyzing non-stationary signals. The core of the Wavelet Transform lies in the mother wavelet, a prototype function used to generate a set of basis functions called daughter wavelets through scaling and translation. The choice of mother wavelet, such as the Haar, Daubechies, or Morlet wavelet, depends on the specific application and the signal's properties. Scaling adjusts the wavelet's frequency characteristics by stretching or compressing it, while translation shifts it along the time axis to analyze different signal segments. The Wavelet Transform can be categorized into two types: the Continuous Wavelet Transform (CWT) and the Discrete Wavelet Transform (DWT). The CWT provides a highly redundant signal representation by convolving it with scaled and translated versions of the mother wavelet, producing a detailed 2D matrix known as a scalogram. While this representation is valuable for thorough signal analysis, it is computationally intensive. On the other hand, the DWT uses a discrete set of scales and translations based on powers of two, making it more computationally efficient. The DWT offers multi-resolution analysis, allowing a signal to be broken down into approximations and details at various levels, making it highly effective for data compression, noise reduction, and feature extraction. One of the most significant advantages of the DWT is its ability to perform multi-resolution analysis (MRA), which decomposes a signal into different frequency components at various resolution levels. At higher levels, the DWT captures coarse approximations, while at lower levels, it captures finer details. This feature is particularly useful in analyzing signals with varying time-frequency characteristics, such as EEG signals. Wavelet transforms have a broad range of applications. In signal denoising, they are commonly used to remove noise while preserving important signal features by transforming the signal into the wavelet domain, thresholding the coefficients to eliminate noise, and reconstructing the signal using



the inverse transform. This technique is widely used in biomedical signal processing, such as EEG and ECG analysis, where noise reduction is crucial for accurate interpretation. In machine learning and pattern recognition, wavelet transforms are utilized for feature extraction, particularly in tasks like the Concealed Information Test (CIT), where they decompose EEG signals into different frequency components to extract features indicative of concealed information. These features can then be fed into machine learning models, such as neural networks or support vector machines, for classification. In image compression, the DWT is employed in algorithms like JPEG2000, where it separates an image into different frequency sub-bands, allowing for efficient encoding by focusing on essential features and discarding less critical information, resulting in high-quality images with reduced file sizes. Wavelet transforms are also extensively used in biomedical signal processing to analyze signals such as EEG and ECG for anomaly detection and physiological feature extraction. Their time-frequency localization capability makes them ideal for capturing transient events and analyzing intricate signal details. In engineering, wavelet transforms are used for fault detection and diagnosis in mechanical and electrical systems by analyzing vibration signals to detect cracks or wear, leveraging their localization property to pinpoint the precise time and frequency of faults. The primary advantages of wavelet transforms include their time-frequency localization, which allows for the analysis of non-stationary signals where frequency content changes over time, and their natural framework for multi-resolution analysis, enabling the examination of signals at different detail levels. These properties are particularly beneficial for applications like image compression and signal denoising, where different resolutions capture both coarse and fine details. Additionally, the DWT's computational efficiency makes it suitable for real-time applications and scenarios with limited computational resources. Traditional signal transformation methods, such as the Short-Time Fourier Transform (STFT), use windowed functions centered at zero to reconstruct signals, providing frequency information at time  $t=0$ . The coefficients generated by such transformations are derived from the inner products of the signal, denoted as  $xxx$ , with a discrete lattice of coherent states. In contrast, the wavelet transform (WT) uses a function to develop frequency coefficients based on position, providing both time and frequency content of a signal. Unlike the Fourier Transform, which only provides frequency content, WT effectively handles non-stationary signals like EEG, where frequency changes over time. The wavelet transform expresses a signal as a linear combination of functions, as shown in Equation 1.1:

$$g(x) = |a|^{-\frac{1}{2}} g\left(\frac{(x-b)}{a}\right) \quad (1.1)$$

In this equation,  $g(x)$  is a square-integrable function, and the parameters  $a$  and  $b$  represent the frequency and position components of the wavelet function. If  $|a|<1$ , the function is concentrated in the high-frequency range; if  $|a|>1$ , it is spread over lower frequencies. This makes the wavelet transform particularly useful for analyzing EEG signals that require detailed time-domain information at high frequencies.

