Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.081437

ARTICLE

A Lightweight Dual-Branch Hybrid CNN for Real-Time Hardness Recognition Using Low-Cost Tactile Sensors

Thossapon Kaewrakmuk and Jakkree Srinonchat*

Department of Electronics and Telecommunication Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Khlong Luang, Pathum Thani, Thailand

*Corresponding Author: Jakkree Srinonchat. Email: jakkree.s@en.rmutt.ac.th

Received: 02 March 2026; Accepted: 28 May 2026; Published: 23 July 2026

ABSTRACT: Robotic systems require reliable tactile perception to evaluate object stiffness during physical interaction. This study proposes a lightweight dual-branch architecture, named Hybrid-CNN-ResVgg, designed to improve hardness recognition using data from a low-cost piezoresistive tactile sensor. The model combines a one-dimensional convolutional neural network (1D-CNN) based on a ResNet8-Lite architecture for learning temporal signal patterns and a two-dimensional convolutional neural network (2D-CNN) based on a VGG6-Lite architecture for learning spatial representations derived from Gramian Angular Difference Fields (GADF). A cross-architecture fusion mechanism is introduced to integrate temporal and spatial features while reducing redundant representation learning. Experiments were conducted on a controlled dataset comprising three hardness levels, with repeated grasp interactions to ensure consistent model evaluation. The proposed Hybrid-CNN-ResVgg achieved the highest accuracy of 89.67% among the evaluated models, including non-CNN baseline models, standard CNN architectures, tactile perception models, and single-domain lightweight CNN models. Despite its improved accuracy, the model requires only 0.039 giga floating-point operations (GFLOPs) and 0.46 megabytes of memory, supporting the computational feasibility of future real-time implementation on resource-constrained robotic platforms. The results indicate that combining temporal and spatial tactile information through lightweight cross-domain architectures can improve hardness recognition performance. This study provides a practical foundation for extending tactile perception toward more complex materials, continuous stiffness estimation, and multimodal sensing in future robotic applications.

KEYWORDS: Tactile sensing; hybrid-CNN; lightweight deep learning architecture; robotic hardness recognition

1 Introduction

The machine's capacity to understand and adapt to its operational environment is the key to human-robot interaction [1,2]. Although visual perception has traditionally been a major focus in robotic perception research, tactile sensing provides complementary contact-based information, enabling robots to interact with objects and their surroundings physically [3]. Robot hands [4,5] must be flexible and adaptive to manipulate goods and understand their mechanical qualities [6]. Traditional robotic hands can manipulate objects, but they lack human dexterity. Robotic perception faces particular challenges in tactile sensing [7]. For robotic applications, piezoelectric, capacitive, and piezoresistive tactile sensing technologies have been investigated, each with varied sensitivity, complexity, and cost [8,9]. Therefore, effective and low-cost tactile sensing methods for real-time robotic applications are gaining popularity. Tactile input is complex and requires rapid processing, which in turn requires sophisticated computational methods to extract useful information from robotic hand-object interactions. Robotic tactile perception is important for estimating object hardness and contact conditions during physical interaction, supporting applications in industrial automation, healthcare and rehabilitation, and other contact-rich robotic tasks [10–12]. Since human tactile perception is crucial in many applications, robotic systems that can better perceive and interpret tactile input are in demand.

Convolutional neural networks (CNNs) gained widespread attention after demonstrating strong performance in large-scale image recognition tasks [13], and they have since been applied to sensor-based gesture recognition and human–machine or human–robot interaction tasks [14–16]. CNN-based models can automatically learn hierarchical feature representations from tactile data and extract spatial and temporal patterns that are useful for robotic tactile perception [17].

This makes them ideal for processing complex tactile information. To study new areas of hardness comprehension, we directly integrate these networks into robotic hands. CNN-based techniques have shown promise for tactile perception tasks; however, most investigations focus on temporal tactile signals or spatial image representations separately. 1D-CNN methods are effective for learning sequential pressure dynamics from tactile signals, whereas 2D-CNN methods are better at capturing spatial relationships in modified tactile representations. However, a single representation may not capture all tactile interactions, especially for hardness-identification tasks that require delicate pressure patterns. In addition, many existing deep learning architectures for tactile sensing employ computationally intensive models that may not be suitable for robotic systems with limited processing resources. Therefore, there remains a need for lightweight tactile-oriented architectures capable of jointly learning complementary temporal and spatial features while maintaining computational efficiency. To address these challenges, this study proposes a lightweight Hybrid-CNN framework that integrates custom-designed 1D-CNN and 2D-CNN architectures for tactile hardness recognition. The proposed framework combines temporal pressure dynamics and spatial representations derived from tactile signals to provide complementary feature information for hardness discrimination. By integrating lightweight feature extraction strategies with multimodal tactile representations, the proposed approach aims to improve recognition performance while remaining suitable for resource-constrained robotic applications.

