IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Tactile Object Recognition With Recurrent Neural Networks Through a Perceptive Soft Gripper

Enrico Donato, *Member, IEEE*, David Pelliccia, Matin Hosseinzadeh, Mahmood Amiri, and Egidio Falotico, *Member, IEEE*

***Abstract*—Soft robot perception integrates information from distributed, multi-modal sensors, broadening their application to active interaction. Our work introduces recurrent learning models for tactile-based object recognition, demonstrating comparable performance in virtual and real-world scenarios. The work focuses on soft grippers, which facilitate adaptation to objects of varying shapes and sizes thanks to passive finger compliance. Our model successfully identifies over sixteen heterogeneous objects. Findings underscore the significance of sensory multi-modality over single. We highlight how spatial distribution and sensory signal dynamics influence overall estimation accuracy, and what the minimal grasp set is to achieve certain recognition.**

***Index Terms*—Tactile sensing, multi-modal integration, object recognition, soft gripper.**

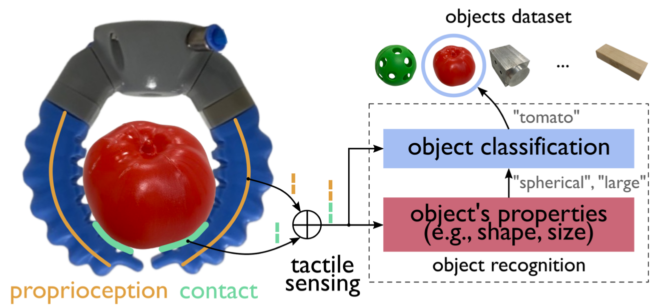

Fig. 1. Tactile object recognition with a multi-modal sensing soft gripper, featuring a dual-stage, learning-based algorithm for object classification.

I. Introduction

UNLOCKING robot potential resides in developing an efficient representation of themselves and the surrounding environment [1]. By mastering data manipulation and overcoming multi-modal sensory fusion, robots can achieve perceptual awareness. It provides a comprehensive representation of the experience, derived from coordinating multiple sources and characterized by completeness and coherence. Such a state renders awareness reliable for closing the control loop or devising perceptive-aware strategies [2].

Perceptual awareness is pivotal while acting on the world, as the robot embodiment shapes the sensory feedback. This is particularly the case of soft robots, which incorporate compliant components to enhance versatility for interaction [3]. These components enable passive adaptability to the external world, leveraging the deformable nature of the materials used in fabrication. Our focus is on soft grippers [4], whose body compliance is at finger level. Previous research has explored various design options and their practical applications for manipulation tasks, including object holding [5], grasping [6], and in-hand manipulation [7]. However, robot control rises as a significant challenge [8], induced by the abundance of passive degrees of freedom and undesired deformations. Integrating onboard sensing capabilities is crucial in enhancing task performance by enabling informed decision-making [9]. Sensors provide valuable feedback to drive actions, assess grasp stability [10] or inform further decision schemes, as per object classification [11].

In this work, we leverage multi-modal information by integrating proprioceptive (finger curvature) and exteroceptive sensing (contact pressure) on a commercial soft gripper. Moreover, our computational design evaluation of different sensor configurations demonstrates how an optimal sensor layout can be determined, balancing performance with system complexity. We propose a tactile-driven learning algorithm for object recognition using a sensorized soft gripper, shown in Fig. 1, whose feedback incorporates proprioception - finger bending - and contact state measurement. The algorithm is implemented using a two-stage architecture and tested on sixteen diverse objects. It first classifies the shape and size of the object, to later boost tactile information to classify the grasped object with higher accuracy in a second stage. Additionally, a single object grasp does not guarantee correct recognition, therefore consecutive grasps of the same object are considered to improve confidence. Our key contributions are:

- a novel dual-stage LSTM-based framework for multi-modal object recognition with a soft robotic gripper;
- a computational design methodology to identify the sensory spatial placement that bolsters object recognition;

Received 20 December 2024; accepted 11 May 2025. Date of publication 22 May 2025; date of current version 2 June 2025. This article was recommended for publication by Associate Editor J. Hughes and Editor Y.-L. Park upon evaluation of the reviewers' comments. This work was supported in part by the European Union's Horizon 2020 Research and Innovation program through PROBOSCIS Project under Grant 863212 and in part by FONDAZIONE PISA - RICERCA SCIENTIFICA E TECNOLOGICA through SHARE-CP Project number 300/22 under Grant CUP J83C24000530007. *(Corresponding author: Enrico Donato.)*

Enrico Donato and Egidio Falotico are with the BRAin-Inspired Robotics (BRAIR) Lab, The BioRobotics Institute, 56025 Pontedera (PI), Italy, and also with the Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, 56127 Pisa, Italy (e-mail: enrico.donato@santannapisa.it).

David Pelliccia is with the Department of Computer Science, University of Pisa, 56127 Pisa, Italy.

Matin Hosseinzadeh and Mahmood Amiri are with the Medical Technology Research Center, Institute of Health Technology, Kermanshah University of Medical Sciences, Kermanshah 56127, Iran.

This article has supplementary downloadable material available at https://doi.org/10.1109/LRA.2025.3572422, provided by the authors.

Digital Object Identifier 10.1109/LRA.2025.3572422

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

- an iterative classification algorithm that leverages experience to bolster object recognition over time;
- experimental and simulation results that validate the robustness and effectiveness of our approach.

We will first outline the object recognition algorithm and its training strategy in Section III, also introducing baseline models for performance comparison. The dataset-gathering process is detailed in Section IV, on both simulated and physical scenarios. Results reported in Section V support explanations on how multi-modal sensing enhances object classification, the significance of temporal dependence in tactile information, and the estimation of the minimal training set required for optimal network performance.

II. Related Works

A. Soft Robot Perception for Interaction

Robot sensing encompasses the collection of both exteroceptive and proprioceptive information [12], also by relying on asynchronous experiences [13]. Transitioning from traditional to soft-bodied robots necessitates a reevaluation of sensor design due to the absence of rigid bodies. This evolution requires shifting from discrete to continuous information along the body [14]. Implementing sensors capable of addressing deformations of soft robots remains a challenge.

