Chapter 5

EMG scalogram-based classification of gait disorders using attention-based CNN: a comparative study of wavelet functions

---

Abstract

In this chapter a robust algorithm has been proposed for classifying gait abnormalities by analyzing scalogram images derived from electromyography (EMG) data. Given the intricate nature of EMG signals, which involve background noise high variability, wavelet transforms and attention-based Convolutional Neural Networks (CNN) were utilized to evaluate classification accuracy. The methodology uses four distinct wavelet functions—Complex Gaussian, Frequency B Spline, Mexican Hat, and Shannon—to construct scalograms, which are subsequently classified using an attention-based CNN model. The proposed model outperformed traditional classifiers, achieving 96.7% accuracy, 95.2% precision, 94.8% recall, and an AUC of 0.99 showing potential for clinical applications in diagnosing gait disorders.

5.1 Introduction

Gait analysis and EMG are crucial in evaluating walking abnormalities resulting from various factors such as age, injury, and disease (Khant et al., 2023). Gait analysis is a valuable tool for diagnosing a wide range of conditions affecting movement, especially those related to neurological and non-neurological disorders. On the other hand, EMG measures muscle activity during walking to help identify deviations from normal muscle activities (Romano et al., 2024). The analysis of EMG signals entails multiple steps, including filtering and feature extraction. One method for extracting frequency domain features from EMG signals is the wavelet transform, which provides detailed insights into signal structure across different frequency bands (Phinyomark et al., 2011).

5.2 Attention Networks

Over the past few years, significant advancements have been made in machine learning and deep learning, particularly in analyzing EMG signals and gait analysis. One critical development has been using Convolutional Neural Networks (CNNs), a specialized deep neural network known for its excellence in image recognition and computer vision tasks (Ketkar et al., 2021). CNNs are particularly effective in learning spatial hierarchies of image features, making them well-suited for classifying complex patterns in images. A typical CNN consists of multiple layers of neurons, with each layer performing specific computations on the input data. The initial convolutional layer applies adaptable filters to the input image, extracting local features such as edges and corners (Wu, 2017). Subsequent layers then further hone in on increasingly complex features by convolving the previous layer's output with additional filters. Ultimately, fully connected layers at the network's end generate predictions based on these learned features (Al-Saffar et al., 2017). Convolutional Neural Networks (CNNs) have advanced image recognition and computer vision but struggle with detecting partially obscured or occluded objects as they process the entire input image uniformly (Rangel et al., 2024). Attention mechanisms such as the Convolutional Block Attention Module (CBAM) have been developed and integrated into CNN architectures to address this (Vaswani, 2017). CBAM consists of two sequential sub-modules: the channel attention module and the spatial attention module. The channel attention module improves the network's ability to prioritize the input image's most informative channels (feature maps), ensuring that the network emphasizes the most critical aspects of the input data (Woo et al., 2018). Subsequently, the spatial attention module focuses on identifying the essential spatial regions within the feature maps, guiding the network's attention to the most pertinent regions of the image. Attention mechanisms like CBAM significantly enhance the performance of CNNs by allowing them to selectively focus on the most relevant features and regions within an image (Baozhou et al., 2021). This improvement in accuracy makes attention-augmented CNNs a powerful tool for various image recognition and classification tasks, contributing to better overall robustness and generalization.

5.3 Research Gap

Accurately classifying human locomotion utilizing EMG signals is a complex task, as it is affected by stochastic properties, sensor variations, and the intricacies of feature extraction (Sorkhabadi et al., 2019). Despite applying various classification techniques, including clustering, learning machines, and support vector machines, to EMG data, these challenges persist for the reasons mentioned above (Pauk et al., 2023). Gait analysis outcomes can be impacted by medication influencing the nervous system, potentially leading to altered results (Sethi et al., 2022). Another issue that arises from EMG signals is cross-talk, i.e., when neighboring muscles' signals interfere due to their proximity, leading to incorrect measurements of muscle activation patterns when muscles have similar activation patterns or are close to each other (Farina et al., 2004). In EMG, normalizing data with a reference signal like maximum voluntary contraction (MVC) or a percentage of gait cycle is common. However, identifying a reliable one can be difficult, and the normalization method can impact data comparability across studies (Sousa and Tavares, 2012). The accurate classification of EMG signals is also hindered by motion artifacts and noise, which can cause signal distortion and inaccurate readings. This presents a notable challenge in the field (Norali et al., 2009). Inter-subject variability in anatomy, muscle architecture, and gait mechanics can result in significant differences in EMG data across individuals. Consequently, the establishment of reliable muscle activation patterns norms for distinct subjects can be a challenging task (Guidetti et al., 1996). As described in Section 1.2, wavelet transform significantly mitigates these limitations (Phinyomark et al., 2011). There has been a constant push to improve the accuracy of classification algorithms. Commonly used classification algorithms suffer from a few limitations. The EMG classification techniques are tailored and trained for individuals or constrained settings. This can hinder their ability to effectively generalize to varying subjects, walking paces, landscapes, or extrinsic factors (Belinkov, 2022). The statistical properties of EMG signals can vary over time, known as non-stationarity. This variability can be attributed to factors such as muscle fatigue, electrode relocation, or variations in gait speeds. Consequently, classifiers may face challenges in effectively adjusting to these changes (Shen et al., 2022). The classification process requires extracting informative features from raw EMG signals. This is an essential step; however, it is challenging to select the appropriate features that adequately capture relevant information while minimizing noise and redundancy (Agarwal et al., 2023). EMG signals are complex as they often contain multiple channels and time-domain samples, resulting in high dimensionality. The high number of dimensions can pose challenges such as increased computational complexity, higher risk of overfitting, and difficulties in identifying significant patterns (Phinyomark et al., 2009).