This research provides a cohesive framework that effectively represents temporal dynamics and spatial properties. Because of this, it will be much simpler to have a flexible, all-encompassing understanding of the difficulty level across various contexts. The primary contribution of this study may be concisely articulated as follows: (1) This paper introduces lightweight CNN architectures inspired by VGG16, ResNet18, and DenseNet201 for tactile hardness recognition. Each architecture is implemented in both 1D-CNN and 2D-CNN configurations to separately investigate temporal signal learning and spatial feature extraction from image-transformed tactile data. This design provides a systematic framework for analyzing the effectiveness of temporal and spatial tactile representations under resource-constrained conditions. (2) This paper presents a lightweight Hybrid-CNN architecture that integrates custom-designed 1D-CNN and 2D-CNN branches using an FC-based feature fusion mechanism. The fusion strategy combines temporal pressure-signal features and GADF-based spatial features into a unified representation, reducing feature redundancy and controlling model complexity to improve hardness discrimination. (3) This study provides a comprehensive comparative evaluation of the proposed lightweight CNN and Hybrid-CNN models against non-CNN baseline models, standard CNN architectures, and tactile perception models. The comparative analysis demonstrates how architectural design influences recognition performance, feature representation, model compactness, and computational efficiency in robotic tactile sensing applications.

2 Related Work

Combining tactile sensing with machine learning and deep learning has improved robotic perception, enabling robots to perceive complex tactile patterns and make sensory-driven judgments. Flexible and compliant tactile sensing systems that adapt to physical encounters have been studied more recently. Drimus et al. [18] proposed a flexible tactile array sensor for object detection that leverages haptic feedback during finger contact, demonstrating the feasibility of low-cost, adaptable tactile-sensing architectures. Deng et al. [19] developed a bio-inspired microfluidic tactile sensor for soft robotic grippers, featuring a compliant sensing structure inspired by human skin to enhance tactile interaction. Recent research has highlighted the necessity of compliant tactile sensing structures for environmental interaction and robotic adaptation. Moreno et al. [20] proposed a soft paw sensor for legged robots that uses flexible, conductive membranes to monitor contact force and location for adaptive tactile perception and environmentally aware interaction in complex terrain. Recent research has investigated biomimetic tactile sensing devices that imitate human skin perception. Lv et al. [21] proposed a hydrogel-based bimodal tactile sensor with triboelectric and ionic supercapacitive sensing for dynamic and static tactile perception in robotic hands. The biomimetic sensing framework shows how intelligent tactile systems can recognize material and hardness in robotics. Tactile representation learning has been studied using image-based tactile representations, learning-based frameworks, and tactile sensor hardware. Pohtongkam and Srinonchat [22] developed a hand tactile sensor system that converts tactile input into image representations for object identification, demonstrating that tactile image representation is feasible for robotic perception. Funabashi et al. [23] proposed a morphology-specific tactile learning framework based on the robotic hand's tactile sensor configuration. Their hierarchical learning technique processes tactile input from small sensor areas to the full hand, proving that structure-aware tactile feature learning works. This research shows that tactile sensing structures are being integrated with learning-based frameworks to enhance robotic tactile perception and object recognition. Machine learning methods for tactile hardness detection and item categorization have been tested to interpret tactile data from these sensing devices. Early tactile recognition experiments used k-nearest neighbors (KNN) and dynamic temporal warping (DTW) to understand tactile input during object contact [18]. Sharma et al. [24] investigated hardness classification in robotic grippers using mechanoreceptor-inspired configurations of commercial off-the-shelf tactile sensors. Their study evaluated different tactile sensor arrangements for multiclass material hardness classification based on the Shore hardness scale using machine learning classifiers. Also, Liu et al. [25] proposed a material identification framework employing a robotic hand's capacitive touch sensor array. Their work used dimensionality reduction and machine learning classifiers to improve material recognition and reduce computational complexity. These approaches demonstrated that tactile hardness could be identified, though they relied on handcrafted feature representations. Such methods may hinder the acquisition of sensor data for complex tactile patterns and temporal correlations. Recently, researchers have used deep learning to automatically learn hierarchical feature representations from tactile inputs.

