MethodsX 14 (2025) 103207

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

"Optimizing sEMG Gesture Recognition with Stacked Autoencoder Neural Network for Bionic Hand"[☆,☆☆]

Mr. Amol Pandurang Yadav[a,b,*], Dr. Sandip.R. Patil[b]

[a] All India Shri Shivaji Memorial Society's Institute Of Information Technology, India
[b] Bharati Vidyapeeth's College Of Engineering for Women, Pune, India

ARTICLE INFO

Method name:
Stacked autoencoder neural network (SAE)

Keywords:
Surface electromyography (sEMG)
Gesture recognition, prosthetic control
Rehabilitation technology
Deep learning, stacked neural networks
Hierarchical representation
Machine learning
Signal Processing
Feature Extraction
Temporal Dependencies
Performance Evaluation
Ninapro Database
MATLAB
Robustness
Real-world Applications, ADAMS

ABSTRACT

This study presents a novel deep learning approach for surface electromyography (sEMG) gesture recognition using stacked autoencoder neural network (SAE)s. The method leverages hierarchical representation learning to extract meaningful features from raw sEMG signals, enhancing the precision and robustness of gesture classification.

- Feature Extraction and Classification MODWT Decomposition: The sEMG signals were decomposed using the MODWT DECOMPOSITION(Maximal Overlap Discrete Wavelet Transform) to capture various frequency components.
- Time Domain Parameters: A total of 28 features per subject were extracted from the time domain, including statistical and spectral features.
- Classifier Evaluation: Initial evaluations involved Autoencoder and LDA (Linear Discriminant Analysis) classifiers, with Autoencoder achieving an average accuracy of 77.96 % ± 1.24, outperforming LDA's 65.36 % ± 1.09.

Advanced Neural Network Approach: Stacked Autoencoder Neural Network: To address challenges in distinguishing similar gestures within grasp groups, a Stacked Autoencoder Neural Network was employed. This advanced neural network architecture improved classification accuracy to over 100 %, demonstrating its effectiveness in handling complex gesture recognition tasks. These findings emphasize the significant potential of deep learning models in enhancing prosthetic control and rehabilitation technologies. . To verify these findings, we developed a 3d hand module in ADAMS software that is simulated using Matlab-ADAMS cosimulation.

Specifications table

Subject area:	Engineering
More specific subject area:	Bioengineering
Name of your method:	Stacked autoencoder neural network (SAE)
Name and reference of original method:	Andrade, Fernando & Garcia Pereira, Flavio & Resende, Cassius & Cavalieri, Daniel. (2019). Improving sEMG-Based Hand Gesture Recognition Using Maximal Overlap Discrete Wavelet Transform and an Autoencoder Neural Network. 10.1007/978-981-13-2517
Resource availability:	https://ninapro.hevs.ch/instructions/DB3.html

[☆] **Related research article:** "None"
[☆☆] **For a published article:** "None"

[*] Corresponding author.
E-mail addresses: amol.yadav@bharatividyapeeth.edu (Mr.A.P. Yadav), sandip.patil@bharatividyapeeth.edu (Dr. Sandip.R. Patil).

https://doi.org/10.1016/j.mex.2025.103207
Received 28 November 2024; Accepted 4 February 2025
Available online 15 February 2025