Our emphasis lies on tactile sensing, encompassing both kinesthetic feedback and localized contact measurement [15]. Regarding proprioceptive feedback, soft sensors can measure bending or axial deformations, with recent advancements enabling the detection of twisting [16]. Various solutions exist for contact sensing, leveraging a broad spectrum of technologies for signal transduction [17], thereby providing information on localized contact pressures or forces.

Control algorithms derive significant advantages from robot perception of its surroundings, paving the way for the development of sensorimotor controllers [2] aimed at enhancing task adaptability, robustness and stability. Indeed, soft robot perception is exploited for robot shape regulation and reconstruction [18], object detection and obstacle avoidance [19], autonomous learning to navigate the environment or learning-from-demonstration by human operators [20].

B. Tactile Object Recognition

Object recognition involves determining the most probable class to which an object belongs by leveraging global features estimation [21]. Objects can differ in several characteristics, making recognition a hard challenge. These characteristics encompass: *shape* - either regular or irregular - *deformability*, indicating the extent to which shape changes under applied load, and *size*. We focus our analysis on rigid objects that differ solely in shape and size. Additionally, object pose can hinder recognition if not experienced.

We focus on blind object recognition, exclusively relying on tactile sensing. Discussions on human tactile sensing highlight how spatio-temporal patterns of neuromorphic tactile signals can discriminate shape and size, with both active and reactive grasping strategies [22]. Further implementations have utilized Gaussian process classification for sparse data [23], covariance matrix estimation of tactile array signals [24], and CNN-Bayes classifiers on soft grippers [25]. However, these methods are limited to regular objects.

Tactile sensors generally perceive only local information, prompting recognition through multiple grasps. This process can be achieved through surface sliding [26] or by integrating different tactile descriptors over time using Bayesian processes [27]. Dynamic exploratory procedures have demonstrated superiority over single-touch recognition [28], driving research towards consecutive multi-grasp object recognition. Our algorithm stands out for simultaneously exploiting multi-modal soft tactile sensing, enhancing mid-stage knowledge to improve recognition accuracy of diverse everyday objects.

III. Tactile Object Recognition Learning Algorithm

We propose a custom Long Short-Term Memory (LSTM)-based learning architecture as a model for tactile object recognition. This selection underscores the significance of temporal correlation in tactile perception, given the LSTM capability to retain state information across extended time intervals. When integrated with spatial sensory distribution, it enhances performance in sensing pattern classification. The novelty of our model lies in the pipeline devised for object recognition and the utilization of multi-modal sensing. Initially, it categorizes objects based on generic properties such as dimension and shape. Subsequently, it identifies the object class by exploiting such prior and the original tactile information. The intermediate output enriches the information fed to the final stage, thereby enhancing object discrimination and increasing classification accuracy. The system operates with a commercially available pneumatic soft gripper equipped with multi-modal off-the-shelf sensors. While the use case focuses on this specific hardware setup, the algorithm can be adapted to other grippers.

The Dual-Stage LSTM (DS-LSTM) in Fig. 2(b) is realised as a sequence of two macro-layers, as depicted in Fig. 2(a). The initial model aims to classify the dimension and shape of the object based on tactile sensing data. The second model, benefiting from a skip connection from the input, continues to process tactile data while also receiving the first stage classification to predict the actual class of the grasped object. To align with the information flow, the model undergoes training in two sequential stages.

The architecture of each LSTM network was informed by an assessment of the model proposed by [29]. Their findings suggested that object classification could benefit from a 4-layer network, incorporating a LSTM layer alongside dropout and fully connected layers. To enhance accuracy, a novel 5-layer architecture was identified through model selection. This new design includes an additional fully connected layer between LSTM and dropout, along with an increased number of units per layer - and the dropout probability is reduced.

To bolster confidence in object classification, an enhanced variant of the model is implemented to ensure that classification confidence exceeds a predefined threshold. In this scenario, following the analysis of an initial grasp, the model would output a discrete probability density distribution indicating the likelihood of an object belonging to various classes. The recognition process terminates if any of these probabilities surpass the threshold. Otherwise, another recognition loop ensues, leveraging the last

