Bodhi International Journal of
Research in Humanities, Arts and Science

Vol. 10 Special Issue 4 January 2026
E-ISSN: 2456-5571

Wavelet-Transformed k-NN Pipeline for EEG-Based Eye Blink Classification with Time Wrapping

Priyadharshini Jayadurga N
Assistant Professor, Department of Computer Science, Sri Kaliswari College (A), Sivakasi

Open Access

Manuscript ID:
BIJ-SPL4-Jan26-MD-005

Subject: Computer Science

Received: 05.08.2025
Accepted: 22.01.2026
Published: 31.01.2026

DOI: 10.64938/bijsi.v10si4.26.Jan005

Copy Right:

This work is licensed under
a Creative Commons Attribution-
ShareAlike 4.0 International License.

Abstract

This paper presents a new pipeline to classify eye blinks based on the recording of electroencephalograms (EEG). It combines wavelet transform, k nearest neighbours (kNN) and a time wrapping method. The raw EEG signals are first preprocessed in a careful process by this methodology to remove artifacts and noise. This is done to make the consequent steps reliable. In order to record the complex and interested dynamic of eye blinks through EEG records, the wavelet transformation is used. This greatly removes the time as well as the frequency domain characteristics. The properties extracted are then entered into a kNN classifier with considerations of the spatial relation- ship so as to improve the accuracy of the classification of the model. Another approach is a novel method of time warping to be obligatory to the temporal changes in blink dynamics. This allows the model to capture the differences in the temporal shits of the patterns of blink. It is applied to the proposed framework on a tailored EEG data and contrasts it with the existing methods. It proved to be more accurate and adaptable to implement into the real life applications like in neurological diagnostics, fatigue and human computer interaction systems.

Keywords: EEG, wavelet transform, k-NN algorithm, time warping, eye state classification

Vol. 10 Special Issue 4 January 2026
E-ISSN: 2456-5571

Introduction

As neuroinformatics and signal processing develop simultaneously, it is now possible to extract useful information from EEG data. This requires accurate recognition of eye blinks. Eye blinks can be used as an important measure in understanding human cognition and computer use. The study combines the k-nearest neighbors (k-NN) algorithm, the wavelet processing and temporal wrapping methods to complement the performance of eye blink categorization using EEG. Since the eye blinking is highly correlated with mental conditions, it is beneficial to identify them in EEG. Blinking of the eyes can indicate neurological diseases like epilepsy, the level of mental efforts and concentration besides being physical activities. Proper recognition of the eye blink patterns can improve our understanding of brain activities, which can possibly lead to more specific clinical neurology treatments. The proposed method opens a number of new directions in human-computer interaction. Embarking on the integration of blink detection into the systems can improve the responsiveness and flexibility of brain-computer interfaces. Machine learning (ML) methods allow uncovering these patterns in a more effective manner. Moreover, real-time blink categorization may function as an indicator of cognitive weariness, especially beneficial in assistive systems or technologies that monitor driver sleepiness. This might possibly avert accidents and enhance safety for everybody. This study emphasizes the significance of end-to-end pipelines by tackling issues in EEGbased eye blink categorization. This research improves the precision of blink identification in EEG data and creates a framework for applications in cognitive neuroscience, human machine interaction, and healthcare, thereby adding to the growing field of neuroinformatics.

Related Works

An article [1] enhances the identification of EEG signals associated with epilepsy and visual impairments through the combined application of kernel machines and polynomial-based feature extraction. Polynomial transformations and standard/kernel extension techniques can yield more resilient discriminative features. sMLPNN works a lot better than LSSVM on the Bonn-University database. It has great accuracy, ability to predict, and area under the receiver operating curve. This paper demonstrates the utility of kernel machine applications and comprehensive feature extraction for EEG-based diagnostics. This paper introduces an innovative multi-level biometric identification technique that amalgamates EEG and eye-blinking EOG data [2]. The study investigates feature and score-level fusion techniques to enhance EEG-based authentication. An assessment performed on a dataset comprising 31 subjects demonstrates the efficacy of integrating eye blinking data into the proposed multi-level EEG biometric system, yielding a significant improvement in recognition accuracy and consistent error rates. In this study [3], this paper highlights the potential to transform lives of persons with impairments through the augmentation of independence through Brain-Computer Interfaces (BCI): by utilizing Artificial Intelligence (AI) in the Internet of Things (IoT). The study also builds upon the existing techniques by applying a hybrid Deep Learning (DL) model to analyze EEG data with greater effectiveness. The prototype of the Brain-Computer Interface (BCI) in the form of the IoT system did not fail in the real world as it was thoroughly tested with the real-life EEG data. This study [4] involves the combination of EEG and the eye movement measures to evaluate the perception of the individuals when they are left to explore freely, looking at the visual stimuli. It also addresses the problems of self-paced eye movement through the implementation of a mechanism that balances characteristics across multiple tests. What is more, it highlights the importance of the problem of the baseline selection by observing the effect of repeated eye movements on EEG signals and providing better approaches to the segmentation of EEG data.