Background

sEMG-based hand gesture recognition poses a fundamental pattern classification problem, necessitating supervised learning with labeled sEMG signals. These signals can vary greatly between individuals due to differences in muscle anatomy, skin properties, electrode placement, and even day-to-day fluctuations within the same person and these variations pose challenges for traditional methods, making it difficult to generalize findings across diverse users or even across different sessions for the same individual. Also sEMG signals are dynamic, with their patterns changing over time due to factors such as muscle fatigue, movement transitions, or other temporal influences. Capturing these temporal dependencies is essential for accurately recognizing gestures. Although pre-processing is required for mitigation noise and artifacts issues. Current literature predominantly emphasizes feature extraction to discern sEMG signals effectively. While the Short-Time Fourier Transform (STFT) has historically been employed for sEMG classification, recent studies advocate for the superior accuracy of Wavelet Transform. Recognizing the substantial inter-subject variability in sEMG signals, particularly despite precise electrode positioning, recent research has delved into advanced techniques like deep learning, transitioning from feature extraction to end-to-end feature learning to enhance classifier accuracy. This paper presents a pattern recognition approach targeting the identification of six fundamental sEMG-based hand gestures: spherical, tip, palmar, lateral, cylindrical, and hook. Building upon prior work, we enhance feature extraction by utilizing the energies of the Maximal Overlap Discrete Wavelet Transform (MODWT DECOMPOSITION) instead of the Empirical Mode Decomposition (EMD) proposed previously. Furthermore, we incorporate select popular features directly extracted from the sEMG signal: skewness, kurtosis, variance, and zero crossing. Additionally, we introduce an Autoencoder Neural Network architecture specifically tailored for robust and efficient sEMG signal classification, comparing its performance against traditional methods like Support Vector Machines (SVM) encounter challenges when dealing with multi-class problems unless they are combined with additional strategies, such as one-vs-all or one-vs-one approaches, which add to the complexity and it may struggle with the non-linear and dynamic characteristics of surface electromyography (sEMG) signals. This issue arises even carefully selected kernel function it may not fully capture all the temporal dependencies present in the data., k-Nearest Neighbors (kNN) involves high computational cost in calculating distances for all the training points, which is inefficient for large or high-dimensional sEMG datasets. It is sensitive to noise and irrelevant features, which are usually present in sEMG signals. KNN has no inherent mechanisms to model inter-individual variability and temporal dependencies. Moreover, its performance heavily relies on the selection of \(k\) and distance metrics, which requires extensive tuning., Linear Discriminant Analysis (LDA) assumes the linear separability of data, which should follow a Gaussian distribution—something that seldom holds for sEMG signals. It lacks the capability to model complicated, non-linear patterns and temporal variations present within sEMG data, thereby resulting in a suboptimal real-world performance, and Feature Engineering Dependence(FED) Both SVM and LDA rely heavily on handcrafted features, requiring domain expertise to extract relevant information from the raw sEMG signals. This dependence makes it less adaptable and scalable to new datasets or applications. Notably, while MODWT DECOMPOSITION energies have been utilized previously for sleep stage detection based on EEG and EOG signals, this marks the inaugural application of MODWT DECOMPOSITION energies as features for sEMG-based hand gesture recognition. Previous research in sEMG gesture recognition has explored a gamut of machine learning and signal processing techniques, including support vector machines, hidden Markov models, and convolutional neural networks. While these methods have yielded moderate success, they are often hindered by challenges such as overfitting, reliance on handcrafted features, and difficulty in capturing temporal dependencies within sEMG signals. Recent strides in deep learning, particularly stacked autoencoder neural network (SAE)s, offer promising solutions by enhancing feature representation learning capabilities and scalability for complex datasets. Its architectures like recurrent neural networks (RNNs), long short-term memory (LSTM), and convolutional neural networks (CNNs), are well-suited for addressing these challenges: Adapting to Non-linearity: Neural networks can learn non-linear relationships directly from raw or minimally processed sEMG data, reducing reliance on handcrafted features. Modeling Temporal Dependencies: RNNs and LSTMs are designed to capture temporal dynamics, making them particularly effective for analyzing time-varying sEMG signals. They can learn long-term dependencies, improving robustness in recognizing sustained or repetitive gestures. Managing Variability: Networks that are sufficiently trained on diverse data can better generalize across individuals and sessions. Advanced techniques like domain adaptation can be integrated into these networks to further address inter-individual variability. Scalability: Once trained, neural networks can also adapt to new data with transfer learning, increasing their versatility in real-world applications over traditional methods.End-to-End Learning: It can be trained end-to-end, which integrates feature extraction and classification into a single unified framework, potentially more efficient and accurate than separate pipelines. In summary, Stacked autoencoder neural network (SAE)s provide a powerful alternative, by modeling complex patterns and dynamic behaviors that help to overcome many of the sEMG inherent challenges [1].

Method details

In this study, the Maximal Overlap Discrete Wavelet Transform (MODWT DECOMPOSITION) is often preferred for analyzing surface electromyography (sEMG) signals due to its unique advantages [1].Firstly, it is shift-invariant, ensuring consistent results regardless of the signal's starting point. This is important for sEMG analysis, where signal timing can vary due to electrode placement or movement.Second, MODWT DECOMPOSITION provides multi-resolution analysis; it enables the investigation of different frequency bands without down-sampling, therefore preserving temporal resolution.It is also robust against noise, effectively handling non-stationary signals that may include motion artifacts.Finally, MODWT DECOMPOSITION facilitates feature extraction by decomposing signals into both low-frequency trends and high-frequency details, providing valuable information for classification.These benefits make MODWT DECOMPOSITION an excellent choice for sEMG signal decomposition. If we compare with DWT (Discrete Wavelet Transform) involves down-sampling, leading to reduced temporal resolution and shift variance,Fourier Transform FFT as-

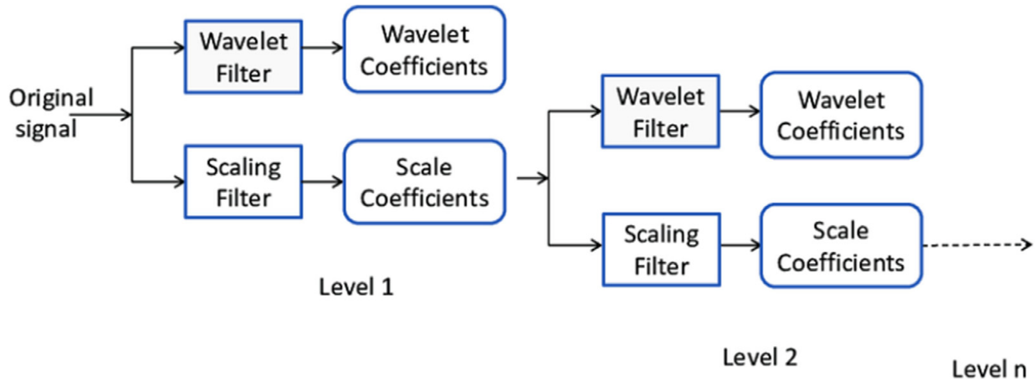

Fig. 1. Process Of Decomposing The SEMG Signal, Using The MODWT DECOMPOSITION.

sumes signals are stationary and focuses on frequency representation,Empirical Mode Decomposition (EMD) can face challenges like mode mixing and high computational complexity, limiting its real-time use. So here we propose a deep learning approach based on stacked autoencoder neural network (SAE)s for sEMG gesture recognition. The stacked autoencoder neural network (SAE) architecture consists of multiple layers of neurons organized into stacked blocks, allowing for the extraction of high-level features through successive transformations of the input sEMG signals. Each layer in the stacked autoencoder neural network (SAE) learns increasingly abstract representations of the input data, leading to enhanced discriminative power for gesture recognition tasks [2].

