ABSTRACT This study aimed to anticipate fractures of fragile food during robotic food manipulation. Anticipating fractures allows a robot to manipulate ingredients without irreversible failure. Food fracture models investigated in food texture fields explain the properties of fragile objects well. However, they may not directly apply to robot manipulation due to the variance in physical properties even within the same ingredient. To this end, we developed a fracture-anticipation system with a tactile sensing module and a simple recurrent neural network. The key idea was to allow the robot to break ingredients during training-sample collection. The timing of fractures was identified via simple signal processing and used for supervision. We performed real robot experiments with three typical fragile foods: tofu, potato chips, and bananas. As the first step toward flexible fragile-object manipulation, we evaluated the proposed method for the fundamental task of object picking. The method successfully grasped the fragile foods without fractures in an online demonstration. In an offline evaluation, the method predicted the fractures with a recall of approximately 80% for all ingredients with 60 breaking trials. We believe that our method can be used to avoid breakage in other types of food manipulation, e.g., holding, pressing, and rolling.

INDEX TERMS Robotic food manipulation, fracture anticipation, tactile sensing.

I. INTRODUCTION Food is a central element of human life and has various industrial demands. In the context of population aging, robotics applications such as home cooking assistance [1] and food-preparing robots in factories and restaurants [2], [3] have high expectations; however, food manipulation is non-trivial because food is edible and thus fragile.

Fracture-free manipulation is achieved by avoiding excessive force when gripping the object. One popular approach is slip detection [4], in which the minimum force needed to move an object is determined by detecting a current slip. In collaboration with these minimum-force identification techniques, knowing the margin force for fracture is beneficial for robotic manipulation. To this end, we propose a new task of fracture anticipation, which identifies the maximum force in fragile-object manipulation before fracturing the object.

Food fractures have been investigated mainly in food texture studies [5], with the aim of establishing models that explain how fracture occurs, according to the relationship between force and deformation. Such models explain the properties of fragile objects well. However, studies have revealed the difficulty of applying such knowledge to robotics applications due to the significant variance in physical properties even within the same ingredient [6].

To overcome the significant differences in physical properties, we designed a learning-based framework for robots to anticipate fractures. It consists of tactile sensors equipped with a two-finger gripper, which determines the physical properties of individual foods. Human infants explore the physical properties of objects by touching them [7]. Similarly,
we let the robot break the food ingredients to determine when the fracture occurs (Fig. 1). We adopted a long short-term memory (LSTM) classifier [8] and an LSTM-based Seq2Seq model [9] to prove the concept of the proposed framework.

We experimentally evaluated the framework with real food data and a robot. Three fragile ingredients were adopted: tofu, potato chips, and peeled bananas. In an online demonstration, we applied the proposed framework to food ingredient picking. The robot trained by our framework successfully grasped the objects without fractures. In an offline evaluation, we evaluated the fracture-anticipation model using samples with known fracture points. We prepared the test samples with within-category differences in physical properties (e.g., different poses and stiffnesses) to evaluate the robustness of the learning-based approach.

The contributions of the study are threefold:
- We propose a new task of fracture anticipation, which aims to estimate the maximum force during object manipulation.
- We developed a data-collection framework by breaking food objects to train the anticipation model. At the data-collection stage, the robot breaks fragile objects, and the fracture timing is automatically identified via simple signal processing. We experimentally validated the concept using simple LSTM models trained with the identified timing.
- We applied the fracture-anticipation model to robotic food-picking tasks. The robot successfully grasped the fragile foods without causing fractures.

The remainder of this paper is organized as follows. Section II introduces related studies. Sections III and IV present the problem formulation and the proposed method. Sections V and VI present our experiments. Section VII discusses the experimental results. Finally, Section VIII concludes the paper.

II. RELATED WORKS

This section introduces studies related to robotic food manipulation and fracture analysis for robot applications.

A. ROBOTIC FOOD MANIPULATION

Robots have performed various food-handling tasks, such as picking [10], [11], [12], cutting [6], rolling pizza dough [13], and flipping a pancake [14], some of which require careful manipulation to avoid breaking ingredients. Many physically soft robotic grippers that allow large deformation have been developed to adapt to food products that vary widely with regard to size, texture, weight, fragility, and shape. Magnetorheological fluids [15], viscoelastic fluids [10], [16], pneumatic fluids [17], jamming structures [18], and elastomers [19], [20], [21] have been used. Although these grippers are suitable for grasping fragile food ingredients, they are not used for other food manipulation. Meanwhile, we use a universal two-finger gripper and tactile sensors [22], which will be applied to various food manipulation.