5.4 Proposed Solution

It is already known that the wavelet transform provides a valuable advantage by allowing the extraction of a specific subset of frequency components or scales from a given signal of interest (Phinyomark et al., 2011). Scalograms use Continuous Wavelet Transform to capture temporal information by turning 1D signal data into a 2D image format. This allows for the implementation of deep learning-based classifiers like CNNs for the classification of images, thus achieving better classification accuracy (Kim and Seo, 2023). The CWT has proven to be a valuable tool in biomedical signal processing, particularly in the analysis of EEG, and EMG (Krishnan and Athavale, 2018). EMG-based scalograms effectively quantified simulated muscle activity even in noisy conditions (Di Nardo et al., 2022). EMG time-frequency activity has been identified in gait abnormalities like Parkinson's disease using scalograms (Romanato et al., 2021). Scalograms are capable of highlighting the difference in muscle recruitment timing for simulated EMG signals. Using scalograms, Romanato et al. (2022) assessed gastrocnemius lateralis muscle recruitment in patients with Parkinson's disease. Hence, Wavelet-based analysis of sEMG signals eliminates noise, leading to better identification of muscle force and classification (Veer and Agarwal, 2014). To further resolve the issues regarding classification, an attention-based classifier is incorporated. Attention mechanisms in deep learning help classifiers focus on essential parts of input data, leading to more accurate results (Niu et al., 2021). Attention mechanisms are a powerful technique for improving the performance and interpretability of classifiers. Attention-based models can capture intricate patterns and relationships within the data by selectively focusing on relevant parts of the input data (de Santana Correia and Colombini, 2022; Ghaffarian et al., 2021). These mechanisms also adaptively adjust the importance of different input elements based on the current context, resulting in more accurate and contextually appropriate predictions. In addition to improving performance and interpretability, attention mechanisms also make the model more robust to noisy or irrelevant data, reduce the risk of overfitting, and effectively combine and weigh information from different modalities of data (Galassiet al., 2020). They also capture complex dependencies between different parts of the input data, making them especially useful in tasks involving sequential or structured data (Veličković et al., 2017). Despite their focus on specific elements, attention mechanisms can still be computationally efficient, particularly when compared to fully connected layers that process the entire input (Zhang et al., 2022). This efficiency, combined with their state-of-the-art performance in various tasks, including machine translation, text generation, image captioning, and speech recognition, makes attention-based models a versatile and effective technique for improving the performance of classifiers across a range of applications (Gonçalves et al., 2022). In this study, an algorithm that utilizes attention-based CNNs and combined scalograms to improve the accuracy of EMG data for classifying gait disorders is proposed. Four wavelet families are used to generate the scalograms: Complex Gaussian, Frequency B Spline, Mexican Hat, and Shannon wavelets. Combining the scalograms allows for feature extraction in all four domains and minimizes the impact of noise and artifacts on the signals. Additionally, the attention-based CNN focuses on specific variations to aid in classification.

5.5 Materials and Methods

5.5.1 Data Collection and Preprocessing

Participants in this study were recruited from Sir Sunderlal Hospital (IMS), Varanasi, and local clinics. Ethical approval was obtained from the Institute Ethics Committee at the Institute of Medical Science (Banaras Hindu University), and all participants provided informed consent. The EMG data was collected from the tibialis anterior (TA) and medial gastrocnemius (MGAS) muscles using two DataLITE Wireless Surface EMG devices. The sampling rate was set to 1000 samples per second, and the participants were instructed to walk on level ground at a comfortable pace of 1-1.4 m/s without any support to ensure their comfort and safety. Figure 5.1 shows the complete workflow of the current study.

