GmClass: Granular Material Classification through Force Feedback of Robotic Manipulation

Zeqing Zhang[1,2], Guanqi Chen[1], Wentao Chen[1], Ruixing Jia[1], Liangjun Zhang[3] and Jia Pan[1,2]

***Abstract*— We propose *GmClass*, a force-based classifier for amorphous granular materials (GMs). Inspired by human perception in the dark, we use force signals from probe-granule interactions along a spiral path. Our multimodal approach combines frequency-domain force data with descriptive text labels, achieving 84.10% classification accuracy. It outperforms traditional supervised learning by 10% and supervised contrastive learning by over 40%, highlighting the benefits of incorporating text modality. Additionally, our method handles unseen particles effectively without fine-tuning. Videos are available at https://sites.google.com/view/gm-class.**

I. Introduction

Understanding the composition of celestial bodies like the Moon or Mars is crucial for human migration. By developing an algorithm to classify GMs, we can gain insights into their properties. This knowledge is vital for planning space missions and ensuring human safety. Traditional vision-based methods face challenges in extreme environments, so, inspired by how humans perceive objects through touch (Fig. 1-(a)), we propose using a robotic arm to interact directly with GMs for classification (Fig. 1-(b)).

The contact model in GMs is highly complex, with mechanical signals originating from force chains between tools and granules [1]. Simplified models based on experimental observations exist [2], [3], but various factors, such as packing density and particle size, influence the contact forces [4], as observed in Fig. 1-(e). These factors lead to variations within classes and similarities between classes, as shown in Fig. 1-(d), challenging GM classification based solely on force feedback. Also, the manner of interaction further affects signal acquisition, which is not accounted for in simplified models.

To this end, we propose ***GmClass***, a classifier based on multimodal supervised contrastive learning (MSCL), which consists of two branches: one extracts features from force signals acquired during GM manipulation, and the other focuses on text information related to the names of GM. This approach effectively learns distinctive granule features by reducing the distance between samples of the same class and increasing the distance between different classes in the feature space. Specifically, we convert the time-series force data into a single frequency spectrum, enhancing class separation. The high-dimensional textual information helps minimize intra-class variance. We also explore the impact of interaction motion and text prompts and model generalization. Our original contributions are:

[1]The University of Hong Kong.
[2]Center for Transformative Garment Production.
[3]Robotics and Autonomous Driving Lab, Baidu Research, USA.

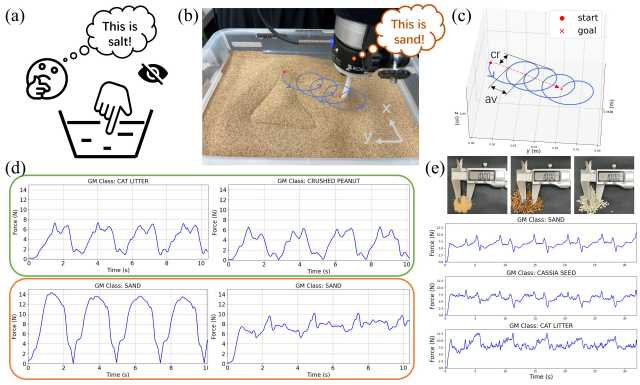


Fig. 1. (a) In the absence of vision, humans can distinguish objects by the touch of their fingers. (b) Inspired by this, we propose ***GmClass*** utilizing force feedback from the probe-granules interaction. (c) Spiral trajectory [5], defined by the circular radius (cr) and advance velocity (av). (d) The complexity of the probe-granules interaction is reflected in the high inter-class similarities (green box) and intra-class variations (orange box) in interaction forces. (e) The enlarging random errors can be observed in the force data with particle size raising.

- We propose a vision-independent classifier, ***GmClass***, using forces from the robot-GM contact to identify GMs and showcase its zero-shot capability.
- We introduce an MSCL framework with frequency and text encoders, to reduce the high variability of forces in the same GM.
- We open source codes and the dataset *GM10-ts*, which is the first work specifically designed for GMs.

II. Methodology

The key idea of our ***GmClass*** is to convert the time-series signals with significant differences into frequency-domain signals, and then leverage MSCL to enable the model to learn the matching relationship between frequencies and classes. The overall framework is illustrated in Fig. 2, whose inputs are the pair of force data and class names, while the output is the similarity between them.

Specifically, the ***GmClass*** consists of two branches: one for the frequency encoder and the other one for the text encoder. The frequency encoder is responsible for extracting frequency features using the Fast Fourier Transform (FFT) from different granules. In our implementation, we employed a one-dimensional convolutional neural network (1D CNN) [6] with different kernel sizes (5, 15, 25, and 50) to interpret the sequence data at multiple resolutions, and compared it with commonly used models good at time-series data, e.g., long short-term memory (LSTM) [7], bi-directional LSTM (BiLSTM) [8], and Transformer [9]. Also, ResNet [10] is considered. On the other branch, the text encoder is designed to extract features from textual information.