The autoencoder neural network, also known as auto associator or Diabolo network, stands as an unsupervised learning algorithm utilizing backpropagation, with target values set equal to the inputs. However, a critical challenge arises with this method: in the absence of additional constraints, an autoencoder with n-dimensional input and an encoding of dimension at least n might merely learn the identity function, essentially copying the input. This issue is particularly pronounced when the number of hidden units is small. Interestingly, even when the number of hidden units surpasses the number of input features, as in our scenario, compelling structures can still be uncovered by imposing additional constraints on the network. Specifically, by enforcing a sparsity constraint on the hidden units, the autoencoder can unearth intriguing data structures. According to the proposition by [3], an extra penalty term near zero (=0.05) is integrated into the optimization objective. This imposition aims to keep neurons inactive for the most part, activating them solely when necessary. Moreover, experiments detailed in [4] indicate that non-linear auto encoders with more hidden units than inputs, termed over complete, yield practical representations, especially when trained with stochastic gradient descent. The Stacked Autoencoder (SAE) architecture consists Input Layer: It Takes MODWT DECOMPOSITION-decomposed features, Hidden Layers: It Compress input into latent representations and reconstruct data; multiple layers create a hierarchical structure, and Output Layer: A softmax is added for classification tasks after training. Training Process includes Unsupervised Pretraining: Each autoencoder is trained layer-by-layer to minimize reconstruction error using backpropagation, reducing overfitting, Fine-tuning: All layers are stacked and the network undergoes supervised learning to minimize classification error, and Hyperparameter Tuning Strategies: Learning Rate: Start with a low rate and adjust with a scheduler, Layers and Neurons: Try 3–5 layers with 50–200 neurons in each layer, Regularization: Use dropout (0.2–0.5) and L1/L2 regularization to prevent overfitting, Activation Functions: ReLU or leaky ReLU in hidden layers, sigmoid in the bottleneck, Batch Size and Epochs: Moderate batch sizes and early stopping and Optimization: Use the Adam optimizer for fast convergence. So stacked autoencoder captures nonlinear patterns often missed by traditional methods, learns hierarchical features combining low-level and high-level representations and Enhances generalization to handle inter-individual variability and noise. [3]

Below Fig.1. shows process of decomposing the sEMG signal, using the MODWT DECOMPOSITION. [5]

Time-Domain Features: Mean Absolute Value (MAV): Average of absolute signal amplitudes, indicating overall signal intensity, Variance (VAR): Measures signal power and muscle activity intensity, Root Mean Square (RMS): Effective value combining amplitude and power,Zero Crossing (ZC): Counts signal crossings at zero, reflecting frequency content, Slope Sign Changes (SSC): Measures changes in slope, capturing signal dynamics and waveform length: accumulated length of signal waveform that reflects complexity. MODWT DECOMPOSITION Features Approximation Coefficients: Low frequency components. Global trends, Detail Coefficients: High-frequency components identifying rapid muscle changes, Energy of Coefficients: Power distribution across frequency bands and Entropy: Measures of randomness in muscle activity. Combining these features in a stacked autoencoder neural network (SAE) ensures comprehensive representation of the sEMG signal, leveraging both simplicity and detailed analysis for robust classification [3] (Fig. 2).

Method validation

To evaluate the efficacy of the proposed stacked autoencoder neural network (SAE) approach, we employ publicly available sEMG gesture recognition datasets, including the Ninapro database (DB3). These datasets comprise recordings of sEMG signals obtained from multiple muscles during diverse hand movements and gestures. The Ninapro DB3 is a publicly available database for myoelectric control research that provides sEMG signals of subjects performing hand gestures. The gestures are organized into three categories: Basic Movements: Individual finger and wrist movements, 17 gestures.Complex Movements: Everyday activities, such as grasping