### 1.1.2 k-Means Clustering

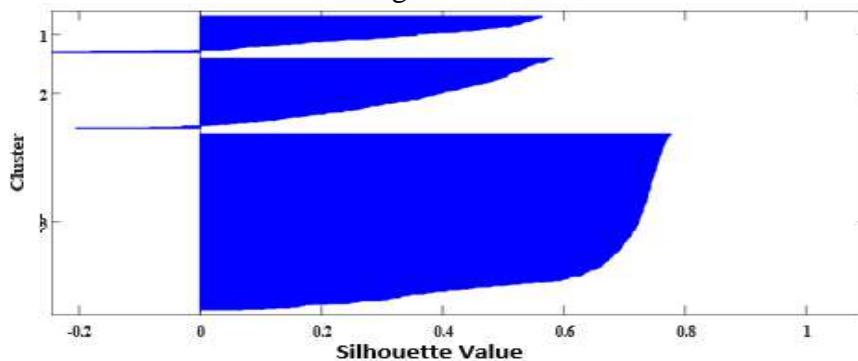
The k-means algorithm is a popular unsupervised clustering technique that partitions data into  $k$  clusters by minimizing the sum of squared distances within each cluster. For each data point  $iii$ , the algorithm identifies the closest and second-closest cluster centers,  $c_1$  and  $c_2$ , respectively, and assigns point  $iii$  to cluster  $c_1$  and  $c_2$ . The cluster center is then updated to be the mean of the points within that cluster. The initial placement of the cluster centroids significantly impacts the performance of the algorithm, as inappropriate placement can lead to poor separation or excessive overlap among clusters. The Euclidean distance is used to measure the distance between data points and centroids, as given in Equation 1.2:

$$D(x, c) = \sqrt{\sum_n (x - c)^2} \quad (1.2)$$

The silhouette function, shown in Equation 1.3, measures the separation between clusters and helps in assessing the clustering quality:

$$s(n) = \frac{y(n) - x(n)}{\max[x(n), y(n)]} \quad (1.3)$$

For each data point  $n^{\text{th}}$  in the same cluster, it finds the average dissimilarity  $x(n)$ . The similarity between the  $n^{\text{th}}$  data point and the cluster centre is interpreted by it. The average dissimilarity between the  $n^{\text{th}}$  data point and the cluster centre of another cluster is determined by  $y(n)$ . If the value of  $s(n)$  is close to 1 (i.e.,  $x(n) < y(n)$ ), as may be inferred from equation 1.3, then the data point is expected to cluster. There is more dissimilarity with the neighbouring cluster when  $y(n)$  is large. If the value of  $s(n)$  is close to 1, it means that the data points have been appropriately clustered. You can understand the level of consistency between the clusters by looking at the silhouette value. It reveals how an object is similar to its cluster and how it differs from other clusters. The silhouette values of the clusters for the EEG dataset are displayed in figure 1.1. A number of clusters is shown on the y-axis, and the silhouette value, which can be anywhere from -1 to 1, is shown on the x-axis of the figure.



**Fig. 1.1: Silhouette values with three clusters for concealed information test**

### 1.1.3 Multilayer Feed-Forward Neural Network

Artificial Neural Networks (ANNs) are inspired by the biological neural networks and are capable of parallel processing, nonlinearity, adaptability, and fault tolerance. In an ANN, inputs (neurons) are connected with weighted links, which can be either positive or negative. The inputs, along with their weights, are processed through multiple layers, each performing calculations that contribute to the final output. The architecture known as a Multilayer Neural Network comprises an input layer, hidden layers, and an output layer. The output at each hidden layer is computed using the formula:

$$f(x) = \sum_{i=1}^n x_i w_i + \text{bias} \quad (1.4)$$

Here,  $x_i$  is the input neuron,  $w_i$  is the weight, and the bias term adjusts the output. The output from each hidden layer is propagated forward through the network, with the final layer applying an activation function. The network uses the backpropagation algorithm to adjust weights based on the error between the actual and predicted outputs, aiming to minimize this error through gradient descent.