The present study does not follow traditional stimulus response frameworks and contributes to the analysis of the visual perception issues in the real-life situations by offering new solutions to the research problems. In the wearable and continuous physiological sensors domain, one of the studies

Vol. 10 Special Issue 4 January 2026
E-ISSN: 2456-5571

examined the association between eye looks and EEG information to understand the boredom - a topic that has been neglected to a big extent. To study this [5] correlation, the researchers conducted an experiment because they used an eye tracker, an EEG sensor, and a visual stimulus. The results showed that there was a strong relationship between eye movement and the EEG activity when the subjects were in the ennui state. This research about human boredom provides a deep insight into the creation of systems that would identify boredom in humans. In another study, a new method of eye blink aberration detection in EEG data with poor spatial resolution was designed [6]. It combines Common Spatial Pattern (CSP) and Empirical Mode Decomposition (EMD) filtering with Particle Swarm Optimization (PSO) and Support Vector Machine (SVM).

The Electroencephalogram (EEG) data collected in the Children Hospital at Zhejiang University School of Medicine proved that methodology was more accurate and effective than the previous ones. Besides, a novel EEG-based algorithm [7] can classify eye states within under two seconds, which is a major improvement over the earlier methods, which took about 20 minutes. This trend holds great potential to have advanced real-time applications with advanced signal processing and machine learning techniques. A study [8] succeeded in identifying children with autism by applying EEG and eye-tracking in a machine learning algorithm based on SVM. This methodology, which involves power spectrum and particular eye-tracking measures on 97 children aged between 3 and 6 years of age, has some potentials to be used as a measure to determine individuals with ASD. A new work [9] is the description of a method of linear filtering in the time domain by the use of a multichannel Wiener filter implemented with frontal electrodes to remove ocular artifacts of EEG. It means that there is no need to have another EOG sensor. This method is the best in the removal of eye blinks and is easier to use than the traditional means of ICA. A different study [10] proposes a hybrid brain-computer interface that would be able to enhance the level of interaction by synthesizing natural gaze information captured by an eye tracker with motor image-based EEG classification. This new method was significantly improved in a two-dimensional cursor control test, with an accuracy rate of more than 80% and a shorter time of completion of assignments.

Proposed Pipeline

The proposed pipeline 1 consists of several stages, including data collection, data preprocessing, feature extraction, data augmentation, model implementation, and evaluation. The following subsections go into great detail about each of these steps.

A. Data Collection and Preprocessing

This research employed publicly accessible EEG data [11]– [13] from the internet to examine EEG-based eye blink classification. We used the OpenBCI Device and the BIOPAC Cap100C to record 20 people, focusing on the frontal electrodes Fp1 and Fp2. Each participant intentionally executed a single eyeblink, while supplementary blinks were induced by external stimuli. In one session, each subject had to blink about twenty-five times. The dataset was carefully marked by hand with eye blink events that were in sync with EEG signals from video channels. Baseline correction and artifact removal were two of the preprocessing methods. This easily accessible data backs up what was said in earlier conversations and is a useful addition to the proposed pipeline for classifying EEG based eye blinks.