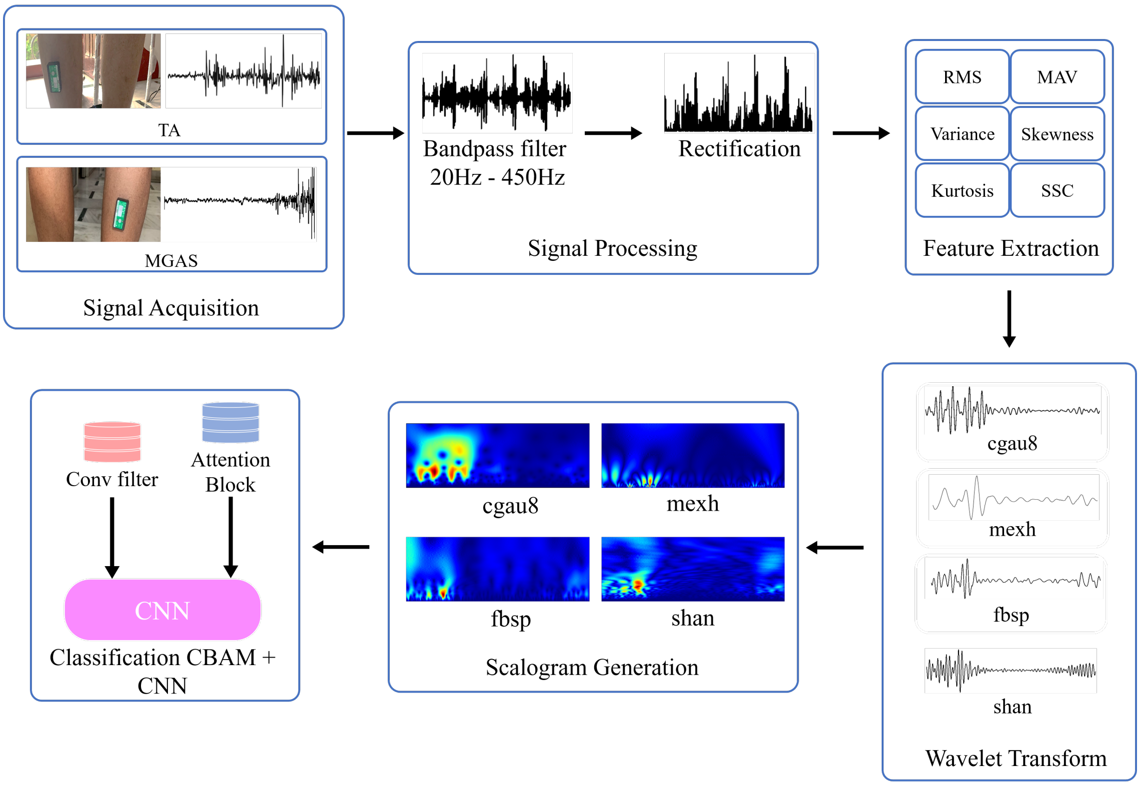


Fig. 5.1 Workflow of the present study

5.5.2 Wavelet Transform and Scalogram Generation

The Continuous Wavelet Transform (CWT) is a mathematical tool for signal analysis in both time and frequency domains. It involves convolving a function with a scaled and translated wavelet function. The scale factor determines the width of the wavelet, impacting the captured frequency components. In contrast, the translation factor shifts the wavelet along the time axis. A scalogram is a visual representation often used in wavelet analysis to illustrate the energy distribution of a signal across different scales or frequencies over time. It provides benefits such as time-frequency localization, pattern recognition, and improved interpretation. In the analysis of Electromyography (EMG), scalograms assist in identifying muscle contractions and frequency content, which can be valuable in diagnosing muscle disorders. The wavelets used in this study are Complex Gaussian (cgau8), Frequency B Spline (fbsp), Mexican Hat (mexh), and Shannon (shan). Table 5.1 lists the formula and brief description of wavelet families used in present study.

Table 5.1 Wavelets used in present study

Wavelet	*Equation*	*Description*
CWT	$CWT(a,b) = \int_{-\infty}^{\infty} x(t) \cdot \varphi_{a,b}(t)dt$	Basic CWT Equation
mexh	$\varphi_{Mexh}(t) = (1-t^2) \cdot e^{-\frac{t^2}{2}}$	Mexican Hat wavelets are suitable for detecting transient features in signals.
shan	$\varphi_{shan}(t) = \text{sinc}(t)$	Shannon wavelets are suitable for analyzing signals with transient features.
cgau8	$\varphi_{Cgau8}(t) = e^{-\frac{t^2}{2}+i8\pi t}$	Complex Gaussian wavelets are effective for detecting oscillatory patterns in signals with complex dynamics.
fbsp	$\varphi_{fbsp}(t) = \int_{-\infty}^{\infty} x(t) \cdot \varphi_{a,b}(t)dt$	Fbsp has flexible frequency response.

5.5.3 Conventional Machine Learning

The extracted features were classified using conventional machine learning algorithms to compare the results with the signal classification approach. These are K-Nearest neighbors (KNN), Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), AdaBoost, Gradient Boost, and XGBoost. K-Nearest Neighbors (KNN) is a classification algorithm that identifies the majority class of the nearest neighbors of data points. Naïve Bayes is a probabilistic algorithm that uses

Bayes' theorem and assumes feature independence to predict class probabilities. Support Vector Machines (SVM) aim to find the hyperplane that best separates classes in high-dimensional spaces. Decision Trees use a tree-like structure to split data based on feature values recursively. Logistic Regression is a classification method that uses the logistic function to estimate probabilities. Random Forest combines multiple decision trees to enhance accuracy and reduce overfitting. AdaBoost focuses on improving weak learners' performance sequentially by adjusting the weights of misclassified instances. Gradient Boost iteratively adds weak learners to construct a strong learner that corrects errors of previous iterations. XGBoost is an optimized gradient-boosting algorithm known for its speed and performance. These algorithms offer unique approaches to solving classification problems across diverse datasets. Machine learning algorithms come with a diverse range of hyperparameters. These parameters help optimize the algorithms' performance. For example, in K Nearest Neighbor (KNN), hyperparameters like the number of neighbors, weighting strategy, and algorithm type play a crucial role. They affect the accuracy of classification results. Similarly, in SVM, the regularization strength is determined by 'C'. The kernel type, gamma, and degree can be altered to tailor the model's ability to adapt to different complexities. In decision trees, parameters like maximum tree depth, minimum samples required for a split, and minimum samples in a leaf node have a significant influence on the model's complexity. Logistic Regression depends on 'C' to manage regularization strength. The penalty choice ('l1' or 'l2') determines the type of regularization applied, and the solver specifies the optimization algorithm. Random Forest explores the number of trees in the forest, maximum tree depth, minimum samples required for a split, and minimum samples in a leaf node. AdaBoost and Gradient Boost tinker with the number of boosting stages and learning rate. These parameters allow fine calibration of the boosting process, finding the spot between model complexity and predictive power. Lastly, XGBoost introduces the number of boosting stages, learning rate, maximum tree depth, and minimum child weight. These parameters ensure the model's optimal performance across diverse datasets. The meticulous parameter selection process caters to the distinctive requirements of each algorithm, facilitating a comprehensive exploration of model performance during the grid search. To ensure robust and accurate classification, an extensive Grid Search and 5-fold cross-validation were implemented with each classifier in this study.