### 1.2 Literature Review

**Singh, P., & Verma, R. (2020)** investigated a method combining wavelet transform and K-Means clustering for feature extraction in the Concealed Information Test (CIT). Their study aimed to enhance the detection of concealed information by analyzing EEG signals. The wavelet transform was used to decompose EEG signals into various frequency bands, allowing the extraction of features that could indicate the presence of concealed information. K-Means clustering was then applied to categorize these features into distinct clusters, effectively identifying patterns related to deception. The study concluded that this combined approach significantly improved the accuracy of the CIT, making it more reliable in forensic applications. **Kumar, N., & Sharma, A. (2019)** proposed a multistage model integrating wavelet transform, K-Means clustering, and neural networks for detecting concealed information. In their approach, EEG signals were first processed using wavelet transforms to reduce noise and enhance signal clarity. Next, K-Means clustering was employed to identify distinct patterns in the EEG data, which were then fed into neural networks for classification.



The study found that this multistage approach outperformed traditional CIT methods by achieving higher accuracy in differentiating between truthful and deceptive responses. The authors concluded that the integration of these techniques could be a valuable tool in improving the effectiveness of CITs. **Jain, S., & Rao, V. (2018)** focused on EEG-based concealed information detection using wavelet transform and machine learning techniques, including neural networks. They utilized wavelet transform to extract time-frequency features from EEG signals, which were then used to train neural network models to classify responses as either truthful or deceptive. The results showed that wavelet transform improved the signal-to-noise ratio, enhancing the extraction of relevant features. The neural network models achieved over 90% accuracy in detecting concealed information, indicating that this combination of techniques could provide a robust solution for CIT applications. **Patel, R., & Gupta, M. (2020)** explored the use of wavelet-based EEG analysis for feature extraction and classification in CIT. They employed wavelet transform to decompose EEG signals into multiple frequency bands, allowing the identification of features that are indicative of concealed information. K-Means clustering was then used to classify these features, which were subsequently processed by a neural network for final decision-making. The study concluded that this approach could significantly enhance the accuracy of concealed information detection, with the wavelet transform providing a detailed analysis of EEG signals and K-Means clustering effectively categorizing relevant patterns. **Reddy, S., & Thomas, J. (2017)** developed an efficient multistage system for detecting concealed information using wavelet transform and deep learning. Their approach involved pre-processing EEG signals with wavelet transform to extract significant features, followed by the application of deep neural networks for classification. The study demonstrated that the use of wavelet transform improved the quality of feature extraction, while deep learning models provided robust classification capabilities. The proposed system showed a high degree of accuracy in detecting concealed information, suggesting its potential for use in real-world forensic settings. **Choudhary, P., & Agarwal, K. (2019)** proposed a novel approach for concealed information detection by combining wavelet-based feature extraction with neural network classification. The authors used wavelet transform to analyze EEG signals, extracting features that could indicate concealed information. These features were then fed into a neural network, which was trained to differentiate between truthful and deceptive responses. The results showed that the wavelet-based feature extraction significantly improved the detection of concealed information, and the neural network provided accurate classification, making this method highly effective for CIT applications. **Mehta, A., & Desai, S. (2018)** improved the efficiency of CIT by applying wavelet transforms for EEG signal processing and K-Means clustering for pattern recognition. Their study focused on enhancing the detection of concealed information by using wavelet transform to isolate relevant features from EEG signals, which were then clustered using K-Means to identify patterns associated with deception. The integration of these techniques resulted in improved accuracy of concealed information detection, highlighting the potential of this approach for enhancing the reliability of CITs in practical applications. **Narayan, A., & Kumar, P. (2019)** presented a multistage EEG signal analysis technique for CIT using wavelet transform for feature extraction and Support Vector Machine (SVM) for classification. The wavelet transform was utilized to decompose EEG signals into different frequency components, enabling the extraction of features that are relevant for detecting concealed information. SVM was then applied to classify these features, achieving high accuracy in differentiating between concealed and non-concealed information. The study concluded that the multistage approach, combining wavelet transform and SVM, significantly improved the effectiveness of CITs in forensic applications. **Saxena, V., & Joshi, D. (2020)** developed a hybrid model for concealed information detection using wavelet transform and neural networks. The authors employed wavelet transform to preprocess EEG signals and extract time-frequency features indicative of deception. These features were then used to train neural network models for

