Real-Time Hand Gesture Recognition System with User Identification for Industrial Applications

Real-Time Hand Gesture Recognition System with User Identification for Industrial Applications

Marco Esposito*, Sara Raggiunto†, Luisiana Sabbatini*, Alberto Belli*, Sara Bruschi*,
Lorenzo Palma*, Emanuele Storti*, Stefano Rossini‡, and Paola Pierleoni*
*Department of Information Engineering (DII)
Università Politecnica delle Marche
Ancona, Italy
m.esposito@pm.univpm.it
†National Inter-University Consortium for Telecommunications (CNIT)
Research Unit of Università Politecnica delle Marche
Ancona, Italy
‡Università Politecnica delle Marche
Ancona, Italy

***Abstract*—Human-Machine interaction plays a pivotal role in contemporary industrial and civil environments. Among other sensing technologies, Leap Motion Controller is a non-contact and reliable one to be used for gesture recognition. In this paper, a dataset composed of more than two thousand samples is presented. The data have been collected by 17 subjects performing multiple repetitions of 26 gestures based on the Italian sign language alphabet. The "LEAP-SDGest: Leap Motion static & dynamic gestures" dataset can be adopted for the development of algorithms for both gesture recognition and user identification, for example, in industrial control applications. An architecture leveraging the MQTT protocol is proposed for this purpose, embedding gesture recognition and access control by exploiting classifier models and MQTT topics structure. In order to validate the dataset and proposed methodology, Machine Learning and Deep Learning models for both control tasks were developed. After a preliminary analysis, the Multi-Layer Perceptron proved to be the best solution and was optimized through grid-search. The resulting models reached 91.6% accuracy for user recognition and 97% accuracy on gesture recognition through very lightweight models and reduced pre-processing. Under strong authentication conditions, an end-to-end success rate of 91% was achieved for the entire system that includes access control and user recognition.**

***Index Terms*—Human-Machine Interface, gesture recognition, Leap Motion, user identification, Machine Learning**

Finanziato
dall'Unione europea
NextGenerationEU

This work has been partially supported by the PRIN 2022 project "HOMEY: a Human-centric IoE-based Framework for Supporting the Transition Towards Industry 5.0", funded by the European Union - Next Generation EU, Mission 4 Component 1 (code: 2022NX7WKE, CUP: F53D23004340006) DOI: 10.1109/MetroXRAINE66377.2025.11340503

I. Introduction

In recent years, advances in Human–Machine Interfaces (HMIs) have driven the development of gesture recognition technologies for diverse applications, highlighting the growing need for secure, real-time systems capable of both gesture interpretation and user identification, particularly in industrial environments. Many different sensing technologies, both wearable and contactless, can be used for gesture recognition [1]. Sensors such as Inertial Measurement Units (IMUs) and infrared systems have become enablers of real-time hand gesture recognition and user authentication, since they offer robust biometric data, critical for secure industrial applications, and are hard to spoof [2].

Among the advanced technologies based on infrared sensors for gesture recognition, Leap Motion Controllers are commonly used. These devices have opened new opportunities in gesture recognition due to their high accuracy in hand tracking (for example, better with respect to wearable Electromyography bands [3]), particularly in capturing the depth of the hands and fingers [4].

The Leap Motion software provides direct 3D hand skeleton data, which includes 3D coordinates for hand direction, palm center, and fingertips' positions. These features are low-dimensional compared to RGB images, but still informative on hand dynamics and characteristics, making it possible to develop simpler, lighter models [5]. Tracking through Leap Motion has generally different advantages compared to standard RGB cameras, as it is less sensitive to lighting conditions, skin color variations, and background interference [5]. As a consequence, skeleton-based tracking has become the standard in many VR applications, although it can be limited by sensor range and precision at wider distances [5].