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

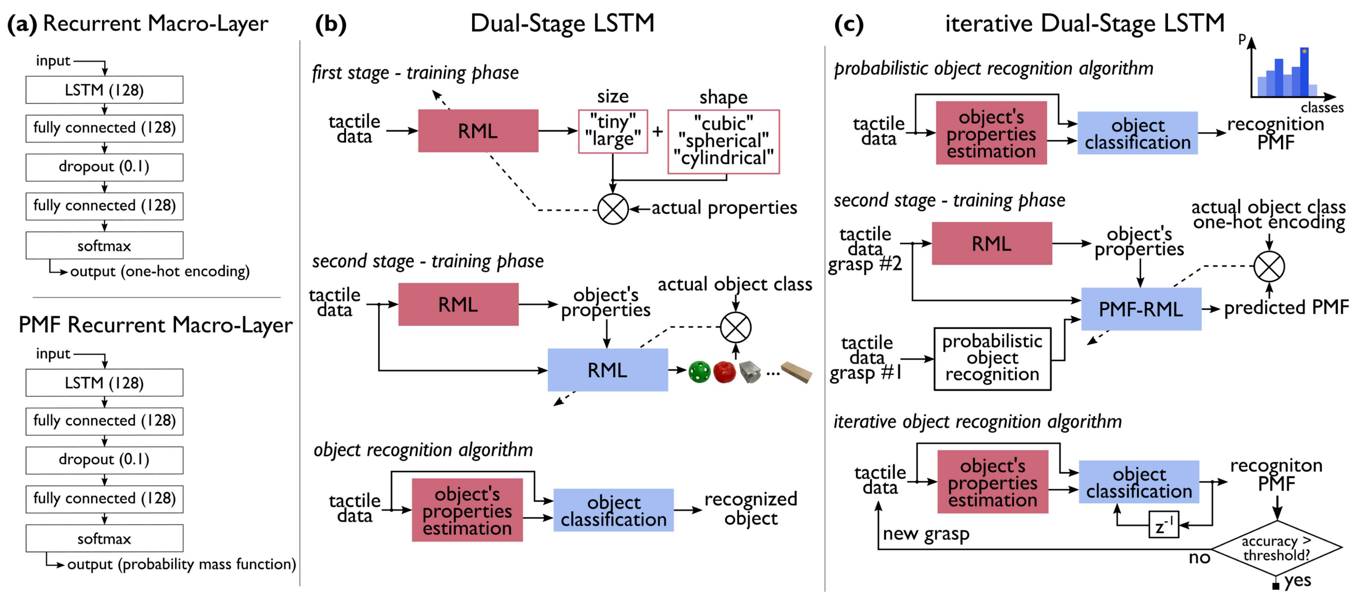

Fig. 2. Tactile object recognition algorithm. (a) Macro-layers utilized in the learning architecture. The Recurrent Macro-Layer and PMF Recurrent Macro-Layer output feature classification as one-hot encoding or Probability Mass Function. (b) Dual-Stage LSTM. Sequential training phases of both stages are outlined, along with the object recognition algorithm. (c) Iterative Dual-Stage LSTM. The first stage training phase mirrors that of the Dual-Stage LSTM, while the second stage employs a modified training methodology leveraging the probabilistic object recognition algorithm. The iterative object recognition algorithm is also illustrated.

probability estimation and the latest sensing sample. This refined approach leads the implementation of the Iterative Dual-Stage LSTM (iDS-LSTM), as in Fig. 2(c), which diverges from the DS-LSTM architecture primarily in terms of the final output and the inclusion of a recurrent connection from the final layer to the mid-stage. Similar to the DS-LSTM, the training of iDS-LSTM unfolds in two sequential stages. One-hot encoding is adopted to learn the distribution of objects class. The training of the network poses a challenge, as it necessitates a mock DS-LSTM architecture that learns the distribution instead of simple object classification. This architecture will be utilized solely during training to generate the input distribution of the previous grasp to the second part of the iDS-LSTM network.

A. *Evaluation and Benchmarking*

The performance of the DS-LSTM is benchmarked against other algorithms to provide a comparative analysis. Various machine learning classification algorithms are categorized into single or ensemble learning models for evaluation. The single model group [30], [31], [32], [33] includes Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and basic LSTM. Ensemble models [34], [35], [36] include Random Forest (RF), Bagged Tree (BT), Gradient Boosting Classifier (GBC), and RUS Boosted Tree (RUSBT).

Each algorithm will undergo evaluation based on metrics such as *macro-accuracy* and *-f1-score*, considering the balanced nature of the dataset. Accuracy measures the proportion of correctly classified cases out of the total number of objects in the dataset. Precision and recall respectively gauge the fraction of instances correctly classified as belonging to a specific class out of all instances the model predicted to belong to that class, and out of all the actual instances in that class. The F1-score represents the harmonic mean of a classification system precision and recall. High values of these metrics indicate good performance.

IV. Experimental Setup

The learning dataset is gathered both in silico and on a physical soft gripper. The simulation has been implemented in SoMo [37], replicating the two-finger soft gripper and gathering sensing information regarding finger curvature and pressure contacts. The test in silico allows for the testing of a wide number of sensorizations of the gripper, focusing on multi-modality and variable spatial patterns. Similar objects have been tested for recognition in both scenarios.

A. *Virtual Scenario*

Each finger of the soft gripper is represented as a Manipulator class entity in SoMo, modeled as planar chains of N spring-loaded rotational joints. The choice for N balances the spatial discretization of the compliant body with the analytical model complexity. We set $N = 10$ to ensure continuous discretization, yet limiting the computational load.

Experimental objects, depicted in Fig. 3(a), are loaded into the environment using their URDF representation. They are categorized by size (tiny, large) and approximated shape (cubic, cylindrical, spherical). To gather a comprehensive dataset, object poses were randomized in terms of both position and orientation relative to the gripper. This strategy ensures that the system is exposed to a large set of tactile signatures, preventing over-reliance on a specific grasp configuration. We recorded 40-55 trials/object.

For gripper control, we implement hard-coded strategies for closing and releasing, to enable object grasping and release at the same position. The simulation loop runs for a fixed duration $T = 4s$, with an actuation frequency $f_a = 10$ Hz and a sampling