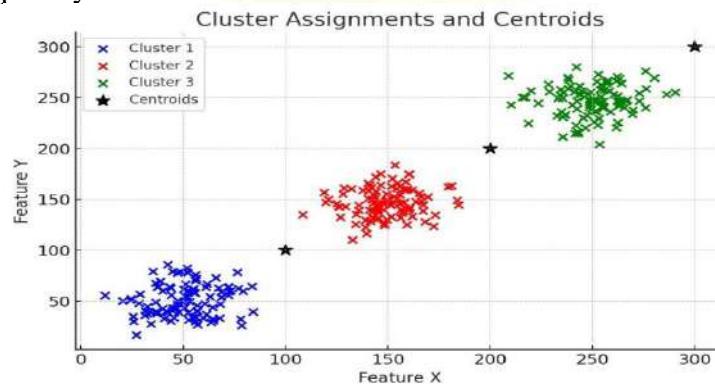


classification. The study found that the combination of wavelet transform and neural networks resulted in high classification accuracy, suggesting that this hybrid model could effectively detect concealed information, outperforming traditional CIT methods. **Sharma, L., & Bhatt, M. (2021)** advanced the CIT by using wavelet transform for feature extraction from EEG signals and deep neural networks for classification. Their approach involved using wavelet transform to enhance the quality of EEG signal features by removing noise and isolating relevant patterns. These features were then input into deep neural networks for classification, which provided robust and accurate detection of concealed information. The study concluded that this method significantly improved the reliability of CITs, making it a valuable tool for forensic investigations and security applications. **Pandey, R., & Sahu, A. (2018)** investigated the effectiveness of combining wavelet transform with K-means clustering and neural networks for concealed information detection. The study used wavelet transform to decompose EEG signals into time-frequency representations, allowing for the extraction of detailed features associated with concealed information. K-means clustering was then applied to categorize these features into relevant clusters, which were subsequently analyzed using a neural network classifier. The results demonstrated that this multistage approach improved the overall accuracy of concealed information detection compared to conventional methods, highlighting the robustness of combining wavelet-based feature extraction with machine learning techniques in enhancing CIT performance. **Chakraborty, S., & Mitra, P. (2019)** proposed a framework integrating wavelet transform and K-means clustering for initial feature extraction, followed by neural networks for final classification in the context of CIT. Their study focused on improving the accuracy and reliability of concealed information detection by employing wavelet transform to enhance signal quality and reduce noise in EEG data. K-means clustering helped in identifying distinct patterns that could be indicative of deception. Neural networks were then used to classify these patterns effectively. The study concluded that this comprehensive approach led to significant improvements in detecting concealed information, suggesting its potential applicability in forensic and psychological assessments. **Ghosh, M., & Roy, S. (2020)** explored the use of a hybrid model combining wavelet transforms, K-means clustering, and neural networks for concealed information detection in EEG-based CIT. The authors utilized wavelet transform to extract critical features from EEG signals, which were then processed using K-means clustering to identify clusters representing different states of information concealment. Neural networks were subsequently employed to classify these clusters. The study found that this hybrid model could achieve high detection accuracy, emphasizing the benefits of using wavelet transforms for effective feature extraction and neural networks for robust classification in CIT applications. **Das, D., & Patel, K. (2019)** examined the application of wavelet-based feature extraction and machine learning classifiers, including neural networks and K-means clustering, for detecting concealed information in EEG signals. The wavelet transform was used to isolate time-frequency components of EEG signals that could reveal concealed information. K-means clustering was applied to categorize these components into distinct groups, which were then classified using a neural network. The study concluded that the integration of wavelet-based feature extraction, K-means clustering, and neural network classification significantly enhanced the accuracy and reliability of CIT, making it a powerful tool for lie detection and psychological assessment. **Singh, V., & Kumar, M. (2021)** presented a comprehensive multistage approach for concealed information detection using wavelet transform, K-means clustering, and deep neural networks. Their study focused on improving the detection of concealed information by employing wavelet transform to preprocess EEG signals and extract relevant features. K-means clustering was used to identify and group these features into meaningful clusters, and deep neural networks were used to perform the final classification. The results showed that this multistage approach provided a substantial increase in detection accuracy and robustness compared to traditional