B. FRAGILE-OBJECT MANIPULATION WITH TACTILE SENSORS

Many studies have addressed the manipulation of fragile objects, including food, using tactile sensors. Romano et al. proposed a heuristic-based grasping strategy inspired by human behavior using tactile sensors and an accelerometer [23]. Misimi et al. and Lillienstjold et al. proposed learning from human demonstrations to grasp food objects [24], [25]. Additionally, feedback control systems using slip detection [26], [27], [28] have been developed. Other papers proposed extracting features or qualities of foods by breaking or touching them [29], [30], [31], [32]; however, these properties were not used for fracture anticipation.

Researchers have proposed detecting the beginning of fracture of fragile foods in a rule-based manner [16] or by using the estimation error of the pressures in fingers based on polynomial models [10] for robotic food manipulation. The proposed task of fracture anticipation can be used in collaboration with these methods, where the detected fracture is used as supervision.
III. DEFINITION OF FRACTURE AND FRACTURE-ANTICIPATION PROBLEM

To explain our framework in detail, we present a definition of food fracture and formulate the fracture-anticipation problem in this section.

A. DEFINITION OF FRACTURE

In this study, a robot arm with a parallel gripper and two distributed tactile sensors was used. Each tactile sensor had 16 taxels. Fig. 2 shows the signals measured by the 16 taxels in one of the tactile sensors when the robot closes its gripper and breaks food ingredients. The horizontal axis indicates time, and the vertical axis indicates the norm of the three-axis force signal. The line colors correspond to the locations of the taxels. As the forces on the taxels increased, the peak force appeared at different times for the different taxels. Although identifying the exact moment of fracture was difficult, we noted that fracture could only be observed visually after the forces had peaked. We observed similar tendencies for tofu, potato chips, and bananas. Thus, in this study, fracture was considered to occur when the first peak force appeared among the 32 taxels of the two tactile sensors. In this breaking trial, fracture occurred at approximately 5.5, 2.5, and 5.0 s for tofu, potato chips, and banana, as shown in Fig. 2. Our framework is not limited to this simple definition of fracture; it is compatible with more sophisticated fracture detection methods [10], [16].

B. PROBLEM FORMULATION

Fig. 3 shows the mathematical notation for the fracture-anticipation problem. The robot must stop before it creates a crack or fracture during food manipulation. To ensure this, we stop the robot’s motion at the timestep $T_p - m$, where $T_p$ represents the timing of the fracture (the peak time) and $m$ is a safety margin. Let $t_w = t - w$ be the first timestep of the input observation and $t_δ = t + δ$ be the target timestep of fracture prediction. Then, the fracture-anticipaton problem is formulated as $y_{t_δ} = f(X_{t_w,t})$, where $X_{t_w,t} = \{x_{t_w}, \ldots, x_t\}$ is a sequence of observations from the tactile sensors and $y_{t_δ}$ is a binary value that indicates whether the robot exceeds the fracture timing (True if $T_p - m \leq t_δ$, False otherwise).

IV. LEARNING-BASED FRACTURE ANTICIPATION FOR ROBOTIC MANIPULATION

This section describes the fracture-anticipation network and its application to robotic manipulation.

A. FRACTURE-ANTICIPATION NETWORK

We aimed to predict the fracture at least $δ$ timesteps in advance. To achieve this, we adopted two simple recurrent neural network (RNN) models: a simple LSTM classifier and a Seq2Seq model [9], which are denoted as Proposed1 and Proposed2, respectively.

Let $E : X_{t_w,t} \rightarrow \{z_t, h^E_t\}$ be an LSTM encoder, where $z_t$ represents the output of the LSTM and $h^E_t$ represents the
hidden state of $\mathcal{E}$ at time $t$. The Proposed1 model consists of $\mathcal{E}$ and a dense layer $\mathcal{M}: z_t \rightarrow \hat{y}_{ts}$, where $\hat{y}_{ts}$ is an estimate of the binary fracture state $y_{ts}$ (Fig. 4).

The Proposed2 model predicts forthcoming observations $X_{t+1:t+k}$ in addition to $y_{ts}$. It consists of $\mathcal{E}$ and an LSTM decoder $\mathcal{D}: \{x_t, h_E^t\} \rightarrow \{\hat{X}_{t+1:t+k}, \hat{y}_{t+k}\}$ (Fig. 5), where $\hat{X}$ (or $\hat{x}$ for an element in $\hat{X}$) is an estimate of observation $X$ (or $x$). The multi-task strategy of Proposed2 may help guide the training and enhance the explainability by estimating $X_{t+1:t+k}$. Generally, the uncertainty of future events gradually increases. Hence, the sequential estimation may reduce the difficulty of predicting subsequent events. Additionally, anticipating $\hat{X}_{t+1:t+k}$ is useful for understanding how the model made a decision of $\hat{y}_{ts}$.