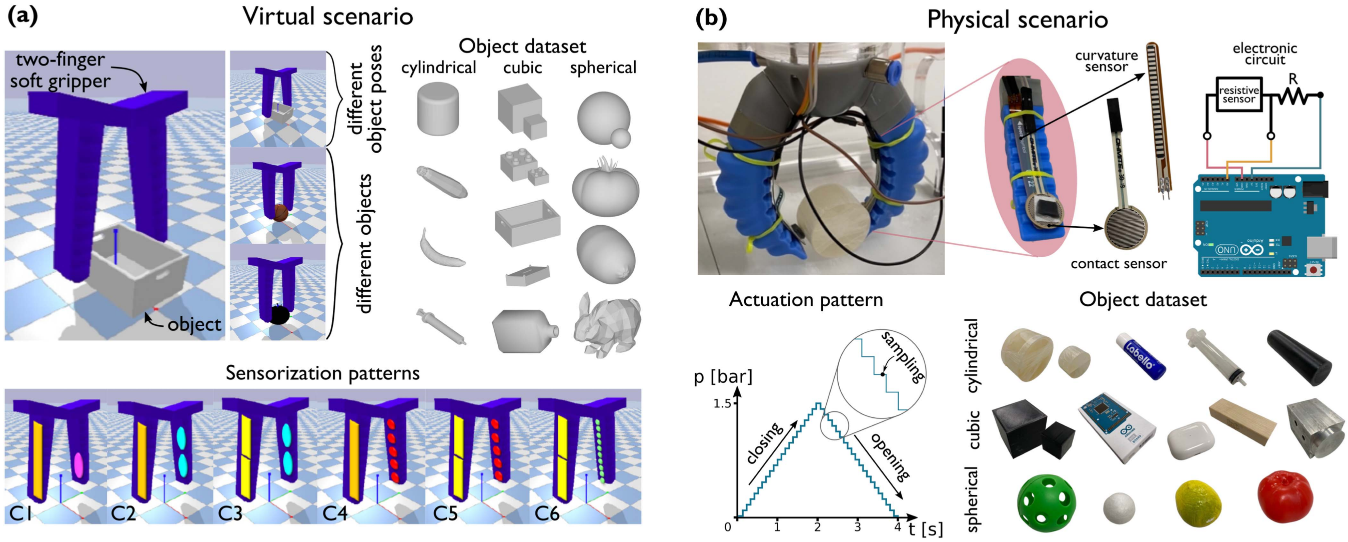

Fig. 3. Experimental setup in virtual and physical scenarios. (a) Two-finger soft gripper simulation in SoMo [37], collecting tactile information from various simulated objects. Spatial patterns of curvature and contact sensors are reported. (b) Sensorized pneumatic gripper equipped with curvature and contact sensors. The sampling electronics is reported, along with the actuation pattern for gripper closing and opening; sensory sampling is performed after the actuation transient to gather more static data.

frequency $f_s = 75$ Hz. At each time step, a commanded torque is applied to the fingers' joints, and sensing information for each finger is stored in a database, including: (a) the state q of proximal, medial, and distal joints, along with their coordinates x, y, z in the world reference frame; (b) if the finger is in contact with the object, the I normal forces $f_{j,i}$, where $i \in [0, I]$, exerted by the object on the finger joint j are stored, concerning contact location and force magnitude.

1) Sensing Simulation: The data initially obtained from the simulation loop are not directly usable for learning models, so they need post-processing. The goal is to mimic sensor outputs for object recognition algorithms.

Concerning proprioception, we aim to monitor finger curvature throughout the simulation loop. At each time step, we interpolate the pose of tracked joints for each finger along a circular curve, computing its radius r. The curvature k of the finger is then estimated as the inverse of the radius: $k = 1/r$. We consider discretizing the curvature along the finger, perhaps opting for two consecutive sensors to track changes in the proximal and distal parts of the finger. We employ a similar approach for curvature estimation using pairs of consecutive joints, leveraging q for curve fitting.

For exteroception, we simulate pressure sensors. For each finger joint j, we compute the area A_j representing the convex hull of contact points on the respective segment, along with the total force magnitude $F_j = \sum_{i=0}^{I} f_{j,i}$ on the same segment. The simulated pressure P_j sensed by a sensor is then computed as $P_j = F_j/A_j$. If longer pressure sensors covering multiple consecutive segments are desired, we replicate the computation while considering all the forces belonging to the chosen segments.

Tactile sensing is provided by combining kinaesthetic and contact information. We evaluate different spatial patterns of sensor placement during simulation experiments, as reported in Fig. 3(a), varying the spatial discretization of the information. Specifically, we explore all combinations of the following proprioceptive and exteroceptive patterns: curvature sensors tracking either the entire finger evolution or disjoint proximal and distal sections; pressure sensors measuring information at a single joint level, in pairs, as proximal and distal sections, or for the entire finger. We have trained learning algorithms to identify the minimal sensor configuration ensuring good performance, which could serve as a reference for physical sensorization.

B. Physical Scenario

The experimental setup in Fig. 3(b) comprises a SoftGripping parallel soft gripper [38] and custom finger sensorization, mirroring the simplest case in the virtual scenario. Thanks to its modular design, components can be easily substituted in case of damage. The gripper, mounted on a custom frame, operates within a pressure range of 0 to 1.5 bar, enabling varying degrees of finger movement. A control box simultaneously actuates both fingers at 0.05 bar steps.

The simulation-based sensorization analysis serves as a blueprint for integrating physical gripper sensors. Specifically, off-the-shelf bending and pressure resistive sensors have been integrated onboard. For proprioception, a Flexible Potentiometer (Spectra Symbol FS-L-0055-253-ST) measures finger deflection. It utilizes conductive ink on a phenolic resin substrate, with a segmented conductor forming a flexible potentiometer. Bending stretches the conductive layer, increasing resistance while straightening returns it to normal. For exteroception, a Force Sensing Resistor (Ohmite FSR07CE) gauges contact pressure on the sensor surface. Increased pressure enhances contact between carbon elements and conductive traces, thus reducing resistance.

Since both sensors function as variable resistors to different physical stimuli, their information can be captured by employing a voltage divider circuit. The voltage at terminals changes

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

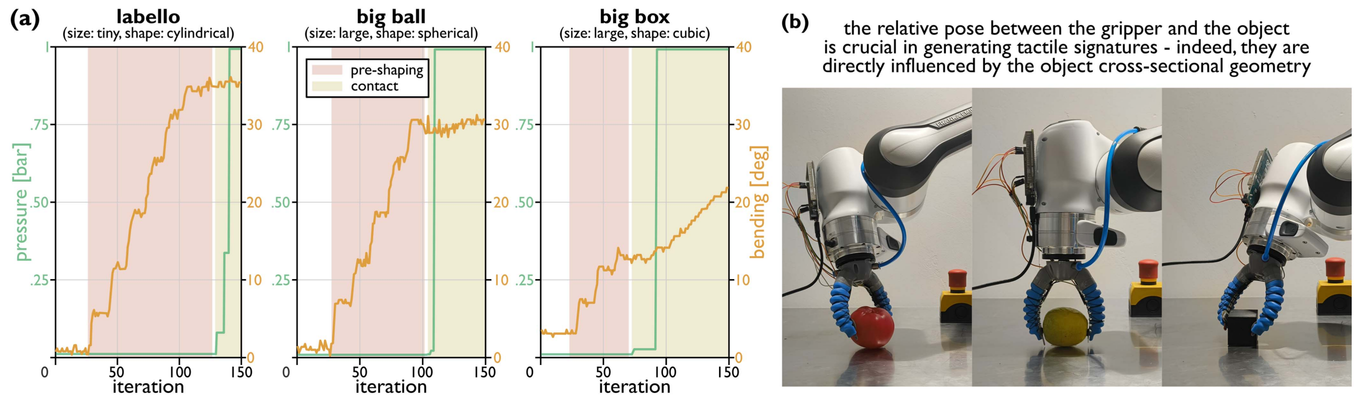

Fig. 4. (a) Tactile signatures for objects with different properties. (b) Examples of grasp configurations around physical objects.