5.5.4 CNN and Convolution Block Attention Module

After obtaining the scalogram images, training and testing datasets were created by splitting the dataset in a 75:25 ratio. The total number of images obtained across all classes was 1000. Convolutional Neural Network (CNN) and Convolutional Block Attention Module (CBAM) were coded in the TensorFlow library. In addition, a CNN filter was used in conjunction with the CBAM module, comprising 5×5, 3×3, and 1×1 convolution layers. The output from the CNN filter was concatenated with the output from the CBAM block and then further fed into the neural network. The use of a CNN filter in tandem with the CBAM module was inspired by Inception Networks, which leverages stacked convolution filters to extract features from images. This design was implemented to capture complementary information from the CBAM module, generating an additional feature map. The CNN filter's ability to extract features at varying scales and resolutions was leveraged to enhance the neural network's overall discriminative power. Figure 5.2 shows the complete network architecture used in this study. Figure 5.3(a) shows the components of the attention block, and Figure 5.3(b) shows the components of the convolution block. Equations 5.1 and 5.2 shows the representation of CBAM involving element-wise multiplication of Channel and

Spatial features with the input.

$$F' = M_c(F) \otimes F \tag{5.1}$$

$$F'' = M_s(F') \otimes F' \tag{5.2}$$

Where F is the input, is element-wise multiplication, M_c and M_s are channel and spatial features, respectively. $F^{'}$ is the product of input with the channel attention map and $F^{''}$ is the output of CBAM obtained by multiplying spatial attention map with $F^{'}$.

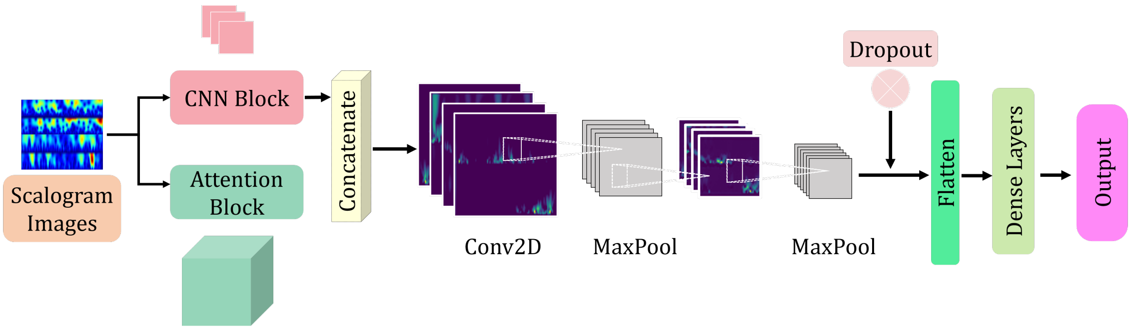
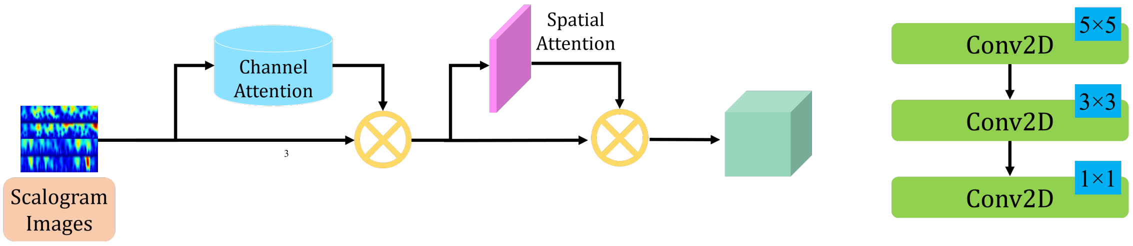


Fig. 5.2 Schematic of the complete network



(a) (b)

Fig. 5.3 (a) Components of Attention Block. (b) Components of CNN Block

Channel Attention Module

The channel attention mechanism captures the interdependencies among channels and learns channel-wise feature dependencies. The mechanism generates a channel attention map by first subjecting the input feature map to global average pooling to obtain a channel descriptor vector. The vector is then passed through two fully connected layers of a shared multi-layer perceptron (MLP) with Rectified Linear Unit (ReLU) activations. The resultant channel attention map is multiplied with the input feature map to highlight informative channels. Figure 5.4 shows schematic of channel attention module. Equation 5.3 represents the channel attention map computation.

$$M_c(F) = \sigma\left(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))\right) \tag{5.3}$$

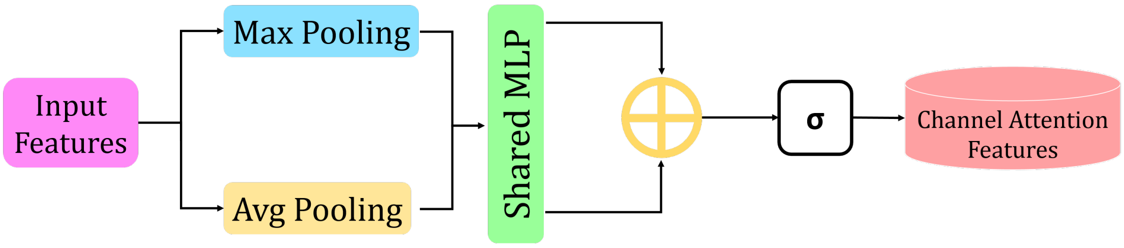


Fig. 5.4 Schematic of channel attention module

Spatial Attention Module

The spatial attention mechanism captures the interdependencies among spatial locations and learns spatial feature dependencies. This mechanism involves applying two convolutional layers with Rectified Linear Unit (ReLU) activations to obtain a spatial attention map. The map is multiplied with the input feature map to highlight informative spatial locations. Computation of spatial attention map is given by Equation 5.4.

$$M_s(F) = \sigma\left(\mathrm{Conv}\left([\mathrm{AvgPool}(F); \mathrm{MaxPool}(F)]\right)\right) \tag{5.4}$$

The spatial attention map is denoted by M_s. The spatial attention mechanism involves concatenating max pooled and average pooled features, which are then subjected to a convolution operation denoted by $Conv()$. The resulting output is passed through the sigmoid activation function σ. Figure 5.5 shows the components of the spatial attention module.