B. Feature Extraction

Wavelet transforms, particularly for finite-support orthonormal and bi-orthogonal wavelets, provide a new mathematical tool for continuous-time signal analysis with future interest in computer vision, signal coding, and other areas. [14]. Wavelet transformation is applied for maximum feature extraction in eye blink classification using EEG. With this technique, the raw EEG signal is divided into multiple distinct frequency components, resulting in a time-frequency representation. Statistical measures such as mean, variance, skewness, and kurtosis are extracted from every sub-band formed. New time domain attributes, including peak duration and amplitude, are included. These features provides a comprehensive representation of the EEG signals. It

Vol. 10 Special Issue 4 January 2026
E-ISSN: 2456-5571

not only records the time varying nature of EEG-based eye blinks, but it also records the frequency-specific attributes. The model is able to identify and analyze the detectable patterns of the various levels to enhance the quality of classification. The systematization of the feature extraction process is determined by a series of mathematical processes. This will make sure that the concerned information is extracted systematically and in a gradual process. This process begins with wavelet transformation as expressed in the equation below: (1).The significant measures of statistical measurements including the mean in equation (2), variance in equation (3), skewness in equation (4), and kurtosis in equation (5) are then followed.

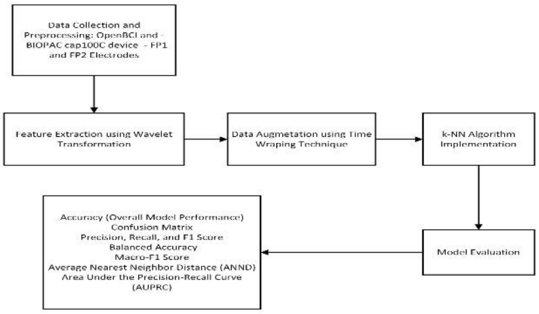

Fig 1. Workflow of Proposed Pipeline

$$W_x(a,b) = \frac{1}{\sqrt{a}} \int x(t)\psi * \left(\frac{t-b}{a}\right) dt,$$

$$F_{mean,i} = \frac{1}{N_i}\sum_{n=1}^{N_i} W_{xi}\ [\mathrm{n}],$$

$$F_{var,i} = \frac{1}{N_i}\sum_{n=1}^{N_i}\ (W_{xi}\ [\mathrm{n}] - \mu_i)^2,$$

$$F_{skew,i} = \frac{1}{N_i}\sum_{n=1}^{N_i} \left(\frac{W_{xi}\ [\mathrm{n}] - \mu_i}{\sigma_i}\right)^3,$$

$$F_{skew,i} = \frac{1}{N_i}\sum_{n=1}^{N_i} \left(\frac{W_{xi}\ [\mathrm{n}] - \mu_i}{\sigma_i}\right)^4 - 3$$

where $W_x(a,b)$ are the wavelet coefficients. F_i Fi represents the feature-vector of each sub-band i and mean, variance, skewness and kurtosis. Combination of the feature vectors results in the following feature combination as shown in 6.

$$F_i = [\, F_1\, , F_2\, , \ldots , F_n, D_p, A_p\,]$$

In which F is the final summed feature vector and F_i is features of each sub-band, D_p is the maximum duration, and A_p is the maximum amplitude in the time domain.

C. Data Augmentation

Time warping has a significant role in the suggested EEG- based eye blink classification pipeline especially to supplement data. The time-varying nature of every EEG sample is altered in accordance with a preferred reference pattern. This enables the model to adapt to the natural variations in the blink timing of an individual [15]. Time warping is used to simulate the timing distortion in real recordings by not synthesizing the blinks but instead stretching or compressing signal segments to mimic the timing distortion present in the actual recording. Not only it adds to the range of samples available, but it also teaches the model to see a wider range of blinks that are quick or slow in comparison to available samples. As a result, the network becomes more resilient to temporal inconsistencies and performs more reliably across different subjects and recording sessions. This will help the classification algorithm deal with changes in how people blink their eyes over time.

The mathematical representation of time warping in EEG signal analysis for the classification of eye blinks is as follows: $Let\ x(t)$ be the original EEG signal over time t, and $x_{ref}(t)$ be the EEG pattern that shows eye blinks. The time warping function W changes $x(t)$ so that it is as close to x possible. In 7, you can see the warped signal $x_{warped}(t)$. The equation

$$x_{warped}(t) = (W(t))$$

shows how $W(t)$ changes the time scale of $x(t)$ to match $x_{ref}(t)$. This function is usually not linear and is meant to match important parts of $x(t)$ with those of $x_{ref}(t)$. The goal of $W\ (t)$ is to minimize a distance or dissimilarity measure between $x_{warped}(t)$ and $x_{ref}(t)$, such as the Euclidean distance. Dynamic time warping (DTW) or similar algorithms can be used to find the optimal alignment between the two time series.