To achieve the above intention, we trained Proposed1 to minimize the loss function $L_{ce}(\hat{y}_{ts}, y_{ts})$, where $L_{ce}$ represents the binary cross entropy. We trained Proposed2 using the following loss function:

$$L(X_{t:t+k}, Y_{t+1:t+k}) = \sum_{k=1}^{\delta} L_{ce}(\hat{y}_{t+k}, y_{t+k}) + L_{mse}(\hat{x}_{t+k}, x_{t+k}),$$  

where $L_{mse}$ represents the mean squared error (MSE), and $\hat{x}_{t+k}$ represents a noise-removed observation. We calculate $L_{ce}$ not only for $y_{ts}$ but for the entire sequence of $Y_{t+1:t+k}$ to optimize identical latent feature extraction, regardless of the difference in the timestep.

To calculate $L_{mse}$, we use a simple moving-average operation to remove the noise from $x_{t+k}$ and obtain $\hat{x}_{t+k}$. Minimizing the MSE implicitly reduces this noise, as predicting a random value is theoretically impossible, and the MSE loss models it as a normal distribution. However, a noisy ground truth may lead to overfitting. We applied explicit noise reduction to prevent overfitting.

### B. ROBOTIC FOOD PICKING APPLICATION

We tested our framework with a food-picking task, which is one of the fundamental tasks in food manipulation. In this context, fracture anticipation helps the robot stop the grasping motion immediately before fracture occurs. This not only prevents the robot from breaking foods but also ensures that the robot grasps them as firmly as possible for subsequent manipulations. In the picking task, the robot closes its gripper while collecting tactile data to anticipate fracture in every timestep. When a fracture is predicted, the robot stops closing the gripper and lifts the gripper upward with the picked object.

### V. EXPERIMENTAL SETUP

We performed experiments with real food ingredients and a robot to validate our framework for sample collection and learning-based fracture anticipation. The objective was to confirm that our method could accurately predict fracture, allowing the robot to pick up different types of fragile objects without fracture.

![FIGURE 4. Simple LSTM classifier Proposed1 for fracture anticipation. We implemented a model with LSTM $\mathcal{E}$ and a dense layer $\mathcal{M}$.](image4.png)

![FIGURE 5. Seq2Seq model Proposed2, as another fracture-anticipation approach. We implemented the model using LSTM networks for $\mathcal{E}$ and $\mathcal{D}$.](image5.png)

### B. DATASET GENERATION

To train and test our model offline, we collected a dataset. During the data collection, the robot repeated the following motions:

1) Move its position to above the target.
2) Open the gripper and lower it to a position where it touches the table’s surface.
3) Completely close the gripper.
4) After waiting for 3 s, open it to release the target object.
5) Return to the initial position.

Because our model was structured to predict forthcoming tactile values at every timestep, we needed ground-truth values.

---

1 These textures correspond to ductile and brittle fractures, respectively.
tactile data to compare the predicted tactile values from LSTM cells with the ground truth. In addition, to create the ground-truth labels when the target object broke during training and offline evaluation, we automatically detected the fracture point (i.e., the force peak point in Fig. 3).

The tactile data obtained by our system were not always recorded at regular intervals, included noise, and could lead to incorrect peak-point detection, making the prediction task difficult. Therefore, we preprocessed the tactile data by resampling the data at regular intervals. Specifically, we first resampled the data at equal intervals of 0.008 s via linear completion and then calculated the L2 norm of the data from the 32 taxels. Then, we converted the L2-norm data into the frequency domain and cut off the frequencies with values below the threshold to reduce the amount of noise. The threshold was set as $5 \times 10^{-5}$. After the noise was removed from the tactile data, we detected each peak point and defined the first peak among the 32 taxels as the ground-truth force peak. To make the peak detection more reliable, we removed the tactile data where the timing of the peak force and the visibly observed fracture are significantly different. We ignored peak points as noise when their L2 norm was <0.004. Finally, we set $T_p$ as the first force peak among the 32 taxels and $y_t^p$ as True under the condition of $T_p - m \leq t_s$. These procedures were implemented using SciPy libraries, i.e., linear interpolation, fast Fourier transform, inverse fast Fourier transform, and signal libraries.

We recorded tactile data for tofu, potato chips, and banana in 60 trials for each object. Diverse samples were used, with different brands, shapes, and positions (Fig. 7). We aimed in 60 trials for each object. Diverse samples were used, with different brands, shapes, and positions (Fig. 7). We aimed to reduce alignment-dependent variance, we used the difference between two consecutive observations as $x_t$.