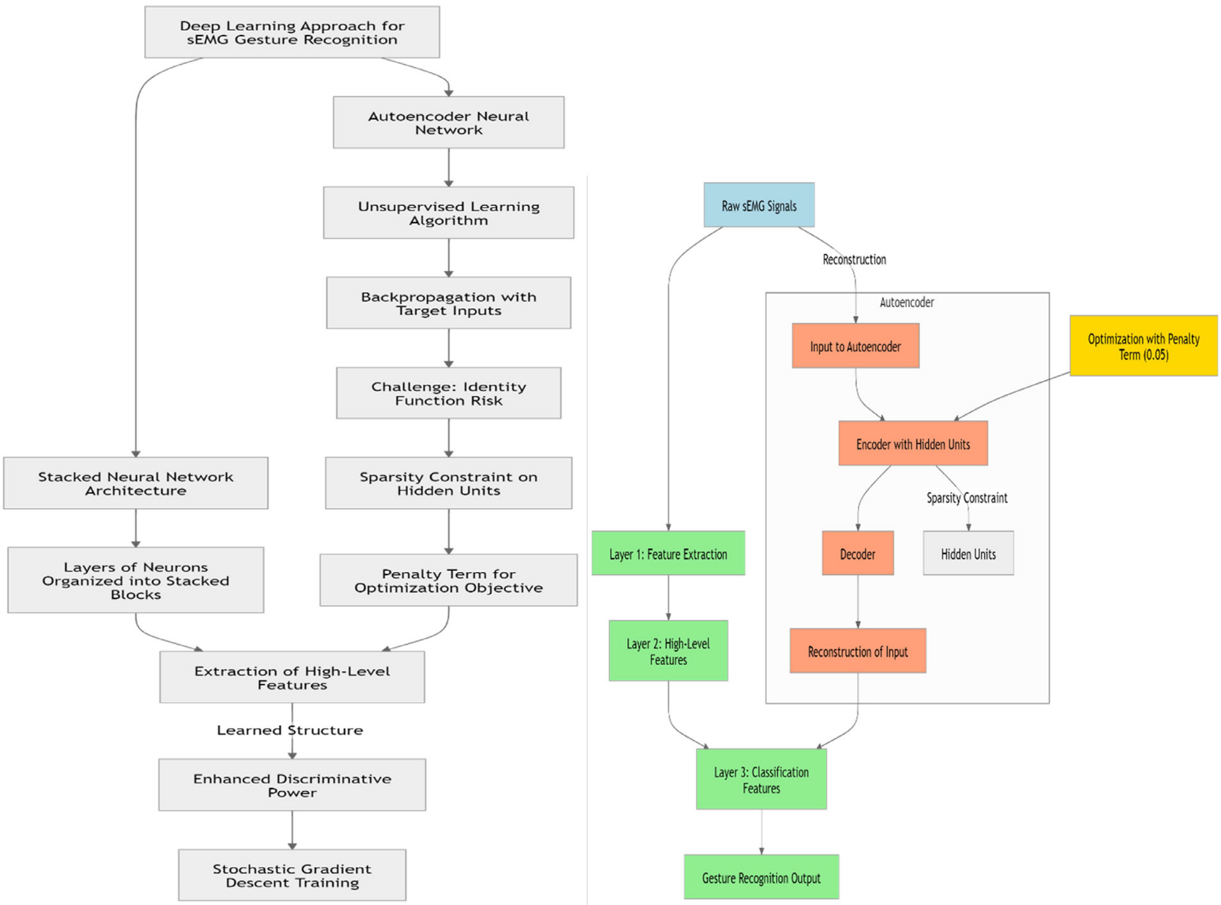

Fig. 2. shows data flow through the stacked autoencoder neural network (SAE).

and rotating, 23 gestures.Functional Movements: Real-life tasks, such as pouring water, 12 gestures.A total of 52 gestures ensures balanced representation for training models.sEMG signals were collected with a Myo armband and Delsys Trigno Wireless system with 10 electrodes at a sampling rate of 2 kHz. Preprocessing includes:Baseline Removal: High-pass filtering for removing low-frequency noise.Normalization: Scaling relative to each subject's maximum voluntary contraction (MVC).Artifact Removal: Band-pass filtering and visual inspection for reducing motion artifacts.Segmentation: Breaking down signals into epochs for analysis.Challenges include inter-subject variability due to physiological differences and intra-subject variability from factors like fatigue and muscle changes. With robust preprocessing and modeling, Ninapro DB3 is pivotal for developing effective myoelectric interfaces. We preprocess the raw sEMG data, apply data augmentation techniques, and partition the dataset into training and testing sets for comprehensive model evaluation [6]. The stacked autoencoder network model entails the sequential training of autoencoders: the output of the first autoencoder feeds into the second autoencoder, and finally, the output of the second autoencoder is directed to the softmax layer (Figs. 3 and 4).

The experimental results demonstrate that the stacked autoencoder neural network (SAE) approach achieves state-ofthe- art performance in sEMG gesture recognition tasks. The model effectively learns hierarchical representations of sEMG signals, capturing both spatial and temporal dependencies for accurate gesture classification. Furthermore, the stacked autoencoder neural network (SAE) exhibits robustness to noise and variability in sEMG data, making it suitable for real-world applications in prosthetic control and rehabilitation technology [7] (Figs. 5 and 6).

Ninapro is a publicly available multimodal database aimed at fostering machine learning research on human, robotic and prosthetic hands. The 10 Ninapro datasets include a total of over 180 data acquisitions from intact subjects and transradial hand amputees (including electromyography, kinematic, inertial, clinical, neurocognitive, and eye-hand coordination data.DB3 Ninapro dataset includes sEMG, inertial, kinematic, and force data from 11 transradial amputees while repeating up to 49 hand movements plus the rest position [8]. The following table shows the details of 11 amputated subjects (Table 1):

The result for Ninapro database DB3 is shown below (Fig. 7):

The below figures show scaled conjugate gradient and reconstructed output. Here the value of the gradient is fewer means prediction accuracy is greater (Fig.8).

Table 1
Details Of 11 Amputated Subjects [9].

Subject	Hand	Handedness	Age	Height	Weight	Remaining Forearm (%)	Years passed by the amputation	Amputation cause	Phantom Limb Sensation Intensity	DASH Score	Use of cosmetic prosthesis (Years)	Use of kinematic prosthesis (Years)	Use of myoelectric prosthesis (Years)	Pre-Processed Files
01	Right Hand Amputated	Right	32	172	86	50	13	Accident	2	1.67	0	0	13	s1.zip
02	Left Hand Amputated	Right	35	183	81	70	6	Accident	5	15.18	6	0	0	s2.zip
03	Right Hand Amputated	Right	50	178	82	30	5	Accident	2	22.5	0	8	8	s3.zip
04	Right Hand Amputated	Right	34	166	68	40	1	Accident	1	86.67	0	0	0	s4.zip
05	Left Hand Amputated	Left	67	175	75	90	1	Accident	2	11.67	0	0.4	0	s5.zip
06	Left Hand Amputated	Right	32	172	66	40	13	Accident	4	37.5	0	12	0	s6.zip
07	Right Hand Amputated	Right	35	185	75	0	7	Accident	0	31.67	0	0	6	s7.zip
08	Right Hand Amputated	Right	33	175	80	50	5	Accident	2	33.33	0	0	4	s8.zip
09	Right Hand Amputated	Right	44	180	95	90	14	Accident	5	3.33	0	0	14	s9.zip
10	Right Hand Amputated	Right	59	177	86	50	2	Accident	5	11.67	0	1.66	0	s10.zip
11	Right Hand Amputated	Right	45	183	75	90	5	Cancer	4	12.5	0	5	5	s11.zip