CNNs can automatically learn hierarchical spatial and temporal tactile representations for object and hardness recognition. Tactile object identification using image-based tactile representations, Bag-of-Words (BoW), and deep convolutional neural networks (DCNNs) for robotic perception tasks has been studied [22]. Morphology-specific Convolutional Neural Network (MS-CNN) frameworks have also been studied based on the layout of robotic hand tactile sensors [23]. Structure-aware tactile feature learning was demonstrated by hierarchically processing tactile input from local sensor areas to the robotic hand using the MS-CNN architecture. Song et al. [26] also used robotic touch input from a flexible tactile sensor array to demonstrate the ability of CNN-based architectures to identify tactile items. Kaewrakmuk and Srinonchat [27] introduced a multisensor data fusion architecture with time-series-to-image encoding for hardness identification, enabling CNNs to learn spatial properties from temporal tactile input. Recent work has examined lightweight, embedded tactile learning systems for real-time hardness identification on resource-constrained devices, demonstrating that CNN-based tactile recognition models can be used in robotic applications [28]. Multimodal and hybrid learning methods are used to enhance tactile object identification by combining complementary sensory data. Gao et al. [29] proposed a multimodal framework for surface categorization that combines optical and haptic inputs, employing 1D-CNN and 2D-CNN architectures. Combining visual and tactile representations increases haptic perception and surface comprehension, according to their research. Zhang et al. [30] used visual-haptic fusion to enhance robotic object description by extracting and classifying multimodal features. Integrating visual and tactile information outperformed either modality alone in categorization. Pastor et al. [31] used a hybrid LSTM to recognize multimodal objects using tactile and kinesthetic inputs. Their work showed that robotic systems may enhance item identification by combining sensory modalities. A bespoke haptic glove with force-sensitive and bend-sensor arrays was used by Lu et al. [32] to develop a 3-D tactile object identification framework. They used CNN and Multi-Layer Perceptron(MLP) MLP models to enhance tactile object identification. Recent research has expanded multimodal tactile perception into cross-modal frameworks that combine touch, visual, and textual information. Li et al. [33] proposed a Transformer-based tactile-visual-textual fusion network to enhance multimodal object identification by leveraging cross-modal feature interactions and semantic alignment. These findings show the growing relevance of multimodal and hybrid learning methodologies for robotic tactile perception, integrating complementary sensory inputs.

Recent research has shown promising tactile object-identification performance using machine learning, CNN-based approaches, and multimodal fusion, but significant limitations remain. Most research concentrates on temporal tactile sequences or spatial tactile representations separately, which may limit their capacity to capture complex tactile interactions. Multiple modal and hybrid learning techniques rely on computationally intensive designs that may not be suitable for resource-constrained robotic systems requiring real-time tactile sensing. Hardness recognition also benefits from temporal tactile signals and spatial image-based representations. Temporal signals maintain sequential pressure fluctuations during tactile engagement, whereas image-encoded tactile representations allow CNN models to infer spatial linkages and structural patterns. Integrating these complementary representations to improve tactile perception while maintaining computing efficiency is difficult. Therefore, this paper proposes a lightweight Hybrid-CNN framework that integrates custom-designed 1D-CNN and 2D-CNN architectures for tactile hardness recognition. By integrating temporal and spatial tactile representations within a unified framework, the proposed approach aims to improve hardness recognition performance while maintaining suitability for real-time robotic applications.

3 Proposed Method

3.1 Lightweight CNN Models

The convolutional neural network (CNN) architectures investigated in this study include VGG [34], ResNet [35], and DenseNet [36]. These architectures were adapted into lightweight variants for tactile sensor input and deployment on resource-constrained devices. Their selection was motivated by previous studies demonstrating the effectiveness of pressure-based hardness recognition [27]. For performance comparison, each architecture was implemented in both 1D-CNN and 2D-CNN configurations to process temporal tactile signals and image-transformed tactile representations, respectively. Batch normalization, Rectified linear unit (ReLU) ReLU activation, and Global Average Pooling (GAP) were applied across all architectures to improve feature learning and reduce model complexity before final classification.

3.1.1 Vgg6-Lite

The Vgg6-Lite model was adapted from the VGG16 architecture and simplified to reduce computational complexity while preserving the hierarchical convolutional structure. The architecture consists of five convolutional layers with 16, 32, 64, 64, and 64 filters. Kernel sizes were set to 3 (1D-CNN) and 3 × 3 (2D-CNN). Max pooling with a kernel size of 2 × 2 and a stride of 2 was applied after Conv1, Conv2, and Conv5 to reduce feature dimensions progressively. After feature extraction, a GAP layer processed the resulting feature maps, followed by a fully connected (FC) layer, and finally by a SoftMax layer for final classification. The overall architecture is illustrated in Fig. 1a.

