Hand Gesture Classification Using Kolmogorov–Arnold Networks (KAN)

E.F Luque
Department of Software Engineer
La Salle University Arequipa
Arequipa
eluquem@ulasalle.edu.pe

Abstract

This paper explores the potential of Kolmogorov-Arnold Networks (KAN) as an alternative to traditional Multi-Layer Perceptrons (MLPs) for hand gesture classification. While MLPs are integral to many AI applications, they have limitations such as inefficiency in parameter usage and a lack of interpretability. KANs address these issues by introducing adaptive, univariate functions that serve as both weights and activation functions, enhancing both the efficiency and flexibility of the network. Experiments were conducted using surface electromyography (sEMG) data from 30 participants performing seven hand gestures. The data were processed through a series of steps, including signal filtering, windowing, and feature extraction. The results showed that KANs performed comparably to, or better than, traditional classifiers such as Neural Networks, Random Forest, and Extreme Gradient Boosting, particularly when larger window sizes were used. However, the training time and risk of overfitting remain challenges. Overall, KANs demonstrate promise for low-resource systems like prosthetics, warranting further exploration in real-world applications and with larger datasets.

1 Introduction

Various methodologies have been used for hand gesture recognition, with surface electromyogram (sEMG) data being particularly effective. sEMG signals, generated by muscle movements, are valuable in developing medical devices like prosthetics and rehabilitation tools(20). These signals provide key information for intelligent systems, often classified using machine learning (ML) methods such as XGBoost, Random Forest, k-nearest neighbors (kNN), and support vector machines (SVM)(21)(23) (22)(24). Features for classification are extracted from sEMG signals across time, frequency, and time–frequency domains. ML-based approaches have achieved accuracies from 70% to 95% (30)(31). In addition to traditional ML approaches, deep learning (DL) methods that are more complex variants of the artificial neural networks (ANNs) commonly used in ML have recently gained popularity for hand gesture classification. DL methods, especially convolutional neural networks (CNNs), have gained popularity for gesture classification due to their ability to automatically extract features. However, studies such as (1), (2), and (3) highlight that although these sophisticated models achieve high accuracy, their computational complexity and longer average processing times make them unsuitable for real-time applications, where rapid response is critical and resources are limited, simpler algorithms with lower computational complexity are often more practical. Various studies, such as (4), (5), (6), (25), and (26), have implemented ML algorithms that, while requiring lower computational power, still achieve relatively strong performance, making them suitable for real-time applications. Given these considerations, this study explores KANs as an alternative neural architecture for classifying hand gestures using sEMG signals. This makes KANs an option for real-time gesture classification.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

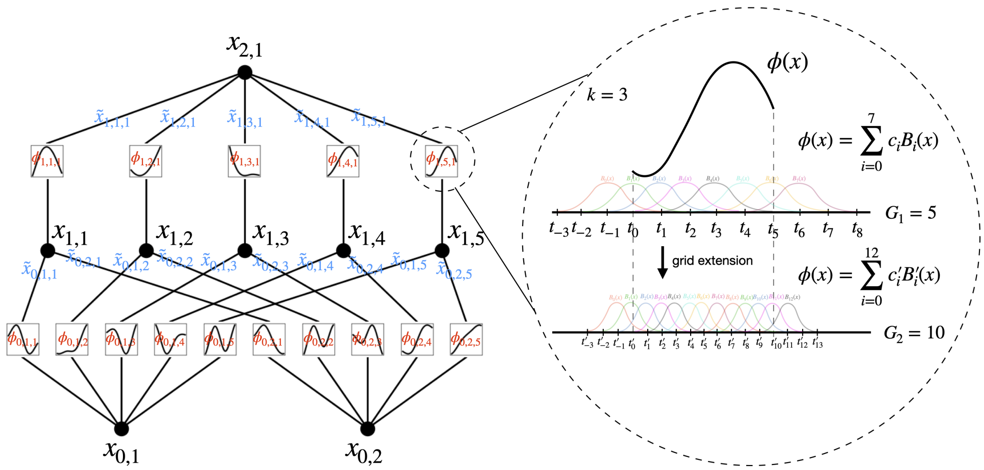

Figure 1: Left: Illustrates the notation for activations flowing through the network. Right: Depicts an activation function parameterized as a B-spline. The parameter k determines the type of B-spline used, enabling the activation function to transition between coarse-grained and fine-grained grids, providing flexibility in capturing data patterns.(8)

The structure of the paper is as follows: Section 2 reviews the key concepts of KANs and sEMG signal processing. Section 3 describes our proposed approach, and the configuration and evaluation of the KAN model. Section 4 presents the results, comparing KAN's performance with traditional machine learning models for sEMG. Finally, Section 5 summarizes the findings and explores future research directions.