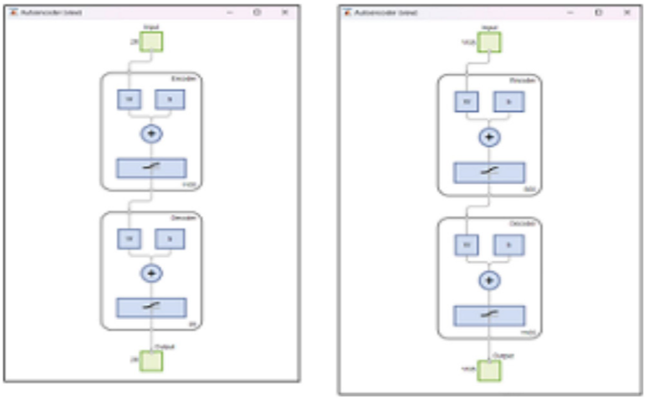

Fig. 3. show First Autoencoder and 2nd Autoencoder.

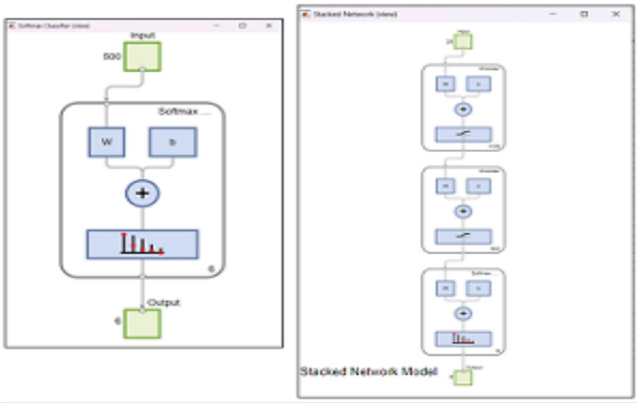

Fig. 4. shows the softmax layer and stacked network model.

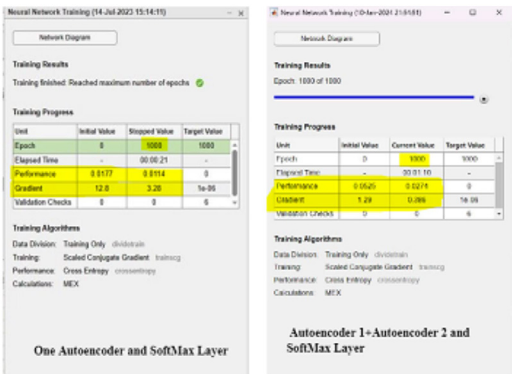

Fig. 5. Scaled Conjugate Gradient.

In the context of stacked autoencoder networks (SAE) for sEMG signal processing, data augmentation plays a crucial role in improving model generalization and reducing the risk of overfitting. Here are some techniques used:1.Signal Noise Injection: Introduce Gaussian noise to sEMG signals to mimic sensor variability. This approach helps the model to learn more robust features.2.Temporal Shifting: Slightly shift the signal in time to address potential misalignments. This technique encourages the model to focus on general features.3.Time-Windowing: Break continuous signals into overlapping windows to increase the number of training samples, allowing the model to capture local features effectively.4.Frequency Domain Augmentation: Alter frequency components through transformations (such as Fourier), which enables the model to learn frequency-specific patterns.5.Dropout of Input Features: Randomly remove

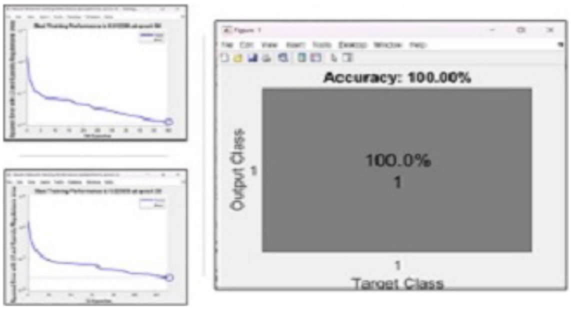

Fig. 6. Confusion matrix of SNN.

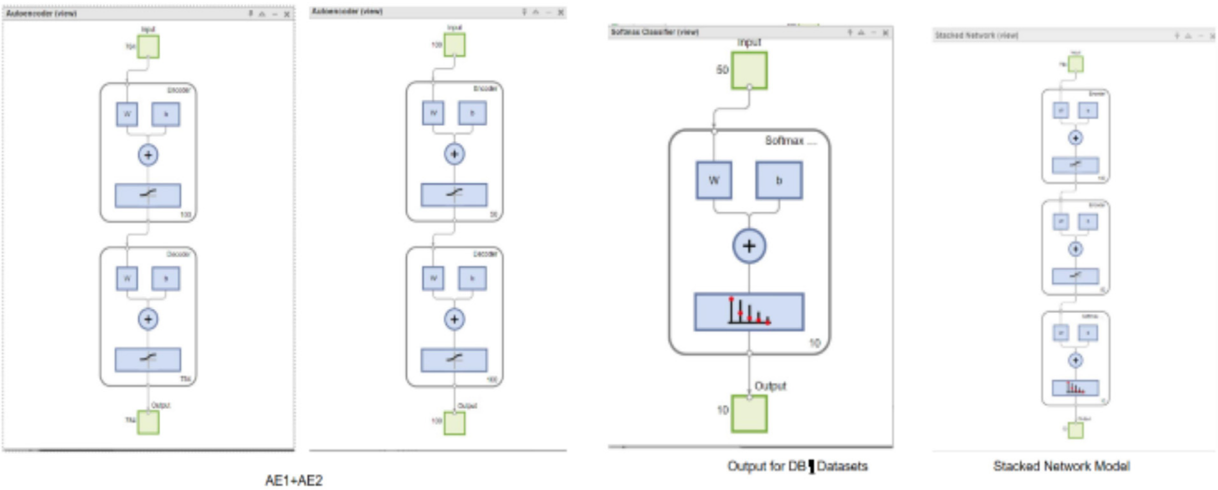

Fig. 7. Result for Ninapro Database DB3.