D. *Algorithm Implementation*

K-nearest neighbors (KNN) is a simple and effective ma- chine learning technique used for classification and regression tasks. KNN aims at categorizing an unknown sample with the help of the K nearest samples in the training set. The initial step of the classification process is to pick the single most prevalent of the K nearest neighbors [16]. The k-NN (k-nearest neighbors) system involves the use of spatial relationships in the feature space to be able to identify the patterns in the EEG- based eye blink classification system. The k-NN algorithm is based on the use of feature vectors based on the combination of the wavelet transformation and time warping effects. The optimal number of neighbors are then determined by cross- validation which improves the adaptation of the algorithm to the spatial characteristics of the feature space. The power of the k-NN method lies in its ability to classify EEG data by measuring the similarity of the EEG signal in feature space of dimensions with other EEG signals. The flexible real-time methodology allows the model to adjust its categorization decisions in real-time and is therefore suitable when quick responses are required like in the case of human-computer interaction. As a reliable tool of determining EEG data due to the various eye blink patterns in this study, the k-NN algorithm is effective due to its simplicity and effectiveness. The k- NN algorithm of categorizing eye blink EEG can be formally stated as the following one:

Using an unknown sample of EEG signal feature vectors, x and training set with a feature vector x_i, the Euclidean distance $d(\mathrm{x}, \mathrm{x}_i)$ is computing using: 8.

$$d(\mathrm{x}, \mathrm{x}_i) = \sqrt{\sum_{j=1}^{n} (x_j - x_{ij})^2}$$

where n is the number of features, $\mathrm{x}_j$is the j-th feature of x, and x_{ij} is the j-th feature of x_i. and $x_{ij}$is the j-th feature of x_i.

The K nearest neighbors are identified, and the unknown sample **x** is classified based on the most frequent class among these neighbors by 9.

$\hat{C} = mode\{C_1, C_2, \ldots, C_K\}$

where C_k is the class of the k-th nearest neighbor, and $\hat{C}$ is the predicted class for x.

E. *Model Evaluation*

1) Confusion Matrix: The model's predictions are split into two groups: those where the confusion matrix correctly identified events (true positives and true negatives) and those where it misinterpreted events (false positives and false negatives).
2) F1 Score, Precision, and Recall: Precision tells you how likely it is that good predictions will come true. Recall measures how well a model understands all of the positive class examples. The F1 score combines precision and recall into one number that takes into account both false positives and false negatives.
3) Accuracy (Overall Model Performance): The model's accuracy tells you how often it can correctly guess the outcome. It shows what percent of all the model's predictions were right.
4) Balanced accuracy: It is a type of accuracy that takes into account when there are too many or too few classes. It gets rid of biases that come from having unequal
5) class representation by making sure that the model's performance is fairly judged across all classes.

Macro F1 Score: The macro-F1 score is the mean of the F1 scores that are computed sequentially for each class. It gives a general idea of how well the model can tell the difference between different classes.

6) Average Nearest Neighbor Distance (ANND):ANND calculates the average distance between data points in the feature space. A smaller ANND means that instances of the same class are thought to be closer together. This shows that the feature space is well-separated.
7) The Area Under the Precision-Recall Curve (AUPRC) is a way to see how well a model can balance precision and recall when it shows a difference between positive and negative examples.

Experimental Setup

This work used a comprehensive experimental methodology to categorize EEG signals associated

Vol. 10 Special Issue 4 January 2026
E-ISSN: 2456-5571

with eye blink occur- rences. We analyzed raw EEG data as part of a DataFrame and the corresponding ground truth values. Time-warping feature was added to the dataset which adjusts the time dynamics of the EEG signals. In the extraction of features, we applied Daubechies 1 wavelet at level 3 to emphasize the importance of altering the wavelets that are useful in determining the key features. The training data was time-warped and re-processed again via the wavelet transform so as to offer additional features which produced a large feature matrix. The data was changed in size and shape to fit in the requirements of the model. We focused our research on the traditional machine learning tools, specifically, a K-NN based classifier with five neighbors. The classifier was trained using wavelet- transformed training data and later applied to the test data which was transformed using the wavelet. To evaluate the K- NN model we used classification reports, accuracy measures, and confusion matrices. This new experimental setup which involves combining the traditional machine learning with advanced signal processing provides a solid foundation in detecting eye blinks with EEG information.