2 Background

2.1 Kolmogorov-Arnold Networks(KAN)

Multi-layer perceptrons (MLPs) are widely used for approximating nonlinear functions(27). However, they come with notable limitations. In transformer architectures MLPs make up the bulk of non-embedding parameters, leading to potential inefficiencies(28)(29).

To fully appreciate the potential of KANs as an alternative, it's important to revisit the role of Multi-Layer Perceptrons (MLPs), which have traditionally served as the backbone of AI applications. MLPs are integral to AI, structuring computations through a series of layered transformations, which can be understood as:

$$f(x) = \sigma(W * x + B) \tag{1}$$

The key to training these networks is optimizing the weights (W). On the other hand, KAN redefines how activation functions σ are used. Unlike the static and non-learnable activation functions in MLPs, KAN utilizes univariate functions that serve a dual role as both weights and activation functions. These functions adapt and evolve as part of the learning process, which can be understood as:

$$f(x_1, x_2) = \phi_2(\phi_{2,1}(\phi_{1,1} * (x_1) + \phi_{1,2} * (x_2)) \tag{2}$$

Where: x_1 and x_2 are inputs, $\phi_{1,1}$ and $\phi_{1,2}$ are specific univariate functions for each input, combined and then processed through another function ϕ_2 in the subsequent layer. The innovations of KANs are clearly illustrated in Figure 1.

KANs introduces non-linearity before summing inputs and relocates activation functions to the edges of the network instead of embedding them within neurons. This shift allows for a more nuanced handling of individual features, making KANs both more intuitive and efficient. These innovations give KANs the potential to excel at complex and dynamic tasks, offering a fundamentally improved architecture that goes beyond incremental improvements seen in traditional models.

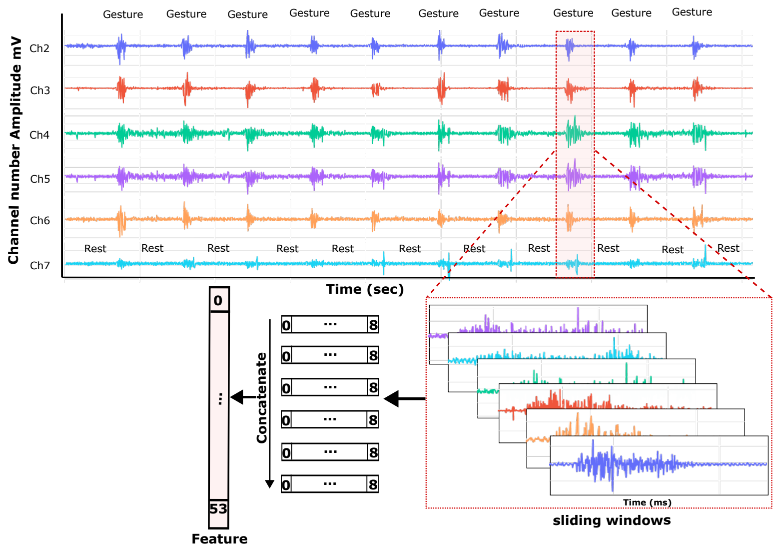

Figure 2: The schematic visualization of the feature extraction process involves 6 input channels, each representing multivariate time series data. Using a step size (s) of 50 ms and a window length (W) of various sizes (e.g., 200 ms), each time series generates a 9-dimensional feature vector, capturing essential characteristics of the data within the specified window.).

3 Proposal

3.1 sEMG Data Collection and Signal Filtering

We used publicly available sEMG data from (17), collected in accordance with the Helsinki Declaration and approved by Izmir Katip Celebi University (IKCU). The dataset consists of 30 healthy participants (15 male, 15 female, average age 22.37 ± 1.47 years) performing seven hand gestures. Signals were recorded from the flexor and extensor muscles of the right hand during 6-second gesture intervals, with 4-second rests, over five 74-second cycles, totaling 490 seconds per participant.