![Robot arm](image1.png)

**FIGURE 6.** Our experimental setup. We used a parallel gripper and two distributed tactile sensors.

![Tofu](image2.png)

**FIGURE 7.** Setup for picking up objects. We prepared tofu, potato chips, and bananas of several brands as examples of slippery or thin fragile objects and placed them in different shapes and poses.

two different brands. We placed them in a concave-upward pose (Pose A) or an upside-down pose (Pose B). For bananas, we prepared three different brands with different levels of maturity (soft and hard). They were always round-sliced, with thicknesses ranging from approximately 0.7 to 1.5 cm. The round surface was in contact with the ground and gripper, but the cut surfaces were untouched (Pose A).

**C. LEARNING SETUP**

To organize the models in Figs. 4 and 5, we used two-layered LSTM networks for the encoder $E$ and decoder $D$. All the layers had 32 hidden units. The output $z_t$ of $E$, which was the input to the dense layer $M$, was set as a 32-dimensional vector.

With the LSTM encoder, we must flatten the $(4 \times 4 \times 3)$-dimensional signals into a 96-dimensional vector. Thus, ConvLSTM [35] networks may be considered more efficient. However, their effect may be small because the size of the tactile map $(4 \times 4)$ is similar to that of the convolution kernel $(3 \times 3)$. Hence, we implemented the model with the simple RNN architecture. Similarly, we avoided using a transformer owing to the limited number of training samples. $D$ had two outputs: a 96-dimensional vector $\hat{y}_{t+k}$ and a logit. We processed the logit with the sigmoid function and obtained $\hat{y}_{t+k}$.

The absolute value of the signals has little meaning under our robot platform because it depends on the alignment between sensing points and the contact surfaces. Thus, to reduce alignment-dependent variance, we used the difference between two consecutive observations as $x_t$. 
We trained a model for 200 epochs using the stochastic gradient descent (SGD) optimizer with a momentum of 0.5. The initial learning rate was set as 0.01, and the rate decayed by 0.8 every 30 epochs. The other parameters for SGD remained at their default values. We experimentally set the safety margin as \( m = 8 \). We used PyTorch 1.9.0 and CUDA 11.1 for the training.

**D. BASELINE AND EVALUATION FOR ROBOTIC PICKING TASK**

We compared our method with a manually designed baseline for the picking task. We expected our method to be more successful than the baseline in picking up the ingredients without fractures.

The baseline used a gradient: the difference of the consecutive observations divided by the time difference. The robot stopped when the gradient exceeded a predetermined threshold; thus, it stopped before collapsing the resisting structure of the object. The threshold was manually tuned via trial-and-error for each food: 0.02 for tofu, 0.05 for bananas, and 0.13 for potato chips.

We tested the different methods with 25 trials for each object. For tofu, we conducted 20 trials for \{Brand1, Brand2\} \times \{Pose A, Pose B\} (five trials for each condition) and five additional trials with the softer tofu (Brand2) cut into a cube with 50 mm. This was an unseen shape for the trained model.

For potato chips, we performed five trials with Pose A for two brands and five with Pose B for one brand (15 trials). In addition, as an unseen location, we shifted the object along the tactile sensor’s horizontal axis for five trials.

For bananas, we performed five trials for each of the three brands with the same pose as the training data (15 trials total). Ten additional trials involving unseen poses were performed: five trials with the cut surface(s) in contact with the gripper (Pose B) and five trials with the cut surface(s) in contact with the ground (Pose C). Poses B and C were unseen.

**VI. RESULTS**

This section presents experimental results of online object picking and offline fracture anticipation.

**A. PICKING PERFORMANCE**

Table 1 presents the picking performance achieved using a rule-based baseline, Proposed1, and Proposed2 under different conditions (i.e., brands, shapes, and poses). We defined successes as cases where the robot picked up and placed the object without causing cracks or dropping the object; other cases were defined as failures. We used the trained models of Proposed1 and Proposed2 that exhibited the highest accuracies among the five holds described in Section VI-B. For picking up bananas, the number of successes was comparable among the baseline and the proposed methods. However, our methods were more successful than the baseline for tofu and potato chips. Overall, our method achieved success rates of \( \geq 80\% \) for all objects. Fig. 8 shows examples of the successful and failed cases of picking up objects.

**B. FRACTURE-ANTICIPATION PERFORMANCE**

Table 2 presents the prediction accuracies of our models, which were used to evaluate the model performance. Here, Accuracy represents the percentage of the agreement between the predicted labels and the ground truth. Precision represents the percentage of timesteps with True (i.e., the target object was fractured) labels among the timesteps the model predicted as True. Recall represents the percentage of timesteps predicted as True by the model among the timesteps with the True label. The F-measure is the harmonic mean of the Precision and Recall scores. Each value was the mean score of fivefold cross-validation.