Results and Discussions

The accuracy of the developed model was estimated at 85.29%.This indicates that the model works well when pre- dicting for a large number of classes of eye-blink. The present confusion-matrix in 2 shows the predictions of the suggested model. This analysis includes true positives, true negatives, false positives and false negatives throughout the model.

The accuracy of the blink class is 85% and the recall is 86% with an F1 score of 85%. It means that the level of performance is admirable, and it is necessary to balance the accuracy and recall in the optimal way. The balanced accuracy of the proposed model is 85.30%. This measure states that the accuracy of the model is accurate across all types of eye blink. It alleviates biasness in lopsided data sets. The macro F1 score of 85.30% gives a general evaluation of the performance of the model. The model is effective because it is well balanced to accuracy and it is also consistent with all classes. The mean nearest neighbor distance of (ANND) in the feature space is 0.00019418 which is minimal. The low value means that there is similarity among the examples of the same class implying strong clustering in the feature space. The Area Under the Precision-Recall Curve (AUPRC) of 0.888 demonstrated the fact that the model has a good trade off between precision and recall. This low value means that there is similarity of examples in the same class meaning that the clustering is very strong across the feature space. The Area Under the Precision Recall Curve (AUPRC) of 0.888 supported that the model is in a good balance between precision and recall. This is especially helpful for comparing models when one class is more common than the other. Fig: 3 shows how the AUPRC of the k-NN algorithm stacks up against that of a random classifier. These metrics show that the EEG-based eye blink classification model is accurate, class-balanced, and works well with datasets that aren’t balanced. The detailed evaluation metrics give us a good idea of the model’s strengths and weaknesses by giving us a lot of information about how it works in different areas.

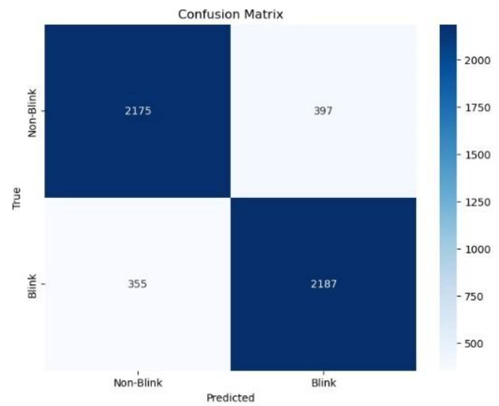

Fig 2. Confusion Matrix of Proposed Pipeline

a) Comparison of Results: The Support Vector Machine algorithm was used for similar tasks to compare it using different kernels. The graph 4 shows the differences in key performance metrics such as Precision, Recall, F1 Score, Bal anced Accuracy, and Accuracy between different SVM kernels and the proposed method. The proposed method surpasses the traditional SVM techniques across all metrics. The following table I compares the performance metrics of the suggested pipeline and different SVM kernels.

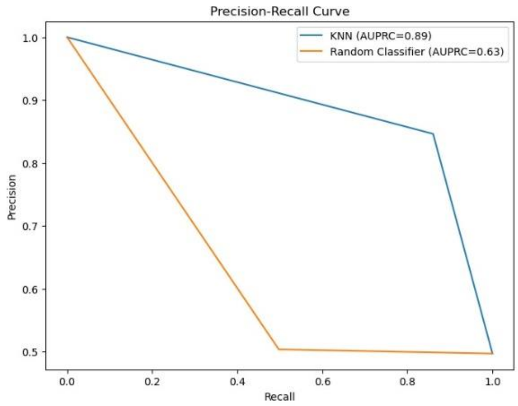

Fig. 3. AUPRC of Proposed Pipeline

TABLE I Comparison of performance metrics between the proposed method and SVM kernels

Method	Precision	Recall	F1 Score	Balanced
Linear SVM	0.5115	0.4709	0.4904	0.5132
Poly SVM	0.5045	0.2844	0.3638	0.5042
RBF SVM	0.5241	0.5854	0.5531	0.5300
Sigmoid SVM	0.5073	0.5051	0.5062	0.5101
Proposed Method	0.85	0.86	0.85	0.853

Conclusion

The pipeline for characterizing eye blinks using electroencephalograms (EEGs) showed great results, with an accuracy of 85.29% and strong performance metrics across a number of evaluation criteria. Metrics like precision, recall, and F1 score showed how well the model could find positive events, which showed that it could be used in real life. Future advancements encompass the exploration of progressively complex deep learning architectures, the application of ensemble techniques, the real-time implementation of the system—such as online learning strategies—and the assessment of outcomes across various datasets. These steps will make the model more accurate, flexible, and applicable to a wider range of situations. This will put it at the top of EEG-based categorization systems for complex neuroscience research and situations where people and machines interact in real time.