To reduce environmental noise, raw sEMG signals were filtered with a sixth-order Butterworth band-pass filter (5–500 Hz). The signals were segmented based on the steady-state muscle activity, excluding transition phases. Windowing was applied using rectangular sliding windows of varying lengths (50–200 ms) with a 25 ms shift, enabling optimized feature extraction and classification performance.

3.2 Feature extraction

According to (10)(30)(31), the most commonly extracted features from sEMG signals include Integrated EMG(**IEMG**), Mean Absolute Value(**MAV**)(15), Modified Mean Absolute Value(**MMAV1**)(12), Willison amplitude(**WAMP**)(13), Root Mean Square(**RMS**)(14), **V-Order**(**V3**)(15), Waveform Length(**WL**)(16), Average Amplitude Change(**AAC**), Slope Sign Changes(**SSC**). These nine features generate a feature vector, which can be formalized by accounting for the multiple channels of sEMG data. If the sEMG signal for M gestures is acquired using a device with C channels over time, and F features are adopted to describe each gesture, the feature vector of $m-th$ gesture D^m can be defined by the following equation:

$$D^m = [d^m_{1,1}, d^m_{1,2}, ..., d^m_{1,F}, d^m_{2,1}, d^m_{2,2}, ..., d^m_{2,F}, d^m_{c,1}, d^m_{c,2}, ..., d^m_{C,F}] \tag{3}$$

Where, $d^m_{c,f}$ represents the f-th feature of the c-th channel for the m-th gesture. This paper uses $M = 3, 7$ to evaluate the performance of the KAN architecture. Initially, we compare the performance using 3 gestures: rest, extension, and flexion. Subsequently, we extend the comparison to 7 gestures by adding ulnar deviation, radial deviation, and punch. Additionally, we use $C = 6$ channels and $F = 9$ features. The schematic visualization of the feature extraction is presented in Figure 2.

Table 1: Summary for the best hyperparameter tuning results of the machine learning models. For KAN model "Hidden layer": refers to the number of basis functions,"Grid": determines the interval over which the activation function operates, "k": This parameter is related to the type of B-spline used in the activation function

Model	3-Gestures	7-Gestures
XGBoost	Maximum tree's depth: 6 Learning rate: 0.01 Columns per tree: 0.9	Maximum tree's depth: 7 Learning rate: 0.01 Columns per tree: 0.9
Random Forest	Estimators: 14 Max depth: 21	Estimators: 34 Max depth: 11
Neural Networks	Activation function: "tanh" Alpha: 0.24 Hidden layer size: (8,)	Activation function: "tanh" Alpha: 0.24 Hidden layer size: (14,)
ExtraTree	Criterion: Gini Max depth: 11 Estimators: 10	Criterion: Gini Max depth: 12 Estimators: 36
KAN	Grid: 3 Hidden layer size: 17 k:3	Grid:3 Hidden layer size:17 k:3

Table 2: Classification scores for 3-gesture and 7-gesture groups improved with increasing window size, showing significant gains up to 200 ms. Beyond 200 ms, the performance improvements plateau, making 200 ms the optimal balance between accuracy and computational efficiency

	Windows	100			150			200		
Gestures	Model	Accuracy	Recal	F1	Accuracy	Recal	F1	Accuracy	Recal	F1
3	NN	0.7750	0.7434	0.744	0.832	0.8344	0.8331	0.8567	0.8444	0.8505
	RF	0.8112	0.8023	0.810	0.8598	0.8677	**0.8637**	0.8758	0.8923	0.8839
	XGB	0.8123	**0.821**	**0.818**	**0.8609**	0.8501	0.8554	**0.9012**	0.8979	**0.8995**
	ET	0.8023	0.7923	0.7923	0.8459	0.8504	0.8479	0.8858	0.8777	0.8817
	KAN	**0.8234**	0.8013	**0.8121**	**0.8602**	**0.8705**	**0.8653**	**0.8969**	**0.9005**	**0.8986**
7	NN	0.6745	0.6069	0.6069	0.8219	0.8214	0.8216	0.833	0.8422	0.8375
	RF	0.7323	0.7545	0.7432	0.8333	0.8298	0.8315	0.8923	**0.901**	**0.8966**
	XGB	0.7512	**0.7628**	0.7569	**0.8513**	0.8485	0.8498	**0.9001**	0.8874	0.8937
	ET	0.7301	0.7247	0.7273	0.84	0.8322	0.8360	0.8717	0.8699	0.8707
	KAN	**0.7554**	**0.7589**	**0.7571**	**0.8495**	**0.8542**	**0.8518**	0.8896	**0.8915**	**0,8905**