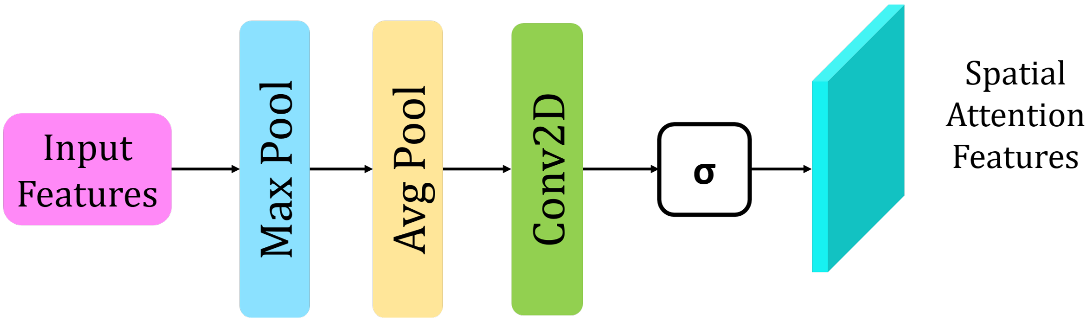
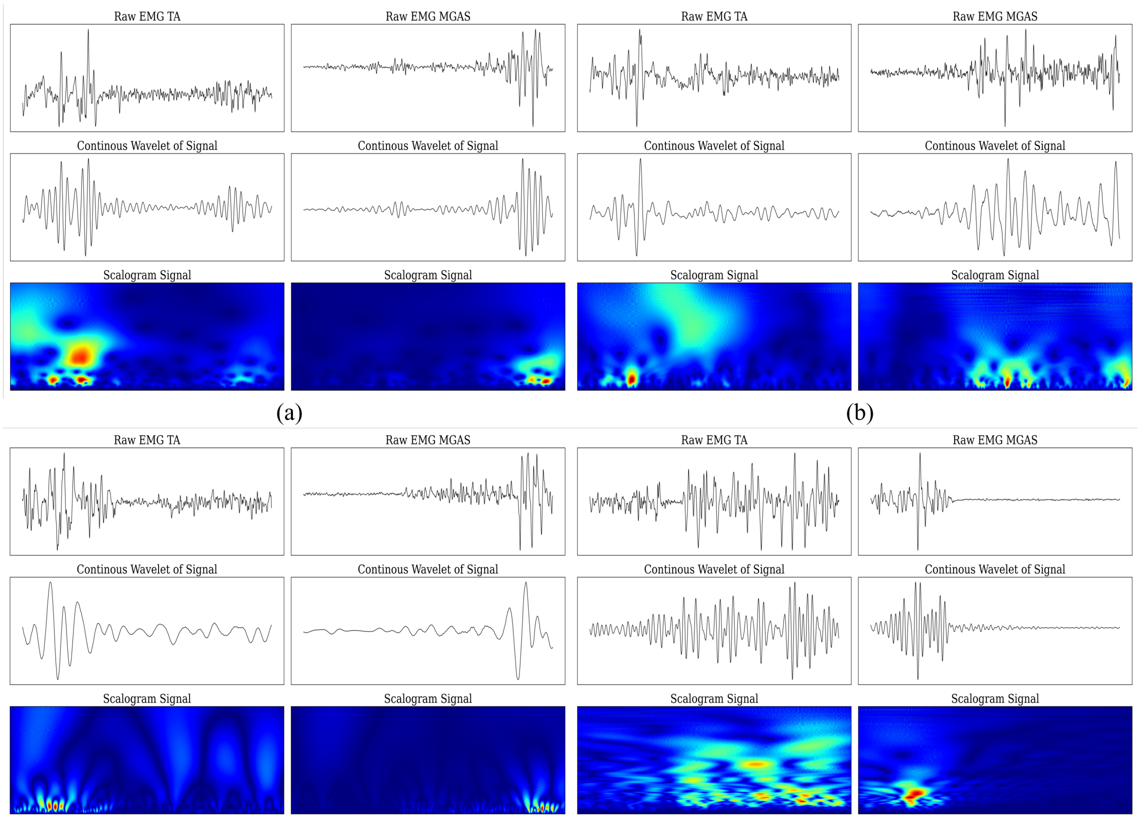


Fig. 5.5 Schematic of channel attention module

CNN is a complex system that relies on several hyperparameters like filter and kernel size, optimizer functions, loss functions, performance metrics, batch size, and epochs that must be optimized and fine-tuned during training. An Early Stopping function and dropout technique were employed to prevent model overfitting. For a thorough comparative analysis, each wavelet was individually classifed, and then classification was done on a combined image with four scalogram families.