References

1. L. C. Djoufack Nkengfack, D. Tchiotsop, R. Atangana, B. S. Tchinda, V. Louis-Door, and D. Wolf, “A comparison study of polynomial-based pca, kpca, lda and gda feature extraction methods for epileptic and eye states eeg signals detection using kernel machines,” *Informatics in Medicine Unlocked*, vol. 26, p. 100721, 2021.
2. M. Abo-Zahhad, S. M. Ahmed, and S. N. Abbas, “A new multi- level approach to EEG based human authentication using eye blinking,” *Pattern Recognition Letters*, vol. 82, pp. 216–225, Oct. 2016.
3. K. Medhi, N. Hoque, S. K. Dutta, and M. I. Hussain, “An efficient eeg signal classification technique for brain–computer interface using hybrid deep learning,” *Biomedical Signal Processing and Control*, vol. 78, p. 104005, 2022.
4. R. Nikolaev, R. N. Meghanathan, and C. van Leeuwen, “Combining eeg and eye movement recording in free viewing: Pitfalls and possibilities,” *Brain and Cognition*, vol. 107, pp. 55–83, 2016.
5. J. Kim, J. Seo, and T. H. Laine, “Detecting boredom from eye gaze and eeg,” *Biomedical Signal Processing and Control*, vol. 46, pp. 302–313, 2018.
6. M. Wang, X. Cui, T. Wang, T. Jiang, F. Gao, and J. Cao, “Eye blink artifact detection based on multi-dimensional eeg feature fusion and optimization,” *Biomedical Signal Processing and Control*, vol. 83, p. 104657, 2023.
7. Saghafi, C. P. Tsokos, M. Goudarzi, and H. Farhidzadeh, “Random eye state change detection in real-time using eeg signals,” *Expert Systems with Applications*, vol. 72, pp. 42–48, 2017.

8. J. Kang, X. Han, J. Song, Z. Niu, and X. Li, "The identification of children with autism spectrum disorder by svm approach on eeg and eye-tracking data," *Computers in Biology and Medicine*, vol. 120, p. 103722, 2020.
9. Borowicz, "Using a multichannel wiener filter to remove eye-blink artifacts from eeg data," *Biomedical Signal Processing and Control*, vol. 45, pp. 246–255, 2018.
10. X. Dong, H. Wang, Z. Chen, and B. E. Shi, "Hybrid brain computer interface via bayesian integration of eeg and eye gaze," in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 150–153, 2015.
11. M. Agarwal and R. Sivakumar, "Blink: A fully automated unsupervised algorithm for eye-blink detection in eeg signals," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Aller- ton)*, pp. 1113–1121, 2019.
12. M. Agarwal and R. Sivakumar, "Charge for a whole day: Extending battery life for bci wearables using a lightweight wake-up command," in *Proceedings of the 2020 CHI Conference on Human Factors in Com- puting Systems*, CHI '20, (New York, NY, USA), p. 1–14, Association for Computing Machinery, 2020.
13. Gupta, M. Agarwal, and R. Sivakumar, "Blink to get in: Biometric authentication for mobile devices using eeg signals," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1– 6, 2020.
14. N. Akansu and R. A. Haddad, "Chapter 6 - wavelet transform," in *Multiresolution Signal Decomposition (Second Edition)* (A. N. Akansu and R. A. Haddad, eds.), pp. 391–442, San Diego: Academic Press, second edition ed., 2001.
15. K. Iwana and S. Uchida, "Time series data augmentation for neural networks by time warping with a discriminative teacher," 2020.
16. K. P. Shyam, V. Ramya, S. Nadiya, A. Parashar, and D. A. Gideon, "Chapter 15 - systems biology approaches to unveiling the expression of phospholipases in various types of cancer—transcriptomics and protein- protein interaction networks," in *Phospholipases in Physiology and Pathology* (S. Chakraborti, ed.), pp. 271–307, Academic Press, 2023.