4 Preliminary Results and Conclusions

The extracted feature vectors are used as inputs for both the KAN architecture and classical classifiers to train and compare their performance. Four classification algorithms were selected for comparison: Neural Networks (NN), Random Forest (RF), Extreme Gradient Boosting (XGB), and Extra Trees (ET) (19), (18). When using machine learning models, key considerations include evaluation metrics such as F1-Score, Accuracy, and Recall, along with 5-fold cross-validation and hyperparameter tuning (18). Hyperparameters, which govern the training process, significantly impact model performance (19). To achieve optimal results, an extensive hyperparameter search was conducted for each model, including KANs, where precise tuning is also necessary. Table 1 summarizes the best hyperparameters obtained for each model. The classification results are presented in Table 2, with the highest scores in each row highlighted in bold.

Key insights from the evaluation of KAN show Table 2 that KAN delivers competitive performance in hand gesture classification, particularly with larger window sizes (e.g., 200 ms). Its accuracy and F1 scores match or surpass those of models like XGBoost and Random Forest, demonstrating KAN's ability to effectively capture complex patterns in gesture data, making it a strong candidate for precise classification tasks. Unlike resource-heavy models such as deep neural networks, KAN excels in scenarios where computational resources are limited. This makes it ideal for applications like prosthetics and embedded systems, where efficiency and low latency are critical. By Other hand, KAN faces challenges with long training times and overfitting when hyperparameters are not finely tuned, highlighting the need for further research to optimize its training process. Finally KAN's straightforward design offers advantages for rapid deployment and use in resource-constrained environments, making it a valuable tool despite its simplicity.

In conclusion, while KAN faces some training challenges, its performance and efficiency in gesture classification demonstrate its potential as a viable alternative to more complex models.

References

[1] Ozdemir, Mehmet Akif and Kisa, Deniz Hande and Guren, Onan and Onan, Aytug and Akan, Aydin: EMG based hand gesture recognition using deep learning, 2020 Medical Technologies Congress (TIPTEKNO).

[2] Rahimian, E., Zabihi, S., Atashzar, S. F., Asif, A., & Mohammadi, A. (2020). Surface EMG-based hand gesture recognition via hybrid and dilated deep neural network architectures for neurorobotic prostheses. Journal of Medical Robotics Research, 5(01n02), 2041001.

[3] Allard, U. C., Nougarou, F., Fall, C. L., Giguère, P., Gosselin, C., Laviolette, F., & Gosselin, B. (2016, October). A convolutional neural network for robotic arm guidance using sEMG based frequency-features. In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 2464-2470). IEEE.

[4] Song, W., Han, Q., Lin, Z., Yan, N., Luo, D., Liao, Y., ... & Bai, S. (2019). Design of a flexible wearable smart sEMG recorder integrated gradient boosting decision tree based hand gesture recognition. IEEE transactions on biomedical circuits and systems, 13(6), 1563-1574.

[5] Benatti, S., Farella, E., & Benini, L. (2014, September). Towards EMG control interface for smart garments. In Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program (pp. 163-170).

[6] Tavakoli, M., Benussi, C., Lopes, P. A., Osorio, L. B., & de Almeida, A. T. (2018). Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier. Biomedical Signal Processing and Control, 46, 121-130.

[7] Liu, X., Sacks, J., Zhang, M., Richardson, A. G., Lucas, T. H., & Van der Spiegel, J. (2016). The virtual trackpad: An electromyography-based, wireless, real-time, low-power, embedded hand-gesture-recognition system using an event-driven artificial neural network. IEEE Transactions on Circuits and Systems II: Express Briefs, 64(11), 1257-1261.

[8] Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., ... & Tegmark, M. (2024). Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756.

[9] Ozdemir, M. A., Kisa, D. H., Guren, O., Akan, A. (2022). Hand gesture classification using time–frequency images and transfer learning based on CNN. Biomedical Signal Processing and Control, 77, 103787.