Although there are many open source hand gesture datasets, most are based on RGB or depth-camera sensing, such as HGM-4 [6], Myanmar Sign Language [7], DHG 14–28 [8], and [9]. The aforementioned datasets are either focused on static gestures or dynamic ones only, and no one considers both types of gesture. They are also often noisy, have limited gesture diversity and/or dataset imbalance [5], in addition to issues specific to the sensing modality, such as skin color problems when using RGB cameras. Only a few major open-source datasets exist for Leap Motion skeleton-based tracking (LMDHG [10], LeapGestureDB [11]), both having different limitations (the first has noisy acquisitions, the second a very limited gesture dictionary). In this work, a new dataset is introduced that addresses these limitations by providing a broader set of gesture classes (both static and dynamic) in a controlled environment and recordings from multiple users. Subsequently, the dataset is validated by developing machine learning (ML) models for gesture recognition and user identification. Many existing approaches have validated deep learning (DL) methods for this task. For example, Caputo et al. [12] showed that skeleton-based hand gesture recognition achieves very strong results with DL methods. However, they remark that the computational load of the deep models they trained still required high-end hardware, and portability to mobile AR/VR systems (e.g., HoloLens, Oculus Quest) remains uncertain. On the other hand, Polsinelli et al. [13] propose a portable setup that combines a Leap Motion Controller, embedded skeleton extraction via LeapC, and a low-cost embedded system (Raspberry Pi). Their proposal, which was aimed at forearm tracking, demonstrates the feasibility of portable, low-cost, and low-power solutions that incorporate skeleton tracking. Based on these considerations, lightweight modeling strategies based on extracted Leap Motion features were explored, and a Raspberry Pi-based system for gesture recognition, user identification, and command transmission towards industrial devices is presented. The proposed system uses the Message Queue Telemetry Transport (MQTT) protocol to control the message exchange between the users and the end devices, taking advantage of MQTT topics' hierarchy to implement an effective form of access control.

This work aims at two main objectives. First, it provides a large dataset that overcomes the main limitations in existing skeleton-based datasets for gesture recognition. Secondly, it introduces a complete end-to-end architecture for human–machine interaction, based on low-cost, accurate, and easy-to-use devices such as the Leap Motion Controller, combined with machine learning techniques for the accurate interpretation of sign language commands. Gesture recognition requires high accuracy and stability to be suitable for applications that demand high precision and operator safety, such as the control of robots and industrial machinery. Authenticated users can control and interact through the interfaces without the need for physical contact.

II. Proposed Gesture Recognition System

A. System Architecture

A complete architecture for real-time hand gesture recognition is proposed. As shown in Fig. 1 the proposed system is composed of a Leap Motion sensor and a Raspberry, moreover the communication architecture is based on the MQTT protocol. The Raspberry used in this work is equipped with a Broadcom BCM2711 CPU (Broadcom Inc., North San Jose, CA, USA), Quad-core Cortex-A72 (ARM v8) 64-bit SoC at 1.8 GHz, and 4 GB of RAM. For the communication between Leap Motion and Raspberry and for data acquisition the LeapC API was used. The LeapC driver and the software to use Leap Motion are installed on the Raspberry, and a model of the hand is created from the acquired data. These data are organized into frames, and the data frame stream includes data on the frame ID, the frame rate, the IDs of the hand, and all the coordinates of the joints of each hand, which are extracted and saved in a CSV file. Moreover, the gesture recognition and user identification models and the MQTT broker are located on the Raspberry. Figure 1 shows the proposed architecture for authentication and gesture classification. At system startup, each smart device subscribes to its own dedicated topic. During runtime, two ML algorithms running on the acquisition device estimate the user identity and the gesture class, respectively. Each gesture class maps a specific command for the target device. The Raspberry device uses the authentication model to determine the MQTT topic where the recognized gesture command is going to be published. By employing separate MQTT topics for each device, and by granting publishing permission to publish on that topic only to a certain user, access control can be enforced and easily configured.