Because our goal was to grasp target objects without fracturing them, the priority for our model was to avoid passing the fracture point. Thus, the Recall score was the most important evaluation metric. As indicated by Table 2, our models achieved recall scores of approximately 80%, and the recall results for Proposed1 and Proposed2 were comparable.

**VII. DISCUSSION**

In this section, our results are discussed, along with future directions. We evaluated two RNN-based models: Proposed1 (a simple RNN-based model) and Proposed2 (the Seq2Seq model). Both models achieved picking success rates of \( \geq 80\% \) and prediction with approximately 80% recall, as shown in Tables 1 and 2. For picking up potato chips, the Seq2Seq model had slightly higher success rates than the simple RNN model. The Seq2Seq model may predict
immediate changes in fracture dynamics because it predicts future observations. Additionally, our method achieved higher success rates than the baseline because of the stable fracture anticipation. For picking up tofu and potato chips, the objects were sometimes broken in the case of the baseline because the gripper was completely closed. The baseline may have been vulnerable to varieties in the objects, causing fracture without exceeding the threshold. We believe that our fracture-anticipation method can be used to avoid breakage in other types of food manipulation, e.g., holding ingredients for cutting, pressing ingredients to investigate their quality [31], and rolling pizza dough [13].

We investigated how the objects were picked. Fig. 9 shows tactile signals for successful and failed trials of tofu picking using Proposed1 and Proposed2. As shown in Fig. 9(a), when the robot grasped tofu without fracturing it, the tactile peak point did not appear before the moment the gripper stopped moving, and the peak points were aligned for all the sensor data in the time-axis direction. Although some peak points appeared (red points), we can ignore them, because the forces were insufficient to break the tofu, as described in Section V-B. However, as shown in Fig. 9(b), when the robot grasped the tofu too strongly and fractured it, some tactile values of the gripper exceeded their peaks and began to decrease.

Fig. 10 shows the prediction accuracy with respect to the number of training data for Proposed1. We prepared 160 banana pieces: 120 for training, 20 for validation, and 20 for testing. The accuracy increased with the number of training samples; however, it stopped significantly increasing when the number of samples reached 50.

Although our method could manipulate food ingredients without fracture, there is room for further investigation and improvement. First, even though grasping was challenging as the softness varies in the same ingredients shown in this study, we need to extend the scalability of our method by tackling more various food objects and brands. Furthermore, we will investigate whether our fracture anticipation network can be used for different tasks, for example, whether the robot can continue grasping the food object without slippage and fracture even though the gripper is moved faster or shaken. Our method can learn the picking strategy more efficiently. To this end, we will consider transfer learning approaches [36], [37] for our fracture-anticipation network. Finally, it would be

### TABLE 1. Picking performance for tofu, bananas, and potato chips with different shapes and poses.

| Toy | Property | Shape and pose | Baseline Th=0.02 Success/try (crush, fail) | Proposed1 Success/try (crush, fail) | Proposed2 Success/try (crush, fail) |
|-----|----------|----------------|--------------------------------------------|------------------------------------|------------------------------------|
| Tofu | Brand1   | Cuboid / Pose A | 5/5 (0.0) | 5/5 (0.0) | 5/5 (0.0) |
| Tofu | Brand1   | Cuboid / Pose B | 1/5 (4.0) | 3/5 (0.2) | 4/5 (1.0) |
| Tofu | Brand2   | Cuboid / Pose A | 4/5 (1.0) | 5/5 (0.0) | 5/5 (0.0) |
| Tofu | Brand2   | Cuboid / Pose B | 0/5 (5.0) | 4/5 (0.1) | 2/5 (0.3) |
| Tofu | Brand2   | Cube (unseen)  | 0/5 (5.0) | 5/5 (0.0) | 5/5 (0.0) |
| Total |         |                | 10/25 (15.0) | 22/25 (0.3) | 21/25 (1.3) |

| Potato chips | Property | Pose | Baseline Th=0.13 Success/try (crush, fail) | Proposed1 Success/try (crush, fail) | Proposed2 Success/try (crush, fail) |
|--------------|----------|------|--------------------------------------------|------------------------------------|------------------------------------|
| Brand6 | Pose A | 5/5 (0.0) | 3/5 (2.0) | 5/5 (0.0) |
| Brand6 | Pose B | 0/5 (5.0) | 5/5 (0.0) | 5/5 (0.0) |
| Brand7 | Pose A | 5/5 (0.0) | 2/5 (3.0) | 5/5 (0.0) |
| Brand7 | Shift (unseen) | 6/10 (1.3) | 10/10 (0.0) | 9/10 (1.0) |
| Total |         |      | 16/25 (6.3) | 20/25 (3.5) | 24/25 (1.0) |