[10] Phinyomark, A., Phukpattaranont, P., Limsakul, C. (2012). Feature reduction and selection for EMG signal classification. Expert systems with applications, 39(8), 7420-7431.

[11] Zardoshti-Kermani, M., Wheeler, B. C., Badie, K., Hashemi, R. M. (1995). EMG feature evaluation for movement control of upper extremity prostheses. IEEE Transactions on Rehabilitation Engineering, 3(4), 324-333.

[12] Sid'El Moctar, S. M., Rida, I., Boudaoud, S. (2023). Time-domain features for sEMG signal classification: A brief survey. JETSAN 2023.

[13] Núñez-Montoya, B., Añazco, M. V., Saravia-Avila, A., Loayza, F. R., Añazco, E. V., Teran, E. (2022, November). Supervised machine learning applied to non-invasive EMG signal classification for an anthropomorphic robotic hand. In 2022 IEEE ANDESCON (pp. 1-6). IEEE.

[14] Kim, K. S., Choi, H. H., Moon, C. S., Mun, C. W. (2011). Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions. Current applied physics, 11(3), 740-745.

[15] Zardoshti-Kermani, M., Wheeler, B. C., Badie, K., Hashemi, R. M. (1995). EMG feature evaluation for movement control of upper extremity prostheses. IEEE Transactions on Rehabilitation Engineering, 3(4), 324-333.

[16] Hudgins, B., Parker, P., Scott, R. N. (1993). A new strategy for multifunction myoelectric control. IEEE transactions on biomedical engineering, 40(1), 82-94.

[17] Ozdemir, M. A., Kisa, D. H., Guren, O., Akan, A. (2022). Dataset for multi-channel surface electromyography (sEMG) signals of hand gestures. Data in brief, 41, 107921.

[18] GuolinKe, Q. M., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst, 30, 52.

[19] Chen, T., Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

[20] Korošec, D. (2000). Parametric estimation of the continuous non-stationary spectrum and its dynamics in surface EMG studies. International journal of medical informatics, 58, 59-69.

[21] Gallón, V. M., Vélez, S. M., Ramírez, J., Bolaños, F. (2024). Comparison of machine learning algorithms and feature extraction techniques for the automatic detection of surface EMG activation timing. Biomedical Signal Processing and Control, 94, 106266.

[22] Fleming, A., Stafford, N., Huang, S., Hu, X., Ferris, D. P., Huang, H. H. (2021). Myoelectric control of robotic lower limb prostheses: a review of electromyography interfaces, control paradigms, challenges and future directions. Journal of neural engineering, 18(4), 041004.

[23] Bhushan, S., Alshehri, M., Keshta, I., Chakraverti, A. K., Rajpurohit, J., Abugabah, A. (2022). An experimental analysis of various machine learning algorithms for hand gesture recognition. Electronics, 11(6), 968.

[24] Karheily, S., Moukadem, A., Courbot, J. B., Abdeslam, D. O. (2022). sEMG time–frequency features for hand movements classification. Expert Systems with Applications, 210, 118282.

[25] Liu, X., Sacks, J., Zhang, M., Richardson, A. G., Lucas, T. H.,Van der Spiegel, J. (2016). The virtual trackpad: An electromyography-based, wireless, real-time, low-power, embedded hand-gesture-recognition system using an event-driven artificial neural network. IEEE Transactions on Circuits and Systems II: Express Briefs, 64(11), 1257-1261.

[26] Pancholi, S., Joshi, A. M. (2019). Electromyography-based hand gesture recognition system for upper limb amputees. IEEE Sensors Letters, 3(3), 1-4.

[27] Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximators. Neural networks, 2(5), 359-366.

[28] Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems.

[29] Cunningham, H., Ewart, A., Riggs, L., Huben, R.,Sharkey, L. (2023). Sparse autoencoders find highly interpretable features in language models. arXiv preprint arXiv:2309.08600.

[30] Kok, C. L., Ho, C. K., Tan, F. K., Koh, Y. Y. (2024). Machine learning-based feature extraction and classification of emg signals for intuitive prosthetic control. Applied Sciences, 14(13), 5784.

[31] Shaw, H. O., Devin, K. M., Tang, J., Jiang, L. (2024). Evaluation of Hand Action Classification Performance Using Machine Learning Based on Signals from Two sEMG Electrodes. Sensors, 24(8), 2383.