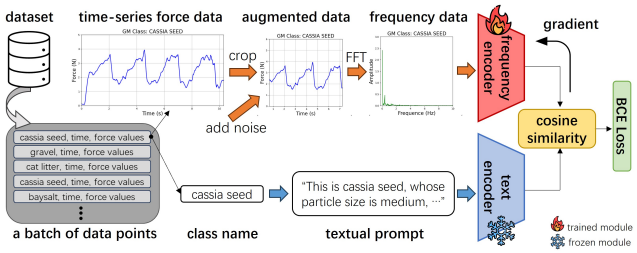


Fig. 2. The framework of ***GmClass***.

In this case, we utilized the pre-trained text encoder from Contrastive Language-Image Pre-training (CLIP) [11]. The textual information is derived from the semantic processing of particulate matter classes. During the test, we input a sequence of force data and all possible GM text labels. Therefore, it is tested whether the trained ***GmClass*** can correctly classify the types of particles from the mechanical signal.

We present our dataset, *GM10-ts* (as shown in Fig. 3), consisting of time-series force signals obtained during the robotic manipulation of 10 commonly found granules in daily life. The dataset contains 5000 data points, with 500 instances for each granule.

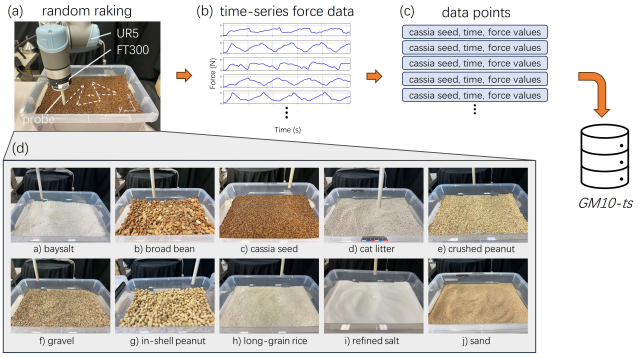


Fig. 3. Experiment setup and dataset generation for *GM10-ts*.

III. Results

In this section, we conduct extensive experiments to test ***GmClass*** and compare it with conventional supervised learning (SL) and supervised contrastive learning (SCL) methods. We also study the effects of the feature extraction model, data modality, ranking trajectory in data collection, text encoder, and textual prompt on classification accuracy.

A. Classification Results

In Tab. II, our ***GmClass*** achieves the highest classification accuracy of 84.10%. This is 40% higher than the SCL (highest at 41.40%) and about 10% higher than the SL (highest at 74.90%). We compare different feature extraction models for GM classification, including 1D CNN, ResNet, BiLSTM, LSTM, and Transformer. We find that 1D CNN and BiLSTM perform well on time-series (TS) data, while 1D CNN shows good performance on frequency-series (FS) data as well. The Transformer model exhibites similar performance on both TS and FS data. Next, we compare the classification accuracy of ***GmClass*** with SCL. We replace the text encoder with the frequency encoder from the lower branch in Fig. 2 in SCL. We also train the model using TS data without performing FFT. We use 1D CNN as both frequency and time encoders in SCL due to its good feature extraction capability. Additionally, we replace the CLIP text encoder with Sentence-BERT [12] in MSCL, therefore, we have MSCL-C and MSCL-B denoting the variants of MSCL using CLIP and Sentence-BERT, respectively. From Tab. II, we observe that MSCL achieves significantly better results than SCL. With TS data, MSCL outperforms SCL by more than 10% (MSCL-B) and more than 30% (MSCL-C). When using FS data, MSCL shows even more improvement, with an increase of approximately 30% (MSCL-B) and over 60% (MSCL-C) in performance. This highlights the importance of utilizing multiple modalities to enhance classification accuracy. The best classification result for SCL is 41.40% on TS data, while MSCL-B performs similarly on TS and FS data (52.90% and 50.50%, respectively). MSCL-C achieves a significantly higher accuracy of 74.00% on TS data, approaching the best result in SL (74.90% from 1D CNN). By incorporating FS data into MSCL-C, the classification accuracy is further boosted by approximately 10%, resulting in the highest performance of 84.10%.

TABLE I
CLASSIFICATION ACCURACY FOR VARIOUS PROMPTS.