5.6 Results and Discussion

Figure 5.6 compares wavelets and scalograms using the four wavelet families for both muscles. It can be observed that each wavelet gives a different representation of the scalogram. The performance metrics of ML classifiers are shown in Table 5.2. The analysis shows that ensemble methods outperform other classifiers, specifically Random Forest (RF) and XGBoost.



(a) (b)

(c) (d)

Fig. 5.6 Wavelet Types and Corresponding Scalograms: (a) Complex Gaussian (cgau8), (b) Frequency B-Spline (fbsp), (c) Mexican Hat (mexh), and (d) Shannon (shan).

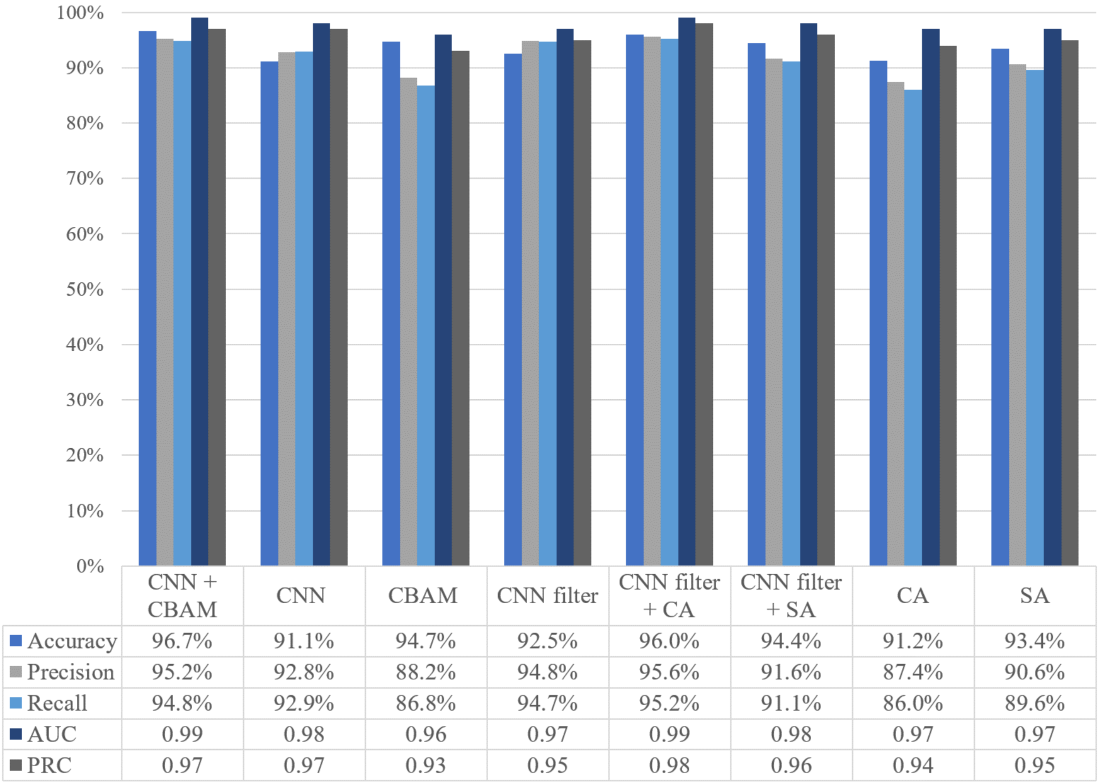
Table 5.2 Performance metrics of ML classifiers

Classifier	***Accuracy***	***Recall***	***Precision***	***F1-Score***
KNN	68.12%	67.42%	66.33%	0.67
NB	64.70%	64.64%	62.09%	0.62
SVM	82.56%	82.35%	81.41%	0.82
DT	77.94%	78.17%	76.85%	0.78
LR	65.38%	64.86%	63.86%	0.65
RF	93.18%	91.13%	90.37%	0.91
AdaBoost	73.32%	72.77%	71.62%	0.72
Gradient Boost	83.12%	82.96%	82.03%	0.83
XGBoost	90.38%	89.36%	89.67%	0.89

RF achieved an accuracy of 93.18%, recall of 91.13%, precision of 90.37%, and an F1-score of 0.91. In comparison, XGBoost attained an accuracy of 90.38%, recall of 89.36%, precision of 89.67%, and an F1-score of 0.89. Support Vector Machine (SVM) and Gradient Boost also demonstrated strong performance, with SVM achieving an accuracy of 82.56%, recall of 82.35%, precision of 81.41%, and an F1-score of 0.82, and Gradient Boost achieving an accuracy of 83.12%, recall of 82.96%, precision of 82.03%, and an F1-score of 0.83. Decision Tree (DT) performed well with an accuracy of 77.94%, recall of 78.17%, precision of 76.85%, and an F1-score of 0.78. K-Nearest Neighbors (KNN), Naïve Bayes (NB), Logistic Regression (LR), and AdaBoost showed moderate performance, with accuracies ranging from 64.70% to 73.32% and F1-scores between 0.62 and 0.72.