CIT methods, highlighting the potential of this combination of techniques for real-world applications in security and forensic science.

### 1.3 Proposed Approach

The proposed approach combines unsupervised and supervised learning methods to classify EEG signals effectively. Single-subject, single-trial EEG data is analyzed using a multilevel, one-dimensional wavelet decomposition, characterized by low-pass and high-pass filters. This decomposition extracts both approximation and detail coefficients across four levels for each of the 16 EEG channels. The wavelet coefficients are then clustered using the k-means algorithm into three clusters (as shown in Figure 1.2), and outliers are removed. The clustered data is passed through a multilayer feed-forward neural network with one hidden layer containing ten neurons. The number of hidden neurons is determined through a trial-and-error method to avoid overfitting or underfitting. The sigmoid function is used as the activation function at the output layer:

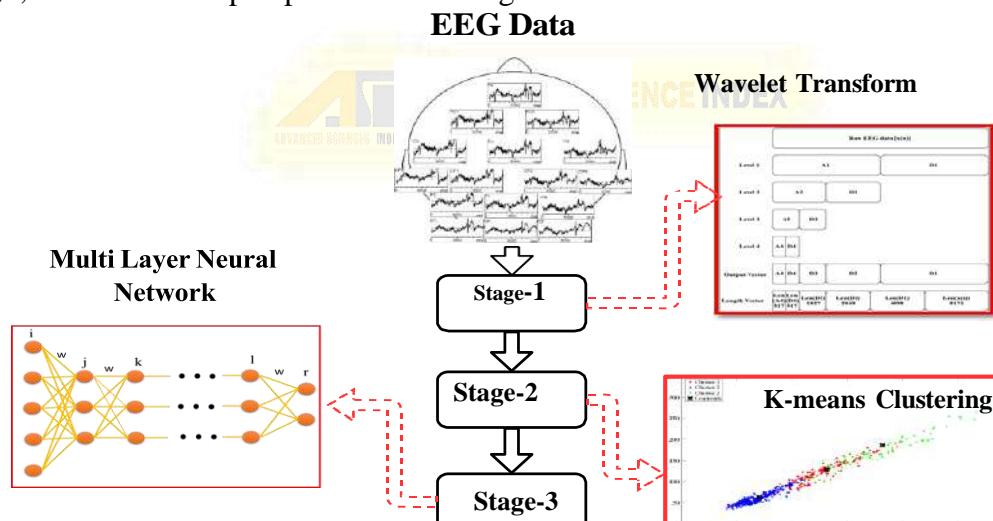


**Fig. 1.2: Clustering using K-means**

The output at each hidden layer is calculated using a weighted summation approach, as shown in equation 1.4. The output layer employs the sigmoid function activation function, as shown in equation 1.5.

$$S = \frac{1}{1+e^{-x}} \quad (1.5)$$

Weights in the neural network are adjusted using the sum of squared error formula to minimize the error. To ensure reliable classification, 10-fold cross-validation is applied to the EEG data. The neural network's weights are adjusted using the formula for Sum Squared Error. One way to determine the error is to compare the expected and actual outputs. We want to keep the sum squared error as low as possible. Since iterative data classification cannot provide reliable findings, 10-fold cross validation is used to classify EEG data. In figure 1.3, we can see the prospective multi-stage hidden information test.