according to their resistance, correlating with stimuli. To facilitate this, a sensing box is designed featuring an Arduino Due for data collection at $f_s = 50$ Hz and electric circuits for voltage dividers. After estimating the respective resistance values, sensors undergo characterization to establish relationships with physical stimuli. Specifically, the bending sensor is characterized while varying the finger angle from $\Theta_{\min} = 0$ deg relative to the palm attachment angle to $\Theta_{\max} = 90$ deg, demonstrating a linear relationship. Conversely, the pressure sensor is characterized against known weights P, ranging from $P_{\min} = 0$ kg to $P_{\max} = 0.5$ kg, revealing a nonlinear characterization curve.

V. Results

This section illustrates the experimental results from both virtual and physical scenarios. Initially, a task-driven approach is employed to evaluate soft gripper sensorization in simulation, aiming to correlate different sensing patterns with recognition accuracy. The objective is to identify sensorization patterns that offer optimal performance while minimizing onboard sensory load. Furthermore, the enhancement in class discrimination achieved through sensory multi-modality is evaluated. The most effective sensorization pattern is selected to assess the DS-LSTM model in both scenarios and benchmark the algorithm against other classification algorithms. Additionally, attention is directed towards determining the minimal training set required for the learning system to maintain performance without degradation. Lastly, the application of iDS-LSTM is demonstrated, highlighting the gains from consecutive grasps object recognition.

A. Raw Tactile Sensing

Tactile data analysis provides helpful insights into the discrimination of tactile events and the dependencies of signals on objects properties. Fig. 4(a) showcases the evolution of kinaesthetic and contact sensing over time for objects of varying shapes and sizes. The pre-shaping phase starts when the bending angle of the deformable finger starts to deviate from its initial value. During this phase, the contact signal remains consistently low, and the bending angle continues to increase until contact is established. For objects with a rounded shape, such as cylindrical or spherical forms, the bending reaches a steady state correlated with the object curvature. In contrast, for objects with a flat surface, the finger cannot conform to the shape, causing the bending angle to increase linearly even after contact is made. Regarding contact pressure, the signal saturates after few iterations following the contact event. The onset of contact pressure marks the beginning of the contact phase, which, after initial saturation, stabilizes at a high steady value due to object stiffness. The iteration at which contact occurs is correlated to the object size, with larger objects leading to earlier contact. Thus, tactile information provides essential insights for object identification, emphasizing multi-modal signals and temporal pattern analysis - arising from different configuration, as the ones depicted in Fig. 4(b).

B. Task-Driven Soft Gripper Sensorization

The sensory patterns proposed in Section IV-A1 have been implemented on the virtual soft gripper. Each pattern is individually employed in the task, and the gathered data are used to train the Recurrent Macro-Layer in Fig. 2(a). Model accuracy is benchmarked against the network by [29].

Examining Fig. 5(a), the highest accuracy is observed with configurations C1 to C3. Oversensorization proposed in configurations C4 to C6 is deemed uninformative for this task. Since most objects primarily contact the distal finger section without significantly altering the finger shape, fewer sensorized segments suffice. Among the top-performing configurations, the variance in performance is minimal. Opting for simplicity without compromising performance, configuration C1 emerges as a suitable compromise between resource demand and performance. In addition, Fig. 5(b) showcases the performance of the model due to different patterns for single modalities, highlighting the performance loss due to the absence of fundamental information and its fusion.

It is worth assessing the advantage of multi-modality sensing over single. The model is trained to recognize objects using either kinaesthetic and pressure sensors alone or both. Fig. 5(c) shows that the fusion of both sensory modalities yields superior performance in accuracy and F1-score, highlighting the advantages of leveraging multiple sources of sensory information. It happens both in virtual and physical scenarios, despite a lower performance in the latter due to signal noise and sensor placement.

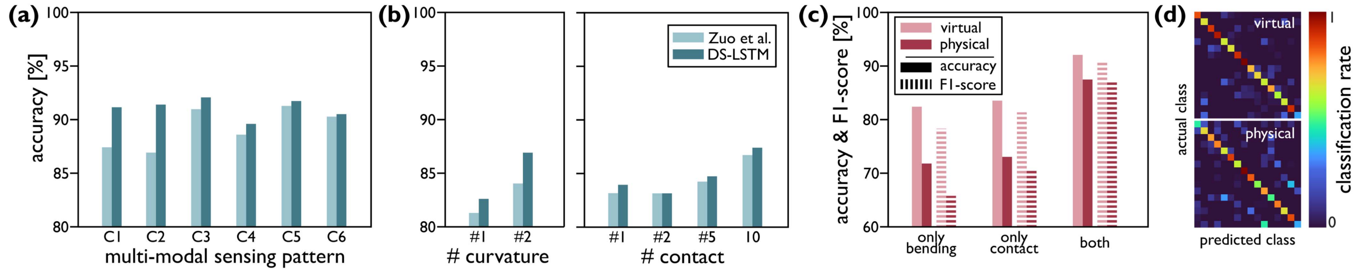

Fig. 5. Evaluation of sensory multi-modality. (a) Accuracy of different sensing pattern, benchmarked with the model of Zuo et al. [29]. (b) Accuracy over single-modalities. (c) Performance comparison between single- and multi-sensory modalities. (d) Object classification confusion matrix.