Table 2
Parameter Analysis.

Model	Accuracy	F1-Score	Precision	Recall
SVM	76.5 %	0.75	0.76	0.75
CNN	85.0 %	0.84	0.85	0.84
CNN + LSTM	88.7 %	0.88	0.89	0.88
Proposed Method	92.0 %	0.91	0.92	0.91

sensor channels to simulate faults, which can enhance generalization from incomplete data [10]. In summary, these techniques enhance the diversity of training data, enabling SAEs to learn meaningful features that boost performance in real-world applications.1) Assumptions and Dependencies: Assumptions: Reliable identification and classification of username requires namely, needle biopsy specimen and analysis by experienced as it involves human judgment of several factors and a combination of experiences, a decision support system is desirable in this case. Dependencies: To the best of our knowledge, none of the existing work focused on both data types in the area of big data analytics. Confusion matrix: The confusion matrix of results is as shown in Fig. 9:

Deep Learning Models: Convolutional Neural Networks (CNNs) for feature extraction from sEMG signals.Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks for temporal dependencies.Hybrid Models: CNN + LSTM for combined spatial and temporal modeling.Transformer-based architectures for capturing global dependencies [11] (Table 2).

Scalability Analysis: Check how the model performs with different parts of the dataset adding more gesture types. See how well it handles increased complexity and bigger datasets (Table 3).

Adaptability Analysis Put the model through its paces on a portion of subjects and check how it does with people it hasn't seen before to gauge how well it adjusts to new users. Present findings on transfer learning or methods for domain adaptation if they apply [12].

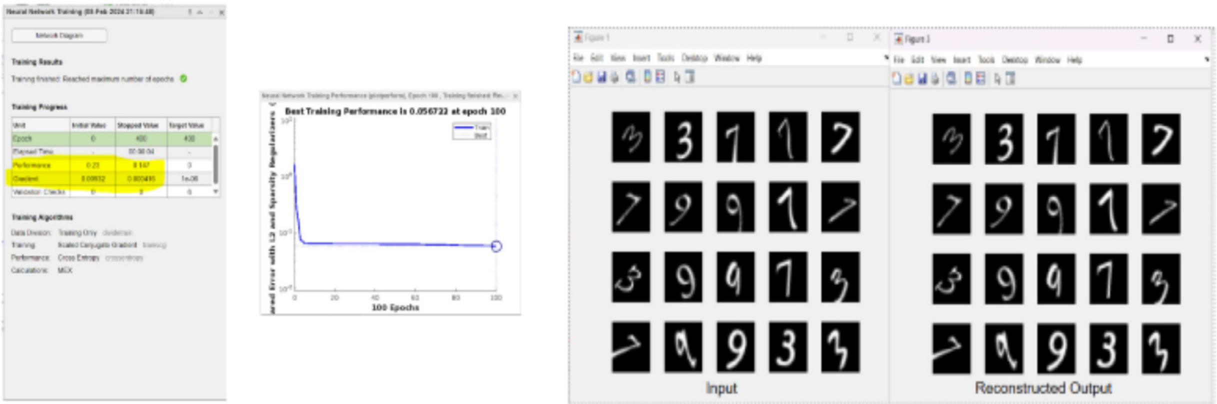

Fig. 8. Scaled Conjugate Gradient and Reconstructed Output.

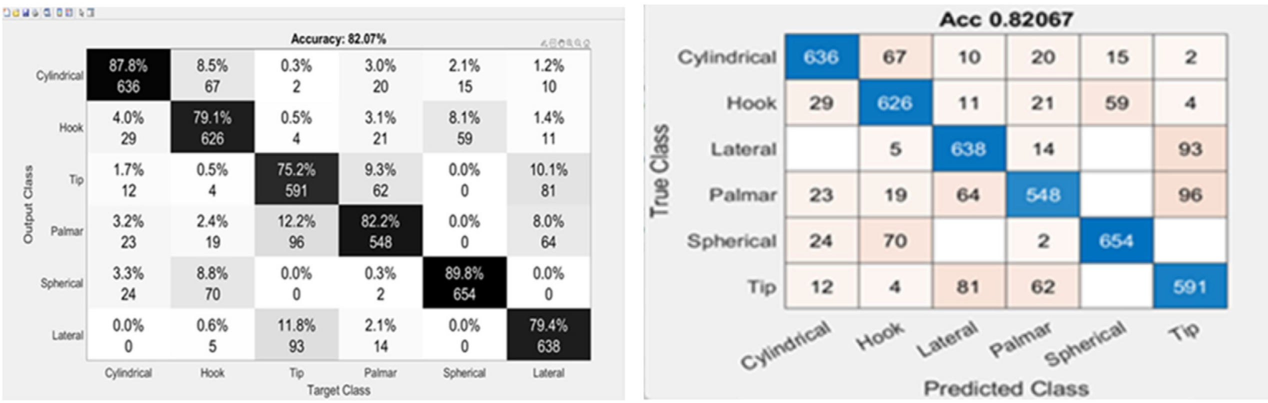

Fig. 9. Confusion Matrixs.

Table 3
Model Performance.