| Banana | Property | Pose | Baseline Th=0.05 Success/try (crush, fail) | Proposed1 Success/try (crush, fail) | Proposed2 Success/try (crush, fail) |
|--------|----------|------|--------------------------------------------|------------------------------------|------------------------------------|
| Brand3 | Pose A | 5/5 (0.0) | 5/5 (0.0) | 5/5 (0.0) |
| Brand4 | Pose B (unseen) | 5/5 (0.0) | 5/5 (0.0) | 4/5 (1.0) |
| Brand5 | Pose A | 5/5 (0.0) | 5/5 (0.0) | 5/5 (0.0) |
| Brand5 | Pose B (unseen) | 5/5 (0.0) | 5/5 (0.0) | 5/5 (0.0) |
| Brand3,4,5 | Pose C (unseen) | 2/5 (0.3) | 5/5 (0.0) | 4/5 (1.0) |
| Total |         |      | 22/25 (0.3) | 25/25 (0.0) | 23/25 (2.0) |

### TABLE 2. Offline evaluation of fracture anticipation.

|          | Accuracy(%) | Precision(%) | Recall(%) | F-measure |
|----------|-------------|--------------|-----------|-----------|
| Proposed1 | 88.748 | 86.947 | 80.250 | 82.979 |
| Proposed2 | 87.727 | 86.375 | 79.986 | 82.600 |

|          | Accuracy(%) | Precision(%) | Recall(%) | F-measure |
|----------|-------------|--------------|-----------|-----------|
| Proposed1 | 88.122 | 86.467 | 78.081 | 81.928 |
| Proposed2 | 87.424 | 84.742 | 79.870 | 82.103 |

|          | Accuracy(%) | Precision(%) | Recall(%) | F-measure |
|----------|-------------|--------------|-----------|-----------|
| Proposed1 | 86.449 | 80.557 | 83.283 | 81.013 |
| Proposed2 | 85.984 | 79.347 | 84.792 | 81.420 |
interesting to change the picking strategies according to the type of ingredient. For example, we can design strategies whereby the robot can grasp stiff food ingredients such as carrots aggressively and soft ones such as potato chips and tofu gently.

VIII. CONCLUSION

We presented a learning-based framework of fracture anticipation for fragile food manipulation using tactile sensors. In our approach, the model is trained by allowing a robot to break the objects. A fracture-anticipation network is learned using the collected data, which the model uses to predict fractures. We performed experiments using a real robot and fragile ingredients to validate our method. The results indicated that our method predicted fractures with approximately 80% recall and reduced the rate of object breakage compared with a rule-based baseline without reducing the picking success rate.

ACKNOWLEDGMENT

The authors would like to thank Professor Hideo Saito for his valuable comments on improving this study.