Prompt	**(a)** 'This is <gm_name>.'	**(b)** '<gm_name>'
Acc.	**84.10%**	81.80%

Prompt	**(c)** 'This is <gm_name>, whose <property.name> is <property.value>.'				
	'particle size'	'particle shape'	'roughness'	'weight'	All
Acc.	82.30%	82.10%	82.10%	81.80%	79.70%

Prompt	**(d)** '<property.name> is <property.value>.'				
	'particle size'	'particle shape'	'roughness'	'weight'	All
Acc.	24.50%	17.80%	17.60%	22.60%	59.60%

B. Effects of Raking Trajectory

In addition, we also generate a dataset sampled from linear motion, instead of spiral trajectory, to compare the effects of different robot-particle interactions on classification results. Results based on the linear path are provided in Tab. II. SL results on FS data are inferior to TS data (Fig. 4-(a)). SCL performs similarly on both modalities, while MSCL performs better on FS data. Spiral trajectory consistently outperforms linear trajectory, with 1D CNN, BiLSTM, and MSCL showing significant improvements. In the frequency domain, models based on spiral trajectories achieve higher classification accuracy. Overall, from Fig. 4-(b)(c), it shows spiral trajectories provide richer information for classification.

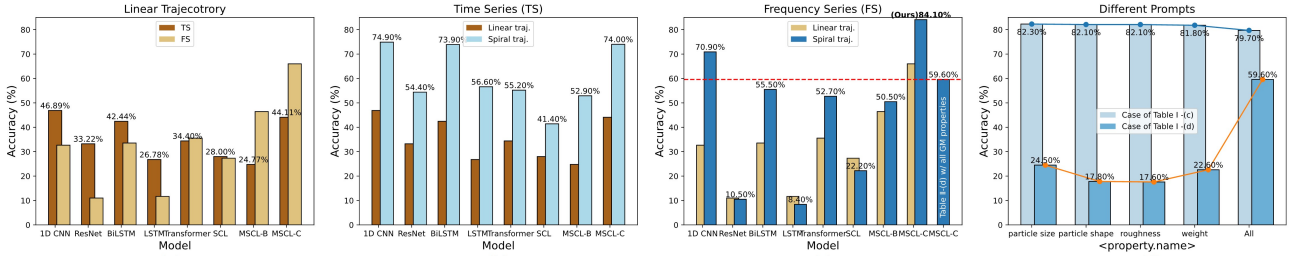
C. Effects of Text Encoder

Comparing SCL and MSCL results, the addition of semantic modality significantly improves classification accuracy. In the linear trajectory case, MSCL-C outperforms MSCL-B by 20% from Tab. II. Considering the spiral trajectory, MSCL-C achieves 20% improvement in the time domain and 30% improvement in the frequency domain, reaching the best performance of 84.10%. This shows that CLIP's text encoder creates a feature space more suitable for classification tasks through contrastive learning compared to BERT.

TABLE II

CLASSIFICATION ACCURACY FOR DIFFERENT LEARNING METHODS, MODELS AND MODALS.

Method	Supervised Learning										SCL		Multimodal SCL			
Model	1D CNN		ResNet		BiLSTM		LSTM		Transformer		1D CNN		1D CNN - BERT		1D CNN - CLIP	
Modal	TS	FS	TS	FS	TS	FS	TS	FS	TS	FS	TS	FS	TS-Text	FS-Text	TS-Text	FS-Text
Linear traj.	46.89%	32.67%	33.22%	11.00%	42.44%	33.56%	26.78%	11.67%	34.40%	35.56%	28.00%	27.33%	24.77%	46.44%	44.11%	66.00%
Spiral traj.	74.90%	70.90%	54.40%	10.50%	73.90%	55.50%	56.60%	8.40%	55.20%	52.70%	41.40%	22.20%	52.90%	50.50%	74.00%	**84.10%**



(a) (b) (c) (d)

Fig. 4. Classification results of various models, data modalities, raking trajectories in data collection, and prompts. (a) Using TS and FS data from the linear trajectory. (b) Using TS data from linear and spiral trajectories. (c) Using FS data from linear and spiral trajectories. (d) Using different prompts based on FS data from the spiral trajectory.

D. Effects of Prompt

1) GM Name: The best performance, 84.10%, is achieved when the prompt is "This is" followed by GM class names (Tab. I-(a)). Directly sending GM class names to the text encoder results in a slight drop in classification accuracy (81.80% from Tab. I-(b)), aligning with previous findings on the impact of different prompts [11].

2) GM Name + GM Property: In addition, we enhance the textual information by considering granular properties such as particle size, shape, roughness, and weight (see Tab. III). The experimental results in Tab. I-(c) and Fig. 4-(d) show that incorporating a single granule property has a similar impact, with accuracy fluctuating by 0.5% (81.80% to 82.30%), all lower than the case without GM property. Incorporating all four properties further decreases classification accuracy.