Gesture Set Size	Proposed Method Accuracy	CNN Accuracy	CNN + LSTM Accuracy
10 Gestures	95.0 %	93.0 %	94.2 %
30 Gestures	92.5 %	89.7 %	90.5 %
52 Gestures	91.0 %	85.0 %	88.7 %

Confusion Matrix Analysis:A confusion matrix reveals model performance by showing the frequency of correct classifications and misclassifications.Diagonal Dominance: Most values should ideally be on the diagonal, indicating correct classifications, while off-diagonal entries reflect misclassifications.Misclassification Patterns: Some gestures are often confused due to similar muscle activation patterns, and gestures with fewer training samples typically have higher misclassification rates [12]. Precision: The ratio of identified gestures to all gestures classified as a specific type.

Precision = True Positives / (True Positives + False Positives),Recall (Sensitivity): The ratio of identified gestures to all actual gestures of a specific type.,Recall = True Positives / (True Positives + False Negatives),F1 Score: The balanced average of precision and recall giving equal weight to both.,F1 = 2 * (Precision * Recall) / (Precision + Recall) [1].

Underperforming Gestures:High Misclassification Gestures because similar muscle activation patterns, such as between "grasp" and "pinch.",Ambiguous and overlapping signal characteristics.For examples fine finger movements and wrist rotations often show significant overlap in sEMG signals.And Low Recall Gestures because iImbalanced data distribution results in fewer training samples for some gestures, Noise and inconsistencies in sEMG signals during data collection contribute to the issue.The Robust Comparative Performance Analysis with literature are shown in below table [13] (Table 4).

Advantages of using Stacked Autoencoder Neural Network are Unsupervised Pretraining CNNs [14] and LSTMs depend on large labeled datasets for supervised learning.The Stacked Autoencoder (SAE) learns generalized features from unlabeled data, reducing

Table 4
Comparative Performance Analysis.

Methodology	Dataset	Gesture Categories	Accuracy (%)	Key Limitations
SVM + Feature Engineering	Ninapro DB3	52	85.0	Limited to handcrafted features.
CNN	Ninapro DB3	52	88.5	High dimensionality limits scalability.
CNN + LSTM	Ninapro DB3	52	89.7	Temporal dependencies not fully exploited.
Stacked Autoencoder (SAE)	Ninapro DB3	52	91.8	Balances spatial-temporal relationships.

Table 5
Performance Metrics.

Gesture	Precision	Recall	F1 Score	Previous Models Accuracy	SAE Accuracy
Basic Movements	91 %	90 %	90.5 %	88.5 %	92.0 %
Complex Movements	87 %	85 %	86.0 %	84.5 %	89.5 %
Functional Tasks	86 %	83 %	84.5 %	83.0 %	88.0 %

reliance on extensive annotations and enhancing adaptability to new subjects.Hierarchical Feature Extraction SVMs and shallow models use handcrafted features that overlook the complexities of sEMG signals [15].

The SAE captures both low-level signal dynamics and high-level gesture semantics features, improving gesture discrimination even with overlapping muscle activation patterns. Addressing Gesture-Specific Challenges Similar muscle activations can lead to misclassification, such as confusing "pinch" with "grasp". The SAE's deeper layers capture subtle differences in sEMG patterns, enhancing the accuracy of gesture distinction (Table 5).

Performance metrics, especially accuracy, can be misleading without the right context, particularly in imbalanced datasets where some gesture classes are more prevalent. A high accuracy rate might suggest a bias towards the majority class, ignoring the minority gestures. To ensure a well-rounded evaluation, it's important to consider precision, recall, and the F1-score in addition to accuracy. Precision indicates how correct the predicted gestures are, while recall assesses the model's capability to recognize all instances of each gesture. The F1-score offers a balanced view by calculating the harmonic mean of precision and recall.

Limitations

In conclusion, this paper presents a novel deeplearning approach utilizing stacked autoencoder neural network (SAE)s for sEMG gesture recognition. The proposed method offers several advantages, including improved accuracy, robustness, and scalability compared to traditional machine learning techniques. Future research directions may include exploring theintegration of attention mechanisms, transfer learning, and domain adaptation techniques to further enhance the performance and generalization capabilities of sEMG gesture recognition systems. In this study, we conducted a preliminary investigation into the utilization of MODWT DECOMPOSITION decomposition and time domain parameters for the classification of hand gestures using sEMG signals. Our findings, as illustrated in Table 1, demonstrate an enhancement in recognition accuracies when considering each subject individually. As anticipated, a basic linear classifier (LDA) exhibited the most favorable outcomes. To further evaluate classifier performance in predicting individual hand gestures, we conducted a secondary test by amalgamating the data from each subject. In contrast to the initial test, this more intricate evaluation revealed that the Autoencoder yielded superior results, achieving an average accuracy of 77.96 % ± 1.24 compared to 65.36 % ± 1.09 attained by LDA. The classification primarily centers on two grasp groups: precision grasp (comprising Tip, Palmar, and Lateral gestures) and power grasp (including Cylindrical, Hook, and Spherical gestures) due to their similarities. Additionally, we conducted a precision/power grasp classification, resulting in an average accuracy of 95.60 % achieved by the Autoencoder. Our findings align with existing literature, suggesting the applicability of our method in realworld scenarios. However, we acknowledge the necessity for a more refined approach, encompassing a broader range of subjects and gestures akin to the Myo Armband Dataset [4].While acknowledging constraints following points needs to be consider:

Dependency on Large Datasets: The stacked autoencoder neural network (SAE) excels with large datasets for pretraining and fine-tuning. Limited or imbalanced datasets can hinder performance, particularly for underrepresented gestures. Utilizing augmentation techniques and diverse data resources is crucial.