REFERENCES

[1] M. Bollini, S. Tellex, T. Thompson, N. Roy, and D. Rus, “Interpreting and executing recipes with a cooking robot,” in Proc. Int. Symp. Exp. Robot., 2013, pp. 481–495.
[2] Z. Wang, S. Hirai, and S. Kawamura, “Challenges and opportunities in robotic food handling: A review,” Frontiers Robot. AI, vol. 8, p. 433, Jan. 2022.
[3] J. Iqbal, Z. H. Khan, and A. Khalid, “Prospects of robotics in food industry,” Food Sci. Technol., vol. 37, no. 2, pp. 159–165, 2017.
[4] W. Chen, H. Khans, I. Biznies, N. F. Lepora, and S. J. Redmond, “Tactile sensors for friction estimation and incipient slip detection—Toward dexterous robotic manipulation: A review,” IEEE Sensors J., vol. 18, no. 22, pp. 9049–9064, Nov. 2018.
[5] C. Swackhammer and G. M. Bornhorst, “Fracture properties of foods: Experimental considerations and applications to mastication,” J. Food Eng., vol. 263, pp. 213–226, Dec. 2019.
[6] I. Lenz, R. A. Knepper, and A. Saxena, “DeepMPC: Learning deep latent features for model predictive control,” in Robotics: Science and Systems, 2015.
[7] N. Vulin, S. Christen, S. Stevšić, and O. Hilliges, “Improved learning of robot manipulation tasks via tactile intrinsic motivation,” IEEE Robot. Autom. Lett., vol. 6, no. 2, pp. 2194–2201, Apr. 2021.
[8] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[9] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 3104–3112.
[10] T. Nishimura, Y. Suzuki, T. Tsuji, and T. Watanabe, “Fluid pressure monitoring-based strategy for delicate grasping of fragile objects by a robotic hand with fluid fingertips,” Sensors, vol. 19, no. 4, p. 782, Feb. 2019.
[11] K. Takahashi, N. Fukaya, and A. Ummadisingu, “Target-mass grasping of entangled food using pre-grasping & post-grasping,” IEEE Robot. Autom. Lett., vol. 7, no. 2, pp. 1222–1229, Apr. 2022.
[12] D. Sepulveda, R. Fernandez, E. Navas, M. Armada, and P. Gonzalez-De-Santos, “Robotic aubergine harvesting using dual-arm manipulation,” IEEE Access, vol. 8, pp. 121889–121904, 2020.
[13] N. Figueroa, A. L. P. Ureche, and A. Billard, “Learning complex sequential tasks from demonstration: A pizza dough rolling case study,” in Proc. 11th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI), Mar. 2016, pp. 611–612.
[14] M. Beetz, U. Klang, I. Kresse, A. Maldonado, L. Mosenlechner, D. Pangerl, T. Ruhr, and M. Tenorth, “Robotic roommates making pancakes,” in Proc. 11th IEEE-RAS Int. Conf. Humanoid Robots, Oct. 2011, pp. 529–536.
[15] A. Pettersson, S. Davis, J. O. Gray, T. J. Dodd, and T. Ohlsson, “Design of a magnetorheological robot gripper for handling of delicate food products with varying shapes,” J. Food Eng., vol. 98, no. 3, pp. 332–338, Jun. 2010.
[16] R. Maruyama, T. Watanabe, and M. Uchida, “Delicate grasping by robotic gripper with incompressible fluid-based deformable fingertips,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Nov. 2013, pp. 5469–5474.
[17] T. Yamanaka, S. Katagi, K. Nabae, K. Suzumori, and G. Endo, “Development of a food handling soft robot hand considering a high-speed pick-and-place task,” in Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII), Jan. 2020, pp. 87–92.
[18] T. Sakuma, E. Phillips, G. A. G. Ricardez, M. Ding, J. Takamatsu, and T. Ogasawara, “A parallel gripper with a universal fingertip device using optical sensing and jamming transition for maintaining stable grasps,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Nov. 2019, pp. 5814–5819.
[19] G.-K. Lau, K.-R. Heng, A. S. Ahmed, and M. Shrestha, “Dielectric elastomer fingers for versatile grasping and nimble pinching,” Appl. Phys. Lett., vol. 110, no. 18, May 2017, Art. no. 182906.

[20] J.-Y. Lee, Y.-S. Seo, C. Park, J.-S. Koh, U. Kim, J. Park, H. Rodrigue, B. Kim, and S.-H. Song, “Shape-adaptive universal soft parallel gripper for delicate grasping using a stiffness-variable composite structure,” IEEE Trans. Ind. Electron., vol. 68, no. 12, pp. 12441–12451, Dec. 2021.

[21] B. S. Homberg, R. K. Katzschmann, M. R. Dogar, and D. Rus, “Haptic identification of objects using a modular soft robotic gripper,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Sep. 2015, pp. 1698–1705.

[22] XELA Robotics. Model XR1944. Accessed: Aug. 6, 2022. [Online]. Available: https://xelarobotics.com/xr1944

[23] J. M. Romano, K. Hsiao, G. Niemeyer, S. Chitta, and K. J. Kuchenbecker, “Human-inspired robotic grasp control with tactile sensing,” IEEE Trans. Robot., vol. 27, no. 6, pp. 1067–1079, Dec. 2011.

[24] E. Misimi, A. Olofsson, A. Eilersen, E. R. Oye, and J. R. Mathiassen, “Robotic handling of compliant food objects by robust learning from demonstration,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Oct. 2018, pp. 6972–6979.

[25] A. Lillenschlojd, R. Rahal, P. R. Giordano, C. Paccherotti, and E. Misimi, “Human-inspired haptic-enabled learning from prehensile move demonstrations,” IEEE Trans. Syst., Man, Cybern., Syst., vol. 52, no. 4, pp. 2061–2072, Apr. 2022.

[26] A. Yamaguchi and C. A. Atkeson, “Tactile behaviors with the vision-based tactile sensor FingerVision,” Int. J. Humanoid Robot., vol. 16, no. 3, Jun. 2019, Art. no. 1940002.

[27] T. Narita, S. Nakagari, W. Conus, T. Tsuboi, and K. Nagasaka, “Theoretical derivation and realization of adaptive grasping based on rotational incipient slip detection,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2020, pp. 531–537.

[28] S. Funabashi, Y. Kage, H. Oka, Y. Sakamoto, and S. Sugano, “Object picking using a two-fingered gripper measuring the deformation and slip detection based on a 3-axis tactile sensing,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Sep. 2021, pp. 3888–3895.