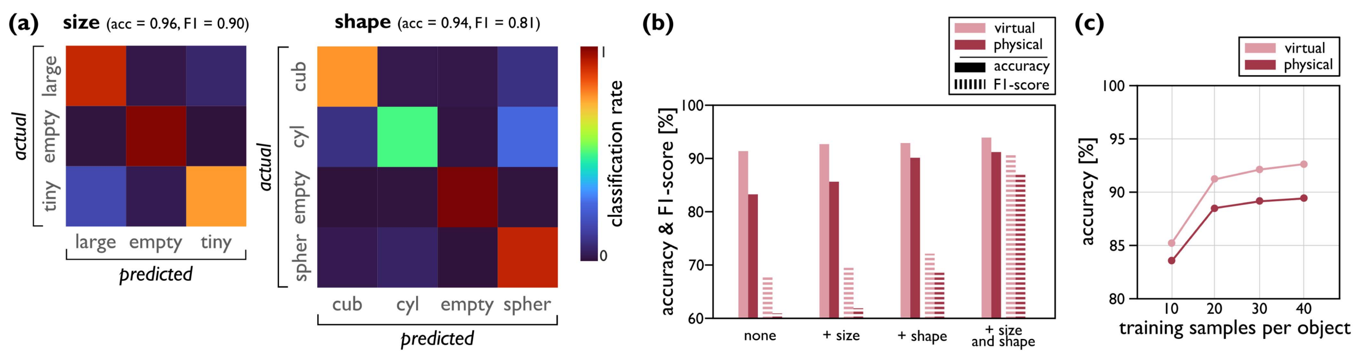

Fig. 6. (a) Dimension and shape estimation confusion matrices. (b) Object recognition performance while knowing only shape, only size or both. (c) Performance while reducing the number of grasps in the training set.

Fig. 5(d) shows a misclassification among specific object classes. We attribute these errors to objects with similar shapes and sizes - e.g., jenga and small cube, tomato and lemon. They inherently generate similar tactile cues, reducing the discriminative power of the current sensor configuration. Classification ambiguities can be mitigated by consecutive grasps of same objects.

C. Dual-Stage LSTM Model Evaluation and Benchmarking

The DS-LSTM model must undergo a dual-stage training phase. To do so, the grasp dataset recorded for each object is split into 40% for first-stage training and 40% for its testing. After the first-stage training, the same data exploited for its testing will be used as training data for the second stage, while the remaining 20% is the test set.

Results for the recognition of object size and shape are illustrated in Fig. 6(a), yielding a F1-score of 0.90 and 0.81, respectively. In terms of object recognition informed by its properties, sensor data were initially combined solely with size or shape information to demonstrate performance improvement when both pieces of information are provided simultaneously, as depicted in Fig. 6(b).

Given the time-consuming nature of dataset gathering, particularly in the physical scenario due to variations in object pose across runs, the aim is to determine the minimum number of grasps required for training with minimal performance degradation. To this end, we considered four training sets with increasing number of grasps per object, while maintaining the dataset split strategy. As illustrated in Fig. 6(c), a smaller number of grasps

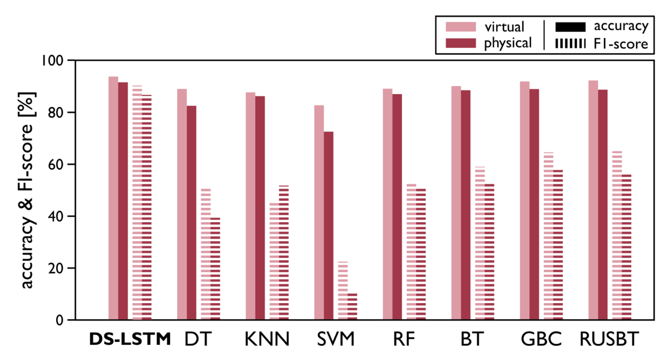

Fig. 7. Comparison of classification performance with single model approaches and ensemble methods. The proposed DS-LSTM model demonstrates superior performance in terms of macro-accuracy and F1-score.

leads to inferior results. However, as we transition towards more densely populated datasets, the impact on performance becomes less pronounced. Furthermore, the DS-LSTM model consistently outperforms other models across all cases, while remaining comparable to ensemble methods.

D. Comparative Analysis of Classification Methods

The DS-LSTM model demonstrates superior performance compared to other single models and ensemble methods for classification. As depicted in Fig. 7, our model exhibits the highest performance, albeit comparable to RUSBT. This architecture proves effective, particularly in classifying imbalanced data, where certain classes in the training data may have fewer

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

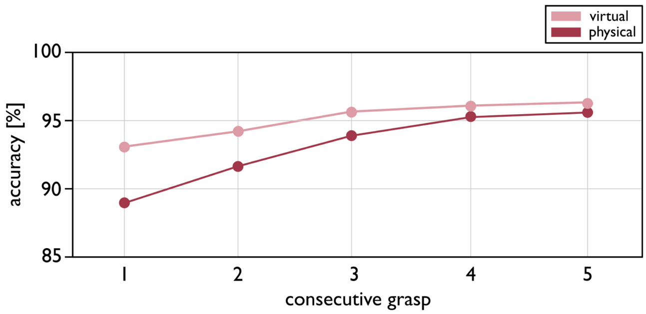

Fig. 8. Consecutive grasp recognition to bolster classification confidence.

instances than others. Additionally, GBC shows promising results, likely attributed to its utilization of time correlation akin to LSTM. Anyway, ensemble methods demand higher spatial and temporal complexity, requiring multiple single models to run concurrently. Overall, the fusion of temporal and spatial information has yielded substantial benefits and notably improved results, especially when compared with single classification models.

E. Iterative DS-LSTM Model

In line with the simple model, the iDS-LSTM model undergoes a dual-stage training phase. The dataset is structured to encompass consecutive grasps on the same object exclusively. The model is trained using the same policy as the DS-LSTM algorithm during the first stage. However, the mock DS-LSTM is employed during the second stage to generate the probability density distribution of previous grasps. Consequently, the dataset is organized as pairs of prior and current grasps of the same object.