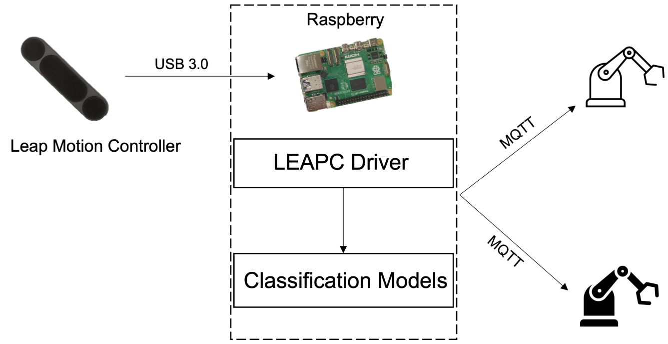

Fig. 1. Proposed system architecture.

B. Data Acquisition

In this study, common hand gestures were used to control the remote interface. To this end, a gesture dataset was collected using the Leap Motion Controller. Our dataset contains 26 gestures performed by 17 subjects (14 men and 3 women). The Italian sign language (LIS), associated to the 26 alphabet letters, has been used as the dictionary. All subjects performed five repetitions of each motion, giving about 2210 acquisitions. In our acquisition protocol, the applied exclusion criteria are the removal from the dataset the gesture recorded in a room with different lighting conditions, obtaining a final dataset of 2036 acquisitions. Each acquisition contains about 120 frames of hand tracking data.

C. Classification Models

1) Preprocessing: For each gesture, a set of relevant hand points features are extracted through LeapC, namely the hand direction, palm position, palm normal, and finger position. All features are computed along X, Y, and Z directions. These features are used for two distinct downstream tasks: i) gesture recognition, which is based on the specific position held by the hand during gestures, and ii) user identification, which is based on the idea that each user's hand has unique characteristics, allowing ML models to match gestures to individuals based on hand characteristics extracted from the Leap Motion frames. Each acquisition was turned into a single feature vector by computing the mean over time. Since gestures are static or fairly simple, this process yields a stable summary of the user's hand pose over the 120 frames, leading to more robust features compared to single-frame classification, while removing redundancy in input features, computational overhead, and the chance of false positives due to continuous analysis of standalone frames [5]. The resulting features are then normalized using Z-score standardization, and the dataset is split into training and testing sets with stratification over a combined user-gesture label to ensure a fair representation of both labels in the dataset.

2) Training and Hyperparameter Search: An initial comparison was carried out using Linear Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and a One-Layer Perceptron model for gesture classification, with the perceptron achieving the highest accuracy (SVM - 0.85, LDA - 0.74, OLP - 0.97). Consequently, the perceptron architecture was chosen as the reference model for subsequent experiments. A grid search with five-fold cross-validation was then conducted to optimize a Multi-Layer Perceptron (MLP) architecture, exploring different hidden layer sizes, activation functions (ReLU and Tanh), and weight decays. The complete hyperparameter space for the grid search is in Table I, which results in 5 folds for each of the 72 candidates, totaling 360 fits. This optimization procedure was performed separately for both gesture recognition and user authentication tasks, resulting in 2 separate models. After hyperparameter search through five-fold cross-validation, the best model was refitted on the entire training set and evaluated on the held-out test set.

TABLE I
GRID SEARCH HYPERPARAMETER SPACE FOR MLP MODELS

Hyperparameter	Values Explored
Hidden layer sizes	(64, 64), (128, 128), (128, 64), (64, 128), (192, 128), (128, 192)
Activation function	ReLU, Tanh
Solver	Adam, SGD
Weight decay	0.0001, 0.001, 0.01, 0.05

3) Testing and Evaluation: Performance was evaluated separately for the two developed models and then for the entire control chain, consisting of user authentication and gesture recognition, under the assumption of strong authentication (i.e., for each of the n users, the system only accepts a control from the user i if the user was correctly classified as the i^{th} user). In this case, the conditional gesture recognition accuracy was first computed, i.e., the gesture classification accuracy only for correctly authenticated samples, and then overall control success, which has been computed as the product of authentication and conditional gesture recognition accuracies, therefore also taking into account missed authentication.