[29] M. C. Gemici and A. Saxena, “Learning haptic representation for manipulating deformable food objects,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Sep. 2014, pp. 638–645.

[30] A. Sawhney, S. Lee, K. Zhang, M. Veloso, and O. Kroemer, “Playing with food: Learning food item representations through interactive exploration,” in Proc. Int. Symp. Exp. Robot., 2020, pp. 309–322.

[31] P. Ribeiro, S. Cardoso, A. Bernardino, and L. Jamone, “Fruit quality control by surface analysis using a bio-inspired soft tactile sensor,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Oct. 2020, pp. 8875–8881.

[32] T. Taunyazov, L. S. Song, E. Lim, H. H. See, D. Lee, B. C. K. Tee, and H. Soh, “Extended tactile perception: Vibration sensing through tools and grasped objects,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Sep. 2021, pp. 1755–1762.

[33] Robots. OpenMANIPULATOR-X E-Manual. Accessed: Aug. 6, 2022. [Online]. Available: https://emanual.robotis.com/docs/en/platform/opmanipulator-x-overview/

[34] M. Hanada, “Food-texture dimensions expressed by Japanese onomatopoeic words,” J. Texture Stud., vol. 51, no. 3, pp. 398–411, Jun. 2020.

[35] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” in Proc. Adv. Neural Inf. Process. Syst., vol. 28, 2015, pp. 1–9.

[36] K. Fang, Y. Bui, S. Hinterstoisser, S. Savarese, and M. Kalakrishnan, “Multi-task domain adaptation for deep learning of instance grasping from simulation,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2018, pp. 3516–3523.

[37] K. Tanaka, R. Ronetani, M. Hamaya, R. Lee, F. V. Drigalski, and Y. Jiiri, “TRANS-AM: Transfer learning by aggregating dynamics models for soft robotic assembly,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2021, pp. 4627–4633.

REINA ISHIKAWA (Student Member, IEEE) received the B.E. degree in information and computer science from Keio University, Japan, in 2021, where she is currently pursuing the M.S.Eng. degree in science and technology. Her research interests include robotics combining computer vision and machine learning using multi-sensors and individual and group activity recognition.

MASASHI HAMAYA (Member, IEEE) received the Ph.D. degree in engineering from Osaka University, in 2019. From 2014 to 2019, he was an Intern Student at the Department of Brain Robot Interface, ATR Computational Neuroscience Laboratories. Since 2019, he has been a Senior Researcher with the Robotics Group, OMRON SINIC X Corporation. His research interests include robotics, soft robotics, human–robot interaction, and robot learning. He received the IEEE Kansai Section Student Paper Award, in 2018, and the IEEE Robotics and Automation Society Japan Joint Chapter Young Award, in 2019. He is a member of RSJ.

FELIX VON DRIGALSKI (Member, IEEE) received the dual M.Sc. degree in engineering from the Karlsruhe Institute of Technology, Germany, and the INSA Lyon, France, and the Ph.D. degree in computer science from the Nara Institute of Science and Technology, Japan. From 2018 to 2021, he was a Senior Researcher at OMRON SINIC X Corporation, Japan. Since 2021, he has been a System Engineer at Mijun Corporation, Japan. His research interests include robotic manipulation, tactile sensing, and autonomous assembly. His teams reached First Place at the Airbus Shopfloor Challenge at ICRA, in 2016, Finalist Prize at the Amazon Robotics Challenge, in 2017, and Third and Fourth Place at the World Robot Summit Assembly Challenge, in 2020 and 2018, along with awards from the Japanese Society for Artificial Intelligence (JSAI) and the Society of Instrument and Control Engineers (SICE).

KAZUTOSHI TANAKA (Member, IEEE) received the Ph.D. degree in information science and technology from The University of Tokyo, in 2017. From 2017 to 2019, he was a Postdoctoral Fellow at the Department of Mechno-Informatics, The University of Tokyo. Since 2019, he has been a Senior Researcher at the Robotics Group, OMRON SINIC X Corporation. His research interests include robotics performing sports, human–robot interaction, mechanics, control, and motion learning. He received the JSJ Young Investigation Excellence Award, in 2018.

ATSUSHI HASHIMOTO (Member, IEEE) received the B.S. degree in engineering and the M.S. and Ph.D. degrees in informatics from the Graduate School of Informatics, Kyoto University, Japan, in 2005, 2008, and 2013, respectively. He is currently a Senior Researcher with Omron SINIC X Corporation. Since 2018, he has been working on human activity analysis and recognition based on human-object interaction, and interface design and development in the field of cooking.

His research interests include pattern recognition, computer vision, and human–computer interaction. He is a member of the ACM.

** **