3) GM Property: We also evaluate a scenario where the prompt provides granule property features instead of class information (Tab. I-(d)). We mask the GM class during training and assess the model's accuracy in predicting property values. Results in Fig. 4-(d) show a sharp drop in classification accuracy when the <gm_name> cue is missing, especially when training on a single property. However, including prompts with all four properties significantly improves accuracy, reaching 59.60% (Fig. 4-(d)). This result ranks third among models trained on FS data from the spiral trajectory, following the 1D CNN model in SL (70.90%) and MSCL-C (84.10%) (Fig. 4-(c)). Comparing Tab. I-(c) and (d), or Fig. 4-(d), reveals the positive impact of GM class names on granule classification. GM names provide direct cues and the CLIP model's text encoder captures semantic information, enhancing classification capabilities by differentiating inter-class similarities.

E. Zero-shot Classification

The unexpected outcome of MSCL in Tab. I-(d) surpassing some SL models prompts further investigation into the model's zero-shot transfer learning. We conduct zero-shot classification experiments using a pre-trained model of Tab. I-(d). Unseen GM classes with *new* textual descriptions (Tab. IV) are defined. From Tab. IV, the pre-trained model shows zero-shot transfer capability, that effectively classifies force information of unseen particles (in the frequency domain) into corresponding property descriptions, e.g., pearl rice, indicating a preliminary understanding of the correlation between mechanical signals and GM properties. However, the model's resolution performance is subpar for certain granule types like mung bean and red bean, as observed in the experiments.

TABLE III

PROPERTY VALUES OF GM, WHICH WOULD BE USED AS ADDITIONAL TEXT INFORMATION IN THE TRAINING PROCESS OF MSCL.

<gm_name>	<property.name>			
	'particle size'	'particle shape'	'roughness'	'weight'
baysalt	medium	rough	non-circular	heavy
broad bean	large	smooth	non-circular	medium
cassia seed	medium	smooth	non-circular	medium
cat litter	medium	rough	circular	heavy
crushed peanut	small	rough	non-circular	medium
gravel	medium	rough	non-circular	heavy
in-shell peanut	large	rough	non-circular	light
long-grain rice	medium	smooth	non-circular	heavy
refined salt	small	rough	circular	heavy
sand	small	rough	circular	heavy

TABLE IV

ZERO-SHOT CLASSIFICATION USING FOUR GRANULAR PROPERTIES TO DESCRIBE UNSEEN PARTICLES WITHOUT GM NAMES. THE *italics* DENOTE THE UNSEEN PROPERTY VALUES IN THE TRAINING.

<gm_name>	<property.name>				Acc.
	'particle size'	'particle shape'	'roughness'	'weight'	
pearl rice	medium	medium	*quite smooth*	*little heavy*	84.20%
small macaroni	large	*irregular*	rough	*very light*	77.20%
large macaroni	*very large*	*irregular*	rough	*very light*	68.80%
sunflower seed	large	*irregular*	smooth	*quite light*	63.20%
mung bean	medium	*regular*	*quite smooth*	*much heavy*	45.20%
red bean	large	*regular*	*quite smooth*	*much heavy*	19.00%

REFERENCES

[1] J. Peters, M. Muthuswamy, J. Wibowo, and A. Tordesillas, "Characterization of force chains in granular material," *Physical review E*, vol. 72, no. 4, p. 041307, 2005.

[2] R. D. Maladen, Y. Ding, C. Li, and D. I. Goldman, "Undulatory swimming in sand: subsurface locomotion of the sandfish lizard," *science*, vol. 325, no. 5938, pp. 314–318, 2009.

[3] Y. Zhu, L. Abdulmajeid, and K. Hauser, "A data-driven approach for fast simulation of robot locomotion on granular media," in *2019 international conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 7653–7659.

[4] R. Albert, M. Pfeifer, A.-L. Barabási, and P. Schiffer, "Slow drag in a granular medium," *Physical review letters*, vol. 82, no. 1, p. 205, 1999.

[5] Z. Zhang, R. Jia, Y. Yan, R. Han, S. Lin, Q. Jiang, L. Zhang, and J. Pan, "GRAINS: Proximity sensing of objects in granular materials," *arXiv preprint arXiv:2307.05935*, 2023.

[6] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," *Mechanical systems and signal processing*, vol. 151, p. 107398, 2021.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural networks*, vol. 18, no. 5-6, pp. 602–610, 2005.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[11] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PMLR, 2021, pp. 8748–8763.

[12] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019.