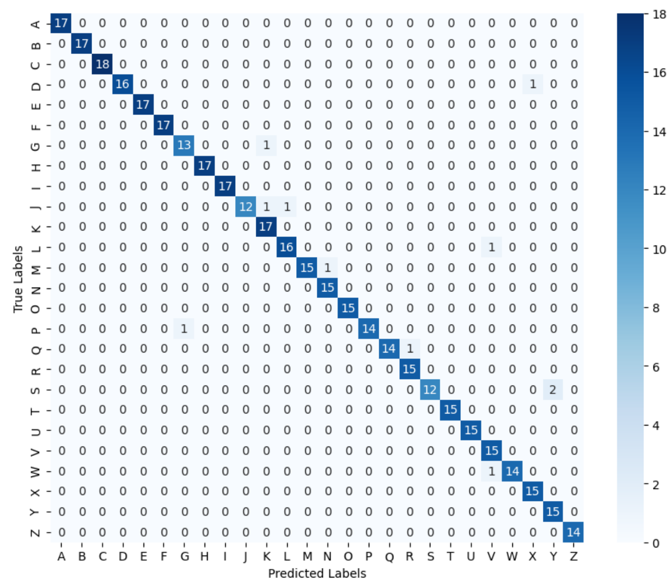

Fig. 2. Confusion matrix for the task of gesture recognition.

III. RESULTS

A. Gesture Recognition

The best configuration identified using grid search was a hidden layer size of (192, 128), Tanh activation, $\alpha = 0.0001$, and Adam optimizer. The cross-validation accuracy was 94.8%. After refitting the model on the entire dataset, a test accuracy of 98.2% was achieved on the hold-out test set, with a macro-averaged precision and recall of 98.0%. Figure 2 shows the test confusion matrix for this model, which highlights classification gaps only for some specific gestures mostly. These gestures are not purely static, thus suggesting as a next step an improved preprocessing strategy to develop a better gesture recognition model.

B. User Identification

The best-performing model found through grid search was the one with a hidden layer size of (128, 192), ReLU activation function, $\alpha = 0.01$, and Adam optimizer, achieving a cross-validation accuracy of 79.9%. After selecting the best hyperparameters, the model was retrained on the complete training set and evaluated on the hold-out test set. The resulting test accuracy was 91.6%, with a macro-averaged precision of 92% and a recall of 91%. The confusion matrix for this model is shown in Figure 3. A summary of the main results is presented

TABLE II
CLASSIFICATION PERFORMANCE SUMMARY FOR USER AUTHENTICATION AND GESTURE RECOGNITION TASKS

Task	**Cross-Val Accuracy**	**Gridsearch Test Accuracy**	**Final Test Accuracy**	**Macro Precision**	**Macro Recall**
User Authentication	79.4%	89.3%	91.6%	91%	91%
Gesture Recognition	97.3%	97.3%	97.3%	97%	97%

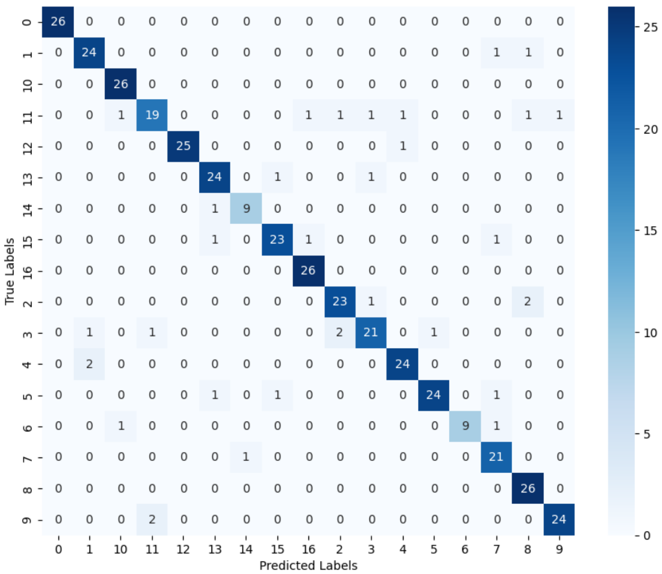

Fig. 3. Confusion matrix for the task of user identification.