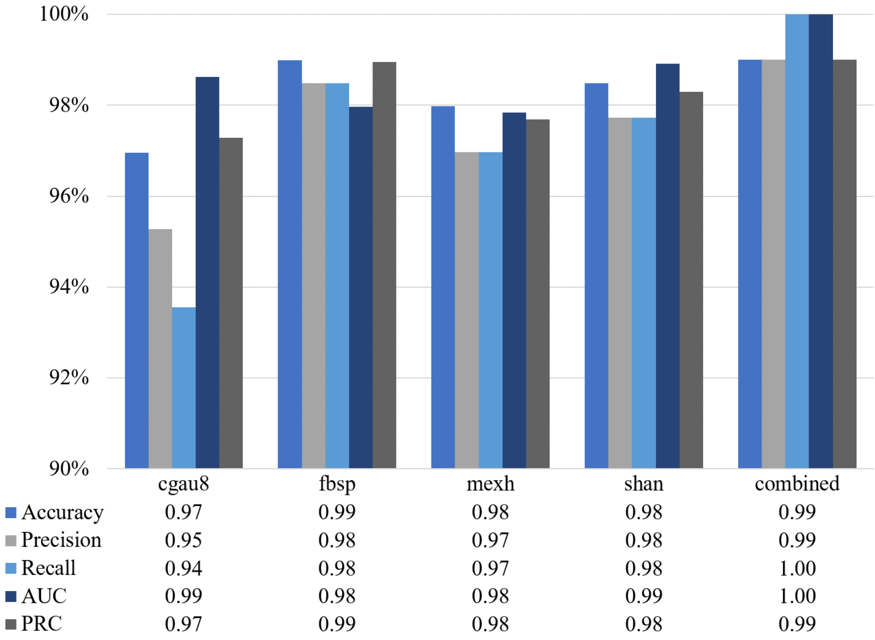
Figure 5.7 summarizes the classification metrics of all eight combinations of classifiers. The best accuracy was achieved when CBAM was combined with a CNN filter. The model gave 96.7% accuracy with a precision of 95.2% and a recall of 94.8%. The AUC was 0.99, and PRC was 0.97. The performance of each classifier was also evaluated when used individually. The CNN filter without any modules gave an accuracy of 91.1%. When used individually, the CNN filter and CBAM gave 92.5% and 94.7% accuracy, respectively. Using the CNN filter with channel attention (CA) and spatial attention (SA) modules separately resulted in accuracy of 96.0% and 94.4%, respectively. Using only channel and spatial attention modules separately gave 91.2% and 93.4% accuracy.

The results of classification using CBAM + CNN are shown in Figure 5.8. Here, classification was performed on individual scalograms and a combined one. It is evident that the performance metrics change upon varying the wavelet function. The attention-based network achieved a decent accuracy on all wavelet types, indicating its robustness. Combining the scalograms achieved an accuracy of 99%, precision of 99%, and recall of 100%. The AUC curve was 1.0, and the precision-recall coverage (PRC) 1.0 was achieved.



	CNN + CBAM	CNN	CBAM	CNN filter	CNN filter + CA	CNN filter + SA	CA	SA
■ Accuracy	96.7%	91.1%	94.7%	92.5%	96.0%	94.4%	91.2%	93.4%
■ Precision	95.2%	92.8%	88.2%	94.8%	95.6%	91.6%	87.4%	90.6%
■ Recall	94.8%	92.9%	86.8%	94.7%	95.2%	91.1%	86.0%	89.6%
■ AUC	0.99	0.98	0.96	0.97	0.99	0.98	0.97	0.97
■ PRC	0.97	0.97	0.93	0.95	0.98	0.96	0.94	0.95

Fig. 5.7 Summary of classification results of different classifier combinations



	cgau8	fbsp	mexh	shan	combined
■ Accuracy	0.97	0.99	0.98	0.98	0.99
■ Precision	0.95	0.98	0.97	0.98	0.99
■ Recall	0.94	0.98	0.97	0.98	1.00
■ AUC	0.99	0.98	0.98	0.99	1.00
■ PRC	0.97	0.99	0.98	0.98	0.99

Fig. 5.8 Classification results for all wavelet types

Figure 5.9 shows the accuracy and loss curves plotted for cgau8. Here, the model quickly goes to more epochs before being stopped. Figure 5.10 shows the curves for fbsp. Here, the model quickly goes to more epochs before being stopped. Figure 5.11 shows the curves for shannon wavelet. Figure 5.12 shows curves for the combination of all wavelets. It can be inferred that the combination of wavelets yields the best results, and the model converges quickly. Even on multiple runs, the results were consistent.