Despite achieving a high level of accuracy in object recognition during the first step, Fig. 8 demonstrates the improvement in object recognition accuracy and F1-score over successive grasps. A steady performance is reached after three or four consecutive grasps, thus additional grasps might not be influential to iterative object classification.

VI. Discussion

The experimental results provide compelling evidence of the impact of the DS-LSTM for tactile object recognition. Our model surpasses other classification algorithms in terms of accuracy and precision, exhibiting capabilities on par with ensemble methods yet with higher computational efficiency.

We propose a methodology to identify minimal, task-specific soft gripper sensorization, enabling evaluation of the simplest sensing pattern that still ensures acceptable performance. This knowledge is crucial, particularly in the physical implementation of gripper sensorization, where bulky sampling hardware may burden the robotic system. Moreover, the randomized object pose during training ensures that no single grasp configuration dominates the performance. Additionally, objects with similar geometry may yield similar tactile signatures, leading to classification ambiguities. While our iterative approach compensates for such variations over successive grasps, the interpretability of such tactile information remains an interesting subject for further investigation.

Given the dynamic nature of interactions with a compliant body, the task cannot be adequately approximated as a sequence of static events. Incorporating temporal correlation of sensory information allows for capturing the evolution of tactile signals over time. Indeed, temporal cues aid in discovering multi-modal temporal patterns that enhance the model ability to discriminate and track contact events. Furthermore, the benefits of sensory multi-modality for the task are demonstrated. The fusion of information from multiple sources enhances object classification performance, underscoring the importance of leveraging both kinaesthetic and contact sensing to provide a comprehensive perceptual state of the finger during interaction with the object. While model performance is comparable in simulation and the physical case, real-world applications exhibit lower accuracy due to signal noise and relative sensors motion.

Concerning robot deployability, the proposed framework was developed and validated using a specific commercial soft gripper with bending and pressure sensors. However, the architecture and the iterative classification strategy are not inherently tied to this particular hardware configuration. In principle, the approach is generalizable to other soft robotic grippers and alternative sensing modalities. Sensor characteristics like resolution, noise, and spatial arrangement can influence absolute performance metrics. Nevertheless, the core approach of fusing multi-modal tactile data and utilizing temporal correlations is still widely applicable with proper calibration and tuning. Specifically to our study, the sensor calibration served as an initial quality control measure, but generic sensor performance might change and be affected by drifts. So, incorporating an intermediate dynamic calibration in future implementations could help mitigate any gradual variations in sensor characteristics, thereby allowing the classification model to remain unchanged.

Finally, the implementation of the iterative tactile object recognition enables the robotic agent to estimate the confidence during object classification. It facilitates improving classification accuracy by providing novel information over time when the current estimation is deemed poor. Such capability is crucial for blind tactile object recognition, where some grasps may not be informative due to incorrect relative positioning between the gripper and the object, resulting in uninformative sensory signals.

VII. Conclusion and Future Works

This work presents a learning-based algorithm for tactile object recognition implemented on a soft gripper. It integrates multi-modal sensors to gather data during object grasping in both simulation and physical implementations. A novel LSTM-based architecture is devised for dual-stage recognition, initially identifying object properties and subsequently determining the object class. This architecture effectively merges multi-modal sensing, yielding more precise and accurate results than other classification algorithms.

Its generalizability offers promising opportunities for future implementations in a variety of application domains. However, despite its competitive performance, our framework has revealed that sensor resolution is a critical factor influencing the system

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

accuracy. The integration of higher-resolution tactile sensors and an improved sensor placement will be a key focus in our future work.

Future improvements of the algorithm include implementing a strategy for self-detection of novel grasped objects, potentially shifting towards unsupervised learning. Additionally, enhanced discrimination between objects could be achieved by incorporating additional sensory modalities. We plan also to investigate the effect of dynamic interactions, by observing how controlled shaking motions could generate additional tactile cues. They might capture subtle variations in object properties that are not evident from static measurements. Lastly, active perception could be implemented to enhance performance and reduce time complexity, enabling efficient sampling by informed action selection.

References

[1] Z. Xia, Z. Deng, B. Fang, Y. Yang, and F. Sun, "A review on sensory perception for dexterous robotic manipulation," *Int. J. Adv. Robot. Syst.*, vol. 19, no. 2, 2022, Art. no. 17298806221095974.

[2] E. Donato, M. Lo Preti, L. Beccai, and E. Falotico, "Sensorimotor control strategies for tactile robotics," 2025, *arXiv:2501.09468*.

[3] J. Jorgensen, K. Bojesen, and E. Jochum, "Is a soft robot more 'natural'? exploring the perception of soft robotics in human–robot interaction," *Int. J. Social Robot.*, vol. 14, pp. 95–113, 2022.

[4] J. Shintake, V. Cacucciolo, D. Floreano, and H. Shea, "Soft robotic grippers," *Adv. Mater.*, vol. 30, no. 29, 2018, Art. no. 1707035.

[5] N. Sinatra, C. Teeple, D. Vogt, K. Parker, D. Gruber, and R. Wood, "Ultragentle manipulation of delicate structures using a soft robotic gripper," *Sci. Robot.*, vol. 4, no. 33, 2019, Art. no. eaax5425.

[6] J. Zhou, S. Chen, and Z. Wang, "A soft-robotic gripper with enhanced object adaptation and grasping reliability," *IEEE Robot. Automat. Lett.*, vol. 2, no. 4, pp. 2287–2293, Oct. 2017.

[7] A. Pagoli, F. Chapelle, J. Corrales, Y. Mezouar, and Y. Lapusta, "A soft robotic gripper with an active palm and reconfigurable fingers for fully dexterous in-hand manipulation," *IEEE Robot. Automat. Lett.*, vol. 6, no. 4, pp. 7706–7713, Oct. 2021.

[8] C. Laschi, T. Thuruthel, F. Lida, R. Merzouki, and E. Falotico, "Learning-based control strategies for soft robots: Theory, achievements, and future challenges," *IEEE Control Syst. Mag.*, vol. 43, no. 3, pp. 100–113, Jun. 2023.