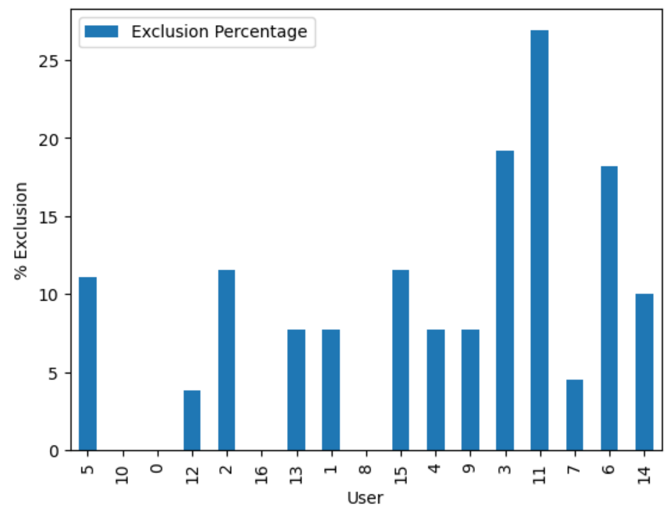

Fig. 4. Percentage of exclusions per user across the entire testing dataset, assuming a strong authentication policy.

in Table II, including cross-validation, test, and refitted model performances for both gesture and user identification tasks.

C. Overall Identification Accuracy

The complete control chain, which combines user authentication followed by gesture recognition, was evaluated to simulate real-world usage. First, authentication is performed using the optimized MLP model trained for user identification, then gesture recognition is performed conditionally, considering only correctly authenticated samples, resulting in a conditional gesture recognition accuracy of 98.6%. The overall system success rate, calculated as the product of authentication and conditional gesture recognition accuracy, is 91%, which means that about 9 out of 10 interaction attempts were correctly processed from an end-to-end point of view. In these cases, a command may not reach the device if the misclassification results in a user being assigned an identity that is not authorized to execute the command. Under the assumption of enforcing strong authentication, i.e., of a command reaching the destination only if the user was correctly identified, the percentage of exclusions per user over the entire testing dataset was computed to highlight users that are more challenging to authenticate. This exclusion metric reflects a real-world scenario where authorizations are granted to a single user at a time. The results are shown in Figure 4. Users 1, 6, and 11 exhibited the highest exclusion rates, indicating lower authentication performance for these subjects. In general, the results for authentication are in line with previous work using non-invasive authentication systems [2], [14]

IV. CONCLUSION

Nowadays, efficient and reliable human–machine interaction is essential to ensure safety at work, enhance decision-making processes, and improve operational efficiency. In this paper, a dataset consisting of 26 hand gestures performed by 17 subjects, each repeated 5 times, was presented. Additionally, two classification models were implemented and tested: one for user identification and the other for gesture recognition. This study aims not only at providing a comprehensive public dataset for future developments but also at delivering algorithms capable of recognizing both gestures and the identity of the user. This approach, applicable in a variety of contexts, enables secure and private interaction for the operator, addresses emerging challenges through the proposed architecture based on the MQTT protocol, and enhances human capabilities by facilitating interaction between humans and machines. The results obtained show high test accuracies, reaching 91.6% and 98.2% for user authentication and gesture recognition, respectively. Although user authentication accuracy is slightly lower than gesture recognition, this is expected given the lower inter-subject variability in hand geometry compared other common or more invasive biometrics. Nevertheless, hand-based authentication offers the important advantages of non-intrusive, contactless sensing, and increased user privacy, favoring its adoption in privacy-sensitive or industrial applications. Future work will focus on exploring more advanced deep learning techniques to enhance the generalizability and robustness of user authentication, building upon the comprehensive open dataset presented in this study.