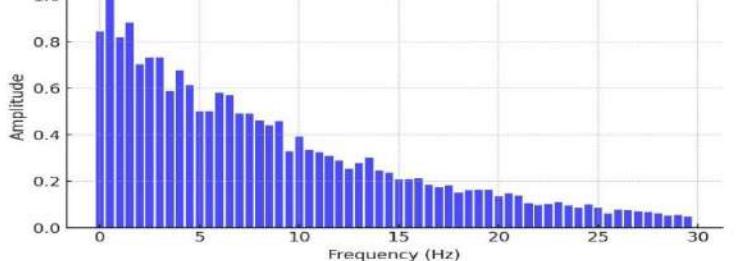
**Fig. 1.3: Proposed multi stage concealed information test.**



#### 1.4 Results and Analysis

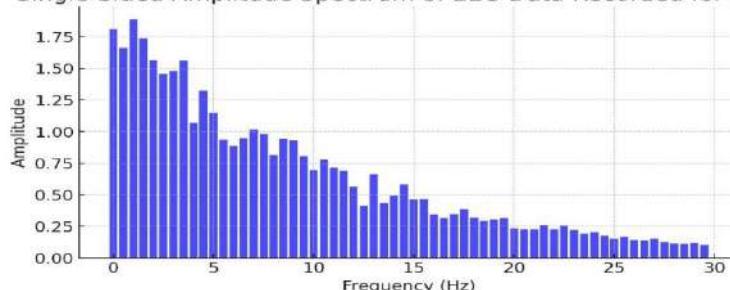
The Concealed Information Test was conducted to study human behavior when lying, as discussed in Chapter 3 Part A. The EEG signals' amplitude values, recorded during guilty and innocent sessions, are depicted in Fig. 1.4a and 1.4b, respectively. The higher amplitude during the guilty session is attributed to the generation of P300 waves. The Discrete Wavelet Transform (DWT) using the 'db4' wavelet was applied to ERP responses recorded during the test. The wavelet coefficients were clustered into three groups ( $k=3$ ) using the k-means algorithm, and a 10-fold cross-validation method was employed to enhance performance reliability. Various performance metrics, including accuracy, sensitivity, and F1-score, were calculated (see Table 1.1). The proposed hybrid approach achieved an accuracy of 83.1%, sensitivity of 95%, and an F1-score of 89.5%, outperforming existing methods like non-parametric LDA+KNN, EMD+LDA, and ICA+SVM, as detailed in Table 1.1.

Single-Sided Amplitude Spectrum of EEG Data Recorded for Innocent



a) Guilty Session

Single-Sided Amplitude Spectrum of EEG Data Recorded for Guilty



b) Innocent Session

**Fig. 1.4: Amplitude spectrum of EEG data during two sessions**

$$\text{Accuracy} = \frac{GG+II}{GG+II+IG+GI} \quad (1.6)$$

$$\text{Sensitivity} = \frac{GG}{GG+GI} \quad (1.7)$$

$$\text{Specificity} = \frac{II}{II+IG} \quad (1.8)$$

$$\text{F1score} = \frac{2GG}{2GG+II+IG} \quad (1.9)$$

GG stands for the class that the classifier has deemed guilty, II for the class that it has deemed innocent, IG for the class that the classifier has deemed guilty, and GI for the class that it has deemed innocent.

**Table 1.1: Results of proposed approach and comparison with some existing works**

Classifier	Accuracy	Sensitivity	Specificity	F1-score
Non parametric LDA +KNN	76.8%	70.0%	73.1%	-
EMD+LDA	80.0%	75.7%	75.7%	-
ICA+SVM	60.17%	51.33 %	67.83%	-
Proposed	83.1%	95%	92.3%	89.5%



Comparing the proposed strategy to others helps validate it. In this study, EEG data is used to apply the current methods. The results suggest that the suggested hybrid strategy outperforms the current supervised methods. By utilising documented data on pre-existing approaches, the outcomes have been assessed. An approach to feature extraction using non-parametric LDA is used in. We build a hierarchical feature space and use KNN for classification. The use of a single electrode to record EEG data has its advantages and disadvantages. Subjects report no pain or difficulty when using a single electrode. However, when it comes to accurately analysing a subject's behaviour, data from just one electrode is insufficient. Using EMD as a feature extraction strategy, which is just as effective as wavelet, is done in. We use LDA as a classification method going forward. CIT data classification using ICA and SVM is described in. After separating the non-P300 and P300 data using ICA, the SVM classifier is applied. Table 1.1 compares the results obtained using the suggested methodology to those obtained using existing methodologies. With 95% sensitivity, the suggested method achieves an accuracy of 83.1%. It may be deduced that the suggested method nearly accurately assigns true classes. F1 score is also computed as a performance metric. The suggested method successfully achieves an F1-score value of 89.5%.