[9] E. Falotico et al., "Learning controllers for continuum soft manipulators: Impact of modeling and looming challenges," *Adv. Intell. Syst.*, vol. 7, 2024, Art. no. 2400344.

[10] P. Khin, J. Low, M. J. Ang, and C. Yeow, "Development and grasp stability estimation of sensorized soft robotic hand," *Front. Robot. AI*, vol. 8, 2021, Art. no. 619390.

[11] Z. Zhou et al., "A sensory soft robotic gripper capable of learning-based object recognition and force-controlled grasping," *IEEE Trans. Automat. Sci. Eng.*, vol. 21, no. 1, pp. 844–854, Jan. 2024.

[12] T. Thuruthel, B. Shih, C. Laschi, and M. Tolley, "Soft robot perception using embedded soft sensors and recurrent neural networks," *Sci. Robot.*, vol. 4, no. 26, 2019, Art. no. eaav1488.

[13] N. Kushawaha, L. Fruzzetti, E. Donato, and E. Falotico, "SynapNet: A complementary learning system inspired algorithm with real-time application in multimodal perception," *IEEE Trans. Neural Networks Learn. Syst.*, early access, Aug. 27, 2024, doi: 10.1109/TNNLS.2024.3446171.

[14] B. Anf and C. Yeow, "A learning-based approach to sensorize soft robots," *Soft Robot.*, vol. 9, no. 6, pp. 1144–1153, 2022.

[15] H. Wang, M. Totato, and L. Beccai, "Toward perceptive soft robots: Progress and challenges," *Adv. Sci.*, vol. 5, no. 9, 2018, Art. no. 1800541.

[16] F. Wang, Q. Jiang, and J. Li, "Shape sensing for continuum robots using FBG sensors array considering bending and twisting," *IEEE Sensors J.*, vol. 24, no. 2, pp. 1546–1554, Jan. 2024.

[17] P. Roberts, M. Zadan, and C. Majidi, "Soft tactile sensing skins for robotics," *Soft Robot.*, vol. 2, pp. 343–354, 2021.

[18] L. Wang, J. Lam, X. Chen, J. Li, R. Zhang, and Y. Su, "Soft robot proprioception using unified soft body encoding and recurrent neural network," *Soft Robot.*, vol. 10, no. 4, pp. 825–837, 2023.

[19] C. Tawk, E. Sariyildiz, and G. Alici, "Force control of a 3D printed soft gripper with built-in pneumatic touch sensing chambers," *Soft Robot.*, vol. 9, no. 5, pp. 970–980, 2022.

[20] P. Wang et al., "Sensing expectation enables simultaneous proprioception and contact detection in an intelligent soft continuum robot," *Nature Commun.*, vol. 15, 2024, Art. no. 9978.

[21] S. Luo, J. Bimbo, R. Dahiya, and H. Liu, "Robotic tactile perception of object properties: A review," *Mechatronics*, vol. 48, pp. 54–67, 2017.

[22] Y. Wei et al., "Human tactile sensing and sensorimotor mechanism: From afferent tactile signals to efferent motor control," *Nature Commun.*, vol. 15, 2024, Art. no. 6857.

[23] M. Jin, H. Gu, S. Fan, Y. Zhang, and H. Liu, "Object shape recognition approach for sparse point clouds from tactile exploration," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2013, pp. 558–562.

[24] H. Liu, X. Song, T. Nanayakkara, L. D. Seneviratne, and K. Althoefer, "A computationally fast algorithm for local contact shape and pose classification using a tactile array sensor," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2012, pp. 1410–1415.

[25] J. Huang and A. Rosendo, "Variable stiffness object recognition with a CNN-bayes classifier on a soft gripper," *Soft Robot.*, vol. 9, no. 6, pp. 1220–1231, 2022.

[26] S. Luo, W. Mou, K. Althoefer, and H. Liu, "Novel tactile-sift descriptor for object shape recognition," *IEEE Sensors J.*, vol. 15, no. 9, pp. 5001–5009, Sep. 2015.

[27] T. Corradi, P. Hall, and P. Iravani, "Bayesian tactile object recognition: Learning and recognising objects using a new inexpensive tactile sensor," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2015, pp. 3909–3914.

[28] E. Kirby, R. Zenha, and L. Jamone, "Comparing single touch to dynamic exploratory procedures for robotic tactile object recognition," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 4252–4258, Apr. 2022.

[29] R. Zuo, Z. Zhou, B. Ying, and X. Liu, "A soft robotic gripper with anti-freezing ionic hydrogel-based sensors for learning-based object recognition," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 12164–12169.

[30] L. Rokach and O. Maimon, "Decision trees," in *Data Mining and Knowledge Discovery Handbook*. Boston, MA, USA: Springer, 2005, pp. 165–192.

[31] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *Proc. Move Meaningful Internet Syst. 2003, CoopIS, DOA, ODBASE*, 2003, pp. 986–996.

[32] M. Awad and R. Khanna, "Support vector machines for classification," in *Efficient Learn. Mach: Theories, Concepts, and Applications for Engineers and System Designers*. Berkeley, CA, USA: Apress, 2015, pp. 39–66.

[33] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, 2020.

[34] C. D. Sutton, "Classification and regression trees, bagging, and boosting," in *Handbook of Statistics*. Amsterdam, the Netherlands: Elsevier, 2005.

[35] J. Speiser, M. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Syst. Appl.*, vol. 134, pp. 93–101, 2019.

[36] Monika, M. Kumar, and M. Kumar, "XGBoost: 2D-object recognition using shape descriptors and extreme gradient boosting classifier," in *Proc. Comput. Methods Data Eng.*, 2020, pp. 207–222.

[37] M. A. Graule, C. B. Teeple, T. P. McCarthy, G. R. Kim, R. C. St. Louis, and R. J. Wood, "SoMo: Fast and accurate simulations of continuum robots in complex environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 3934–3941.

[38] SoftGripping, "Product catalogue," 2023. [Online]. Available: https://softgripping.com/