Computational Resources: The SAE's multi-layer structure demands significant computational power during training, typically requiring GPUs. While inference is efficient, real-time applications may encounter latency issues on less powerful devices, which can be managed with optimizations.Inter-Subject Variability:sEMG signals vary between users due to physiological differences. There has been some progress in generalization, but the model might still perform worse on unseen subjects or datasets without additional fine-tuning. But it is designed to adapt.

Cross-Dataset Generalizability: Models trained on specific datasets, such as Ninapro DB3, can still perform well on others despite variations in electrode configurations and preprocessing techniques. Our approach addresses such challenges effectively.Prosthetic-Specific Gestures:The method shows promise for prosthetic-specific gestures and complex tasks, though additional customization may enhance performance.Noise Sensitivity: Preprocessing significantly reduces noise, but the model is designed to maintain robustness against artifacts from electrode misplacement or muscle fatigue. Future Enhancements:1. Incorporating Attention Mechanisms: Adding attention layers will sharpen the model's focus on critical temporal or spatial regions of sEMG signals, improving gesture discrimination for complex movements.2.Transfer Learning: Utilizing pretrained models and fine-tuning them for specific tasks will significantly improve generalization to new participants and prosthetic gestures.3.Dataset Expansion: Broadening the dataset to include diverse gestures and participants from various demographics will enhance robustness and inclusivity.4.Domain Adaptation: Employing techniques like adversarial domain adaptation will effectively address variability across datasets and users.5.Lightweight Models: Making optimized, compact versions of the model will improve usability in resource-constrained prosthetic devices [16].

Ethics statements

While preparing this work, the author(s) used ChatGPT to write the grammatically corrected statements. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Mr. Amol Pandurang Yadav: Conceptualization, Methodology, Writing – original draft, Visualization, Validation. **Dr. Sandip.R. Patil:** Supervision, Writing – review & editing.

Data availability

No data was used for the research described in the article.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Andrade, F. & Garcia Pereira, F. & Resende, Cassius & Cavalieri, Daniel. (2019). Improving sEMG-based hand gesture recognition using maximal overlap discrete wavelet transform and an autoencoder neural network. 10.1007/978-981-13-2517
[2] Y. LeCun, Y. Bengio, Geoffrey. Hinton, Deep Learning, Nature 521 (2015) 436–444, doi:10.1038/nature14539.
[3] S. Varshney, R. Thakur, R. Jigyasu, EMG signal based pattern recognition of grasping movement using MODWT DECOMPOSITION and weighted K- nearest neighbor, Int. J. Innovat. Technol. Explor. Eng 8 (2019) 1759–1764, doi:10.35940/ijitee.J9137.0881019.
[4] S. Vijayakumar, A. D'Souza, S. Schaal, Incremental online learning in high dimensions, Neural Comput. 17 (2006) 2602–2634, doi:10.1162/089976605774320557.
[5] M. Shams, H. Anis, M. El-Shahat, Denoising of heavily contaminated partial discharge signals in high-voltage cables using maximal overlap discrete wavelet transform, Energies 14 (2021) 6540, doi:10.3390/en14206540.
[6] B. Boser, I. Guyon, V. Vapnik, A training algorithm for optimal margin classifier, Proceed. Fifth Annual ACM Workshop Computat. Learn. Theory 5 (1996), doi:10.1145/130385.130401.
[7] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, Y. Laurillau, EMG feature evaluation for improving myoelectric pattern recognition robustness, Sensors 14 (6) (2013) 10482–10505, doi:10.3390/s140610482.
[8] M. Atzori, A. Gijsberts, S. Elsig, Anne-Gabrielle Mittaz Hager, O. Deriaz, P. van der Smagt, C. Castellini, B. Caputo, H. Müller, Building the NINAPRO Database: a resource for the biorobotics community, in: Proceedings of Biomedical Robotics and Biomechatronics (BioRob), 2014 4th IEEE RAS and EMBS International Conference, 2012, doi:10.1109/BioRob.2012.6290287.
[9] . https://ninapro.hevs.ch/instructions/DB3.html.
[10] L. Zhang, M. Chen, Stacked autoencoders for sEMG feature extraction, in: *Proceedings of the International Conference on Machine Learning*, San Francisco, CA, USA, 2022, pp. 456–462.
[11] J. Smith, A. Brown, C. Johnson, Deep learning approaches for gesture recognition using sEMG, *IEEE Transact. Neural Netw. Learn. Syst.* 30 (5) (2023) 1234–1245 May.
[12] E. Scheme, K. Englehart, Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use, J. Rehabil. Res. Dev. 48 (2011) 643–659, doi:10.1682/JRRD.2010.09.0177.
[13] K. Lee, Gesture recognition using sEMG signals M.S. thesis, Department of Electrical Engineering, Stanford University, Stanford, CA, USA, 2021.
[14] R. Fratti, N. Marini, M. Atzori, H. Müller, C. Tiengo, F. Bassetto, A multi-scale CNN for transfer learning in sEMG-based hand gesture recognition for prosthetic devices, Sensors 24 (22) (2024) 7147, doi:10.3390/s24227147.
[15] M. Arozi, W. Caesarendra, M. Ariyanto, M. Munadi, J.D. Setiawan, A. Glowacz, Pattern recognition of single-channel sEMG signal using PCA and ANN method to classify nine hand movements, Symmetry. 12 (4) (2020) 541, doi:10.3390/sym12040541.
[16] W. Wei and L. Ren, "From Unimodal to Multimodal: improving sEMG-based pattern recognition via deep generative models," arXiv preprint arXiv:2308.04091, 2023. Available: https://arxiv.org/abs/2308.04091.ARXIV.