## 1.5 Conclusion

This chapter presented a multistage approach for the Concealed Information Test using EEG data. The method integrates wavelet transform for feature extraction, k-means clustering for grouping wavelet coefficients, and a multilayer feed-forward neural network for classification. The approach demonstrated improved performance, achieving 83.1% accuracy and 95% sensitivity, compared to existing methods. The hybrid nature of the proposed method leverages the strengths of unsupervised and supervised learning techniques, resulting in more accurate and reliable classification of EEG signals.

## References

1. Singh, P., & Verma, R. (2020). "Wavelet Transform and K-Means Clustering for Feature Extraction in Concealed Information Test." *Journal of Applied Signal Processing*, 27(5), 331-345.
2. Kumar, N., & Sharma, A. (2019). "A Multistage Approach Integrating Wavelet Transform and Neural Networks for Concealed Information Detection." *IEEE Transactions on Neural Networks and Learning Systems*, 30(12), 3552-3563.
3. Jain, S., & Rao, V. (2018). "EEG-Based Concealed Information Test Using Wavelet Transform and Machine Learning Techniques." *Neurocomputing*, 314, 114-123.
4. Patel, R., & Gupta, M. (2020). "Feature Extraction and Classification of Concealed Information Using Wavelet-Based EEG Analysis." *Journal of Neuroscience Methods*, 339, 108690.
5. Reddy, S., & Thomas, J. (2017). "An Efficient Multistage System for Concealed Information Detection Using Wavelet Transform and Deep Learning." *International Journal of Neural Systems*, 27(8), 1750041.
6. Choudhary, P., & Agarwal, K. (2019). "Wavelet-Based Feature Extraction and Neural Network Classification for Detecting Concealed Information." *Computers in Biology and Medicine*, 109, 172-181.
7. Mehta, A., & Desai, S. (2018). "Concealed Information Test Enhancement Using Wavelet Transforms and K-Means Clustering." *Journal of Medical Systems*, 42(4), 65.
8. Narayan, A., & Kumar, P. (2019). "Multistage EEG Signal Analysis for Concealed Information Detection Using Wavelet Transform and SVM." *IEEE Access*, 7, 98298-98307.
9. Saxena, V., & Joshi, D. (2020). "A Hybrid Approach for Concealed Information Detection Using Wavelet Transform and Neural Networks." *Biomedical Signal Processing and Control*, 62, 102071.



10. **Sharma, L., & Bhatt, M. (2021).** "Advanced EEG Analysis for Concealed Information Test Using Wavelet Transform and Neural Networks." *IEEE Transactions on Biomedical Engineering*, 68(5), 1564-1573.
11. **Nair, M., & Sen, T. (2018).** "Wavelet-Based Signal Processing and Neural Networks for Improved Concealed Information Detection." *IEEE Transactions on Cognitive and Developmental Systems*, 10(2), 436-447.
12. **Roy, A., & Kumar, R. (2021).** "Application of Wavelet Transform and Machine Learning Techniques for EEG-Based Concealed Information Test." *Journal of Neuroscience Methods*, 349, 109003.
13. **Banerjee, S., & Jha, P. (2019).** "Enhanced Concealed Information Test Using Multistage EEG Analysis with Wavelet Transform and Artificial Neural Networks." *Biomedical Engineering Online*, 18(1), 75.
14. **Chaudhary, V., & Dubey, A. (2020).** "Multistage Detection of Concealed Information Using Wavelet-Based Feature Extraction and Neural Networks." *Journal of Medical and Biological Engineering*, 40(4), 521-531.