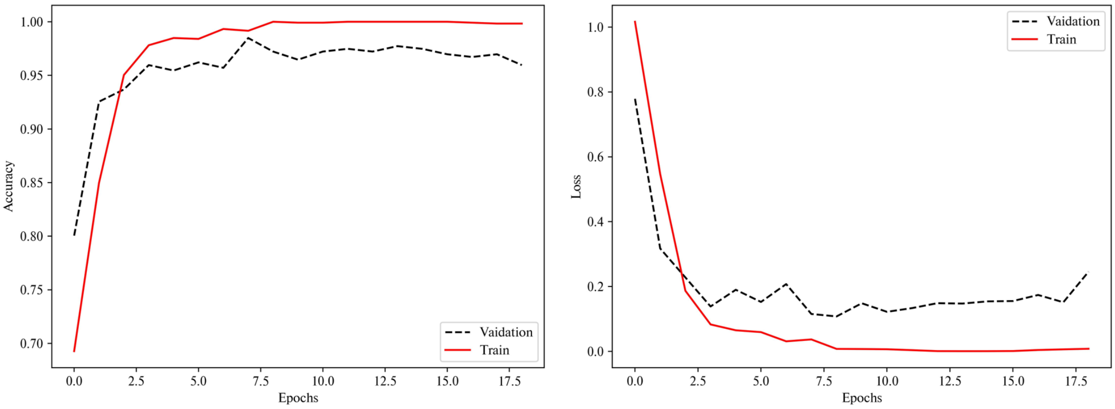


Fig. 5.9 Accuracy and loss curves for cgau8

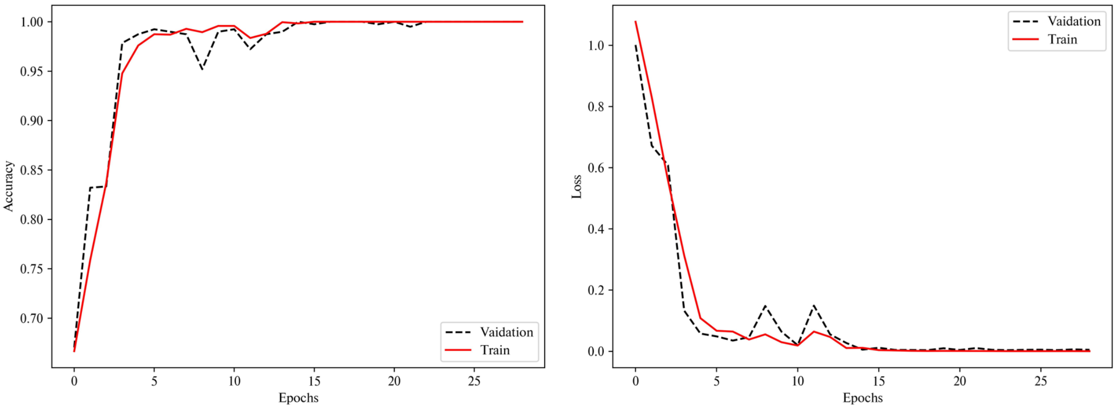


Fig. 5.10 Accuracy and loss curves for fbsp

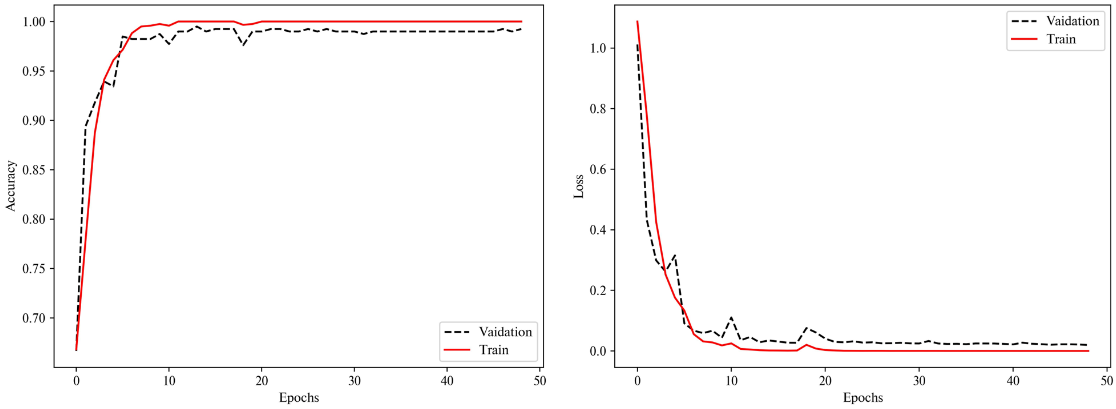


Fig. 5.11 Accuracy and loss curves for shan

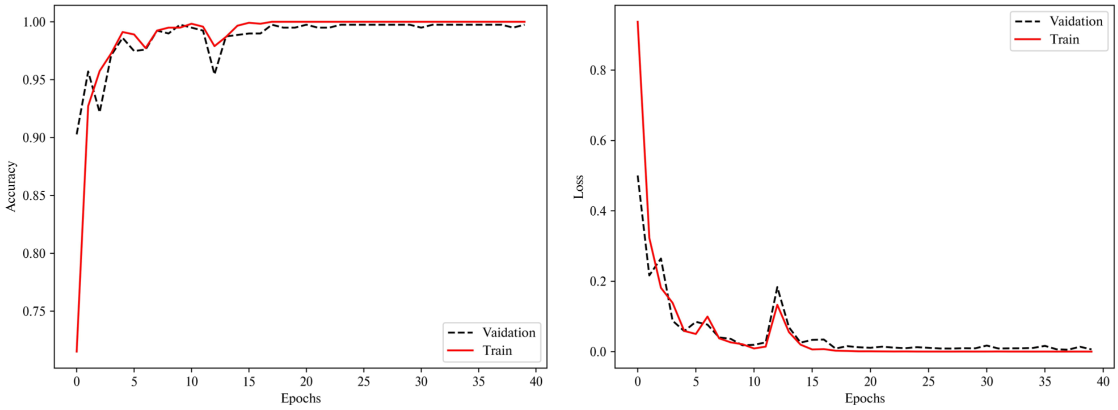


Fig. 5.12 Accuracy and loss curves for combination of all wavelets

5.7 Conclusion

Gait analysis provides a valuable window into the effects of various conditions on an individual's gait patterns. By identifying these patterns, clinicians can better evaluate the impact of a given condition and develop targeted interventions to improve function and mobility. This study used the Convolutional Block Attention Module (CBAM) architecture combined with Convolutional Neural Networks (CNN) to classify individual and combined scalograms generated from different wavelet functions. The CBAM + CNN model achieved a decent accuracy on all wavelet types. Upon combining the scalograms, an accuracy of

99%, precision of 99%, recall of 100%, area under ROC curve of 1.0, and precision-recall coverage (PRC) of 1.0 is achieved. By incorporating the scalograms generated from multiple wavelet families, the proposed approach seeks to leverage the strengths of each wavelet while mitigating the loss of signal features in each wavelet type. This technique holds significant promise for improving the accuracy and precision of medical diagnoses and rehabilitation plans.