Data Availability

The data collected and used for the development of this work are available on Kaggle, searching for "LEAP-SDGest: Leap Motion static & dynamic gestures" (link: Kaggle - LEAP-SDGest: Leap Motion static & dynamic gestures)

References

[1] L. Guo, Z. Lu, and L. Yao, "Human-machine interaction sensing technology based on hand gesture recognition: A review," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 4, pp. 300–309, 2021.

[2] S. Bianco, P. Napoletano, A. Raimondi, and M. Rima, "U-wear: User recognition on wearable devices through arm gesture," *IEEE transactions on human-machine systems*, vol. 52, no. 4, pp. 713–724, 2022.

[3] K. Erin, M. Ç. Kutlu, and B. Boru, "Comparison of gesture classification methods with contact and non-contact sensors for human-computer interaction," *Sigma Journal of Engineering and Natural Sciences*, vol. 40, no. 2, pp. 219–226, 2022.

[4] J. Huo, K. L. Keung, C. K. Lee, and H. Y. Ng, "Hand gesture recognition with augmented reality and leap motion controller," in *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE, 2021, pp. 1015–1019.

[5] N. M. Bhiri, S. Ameur, I. Alouani, M. A. Mahjoub, and A. B. Khalifa, "Hand gesture recognition with focus on leap motion: An overview, real world challenges and future directions," *Expert Systems with Applications*, vol. 226, p. 120125, 2023.

[6] V. Truong Hoang, "Hgm-4," 2020. [Online]. Available: https://doi.org/10.17632/jzy8zngkbg.2

[7] G. R. Sinha and P. S. Oo, "Myanmar sign language," 2019. [Online]. Available: https://dx.doi.org/10.21227/vq7b-e978

[8] Q. De Smedt, H. Wannous, and J.-P. Vandeborre, "Skeleton-based dynamic hand gesture recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 1–9.

[9] G. Fronteddu, S. Porcu, A. Floris, and L. Atzori, "A dynamic hand gesture recognition dataset for human-computer interfaces," *Computer Networks*, vol. 205, p. 108781, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128622000172

[10] S. Y. Boulahia, E. Anquetil, F. Multon, and R. Kulpa, "Dynamic hand gesture recognition based on 3d pattern assembled trajectories," in *2017 seventh international conference on image processing theory, tools and applications (IPTA)*. IEEE, 2017, pp. 1–6.

[11] S. Ameur, A. Ben Khalifa, and M. S. Bouhlel, "Leapgesturedb: A public leap motion database applied for dynamic hand gesture recognition in surgical procedures," in *Soft Computing Applications: Proceedings of the 8th International Workshop Soft Computing Applications (SOFA 2018), Vol. II 8*. Springer, 2021, pp. 125–138.

[12] A. Caputo, A. Giachetti, S. Soso, D. Pintani, A. D'Eusanio, S. Pini, G. Borghi, A. Simoni, R. Vezzani, R. Cucchiara *et al.*, "Shrec 2021: track on skeleton-based hand gesture recognition in the wild (2021)," *URL: https://arxiv. org/abs/2106.10980*.

[13] M. Polsinelli, A. Di Matteo, D. Lozzi, E. Mattei, F. Mignosi, L. Nazzicone, V. Stornelli, and G. Placidi, "Portable head-mounted system for mobile forearm tracking," *Sensors*, vol. 24, no. 7, p. 2227, 2024.

[14] L. Hernández-Álvarez, J. M. de Fuentes, L. González-Manzano, and L. Hernández Encinas, "Privacy-preserving sensor-based continuous authentication and user profiling: A review," *Sensors (Basel)*, vol. 21, no. 1, p. 92, Dec. 2020.