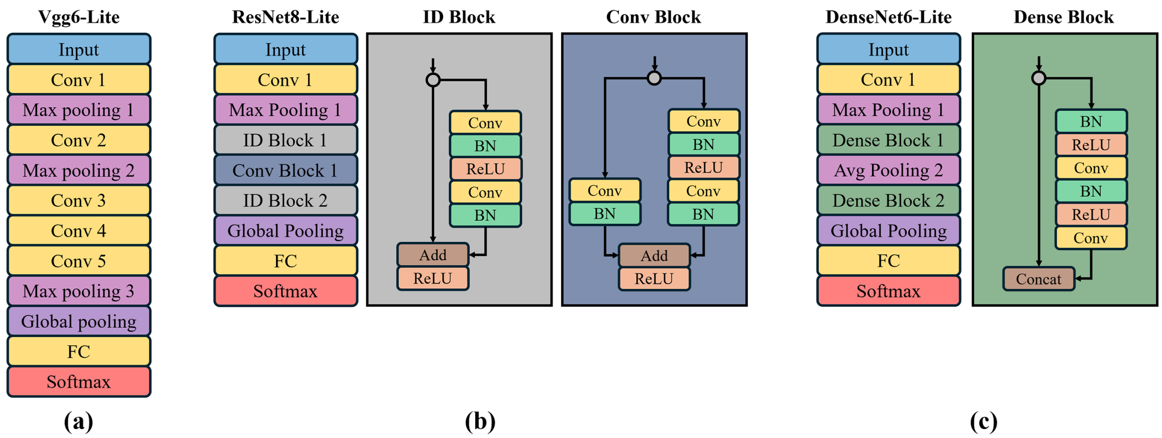


Figure 1: Architecture diagrams of the lightweight CNN models used in this study: (**a**) Vgg6-Lite; (**b**) ResNet8-Lite, including identity and convolutional residual blocks; (**c**) DenseNet6-Lite, including the dense block structure.

3.1.2 ResNet8-Lite

The ResNet8-Lite model was adapted from ResNet18 by reducing the number of layers and filters better to suit tactile data with limited temporal and spatial resolution. The model begins with a convolutional layer containing 16 filters, with kernel sizes of 5 (1D-CNN) and 5 × 5 (2D-CNN), followed by max pooling with kernel sizes of 3/3 × 3 and a stride of 2. The residual learning stage consists of ID Block 1, Conv Block, and ID Block 2, as illustrated in Fig. 1b. The Identity Block (ID Block) contains two convolutional layers with kernel sizes of 3 (1D-CNN) and 3 × 3 (2D-CNN), together with shortcut connections that preserve important features from earlier layers while enabling deeper feature learning. ID Block 1 uses 16 filters for initial feature extraction, whereas ID Block 2 uses 32 filters for higher-level feature representation. The Convolutional Block (Conv Block) downsamples features with a stride of 2 and increases the number of filters to 32. Shortcut connections with 1 × 1 convolutions were employed to match feature dimensions before residual addition. Following the residual learning stage, the extracted features were processed using GAP and FC layers before final classification.

3.1.3 DenseNet6-Lite

The DenseNet6-Lite model was created by reducing the DenseNet201 architecture to make it work with touch data. At the beginning of the model, the input is convolved with 16 filters. with kernel sizes of 5 (1D-CNN) and 5 × 5 (2D-CNN), followed by max pooling with a kernel size of 3/3 × 3 and a stride of 2. The architecture consists of Dense Block 1, a Transition Layer, and Dense Block 2, as illustrated in Fig. 1c. Each dense block employs a bottleneck structure comprising a 1 × 1 convolution layer followed by a 3 × 3 convolution layer. The 1 × 1 convolution layer uses 64 filters for channel reduction, whereas the subsequent

3/3 × 3 convolution layer uses 16 filters for feature extraction. Dense connections concatenate feature maps from previous layers, encouraging feature reuse and improving information flow throughout the network. The Transition Layer applies average pooling with a kernel size of 2/2 × 2 and a stride of 2 to reduce feature dimensions while maintaining computational efficiency. After feature extraction, the resulting feature maps were processed through GAP and FC layers before final classification using the SoftMax layer.

3.2 Hybrid CNN Method

The Hybrid-CNN architecture aims to combine the benefits of both 1D-CNNs and 2D-CNNs. The 1D-CNN branch captures temporal dynamics in sequential sensor signals, while the 2D-CNN branch focuses on spatial correlations after converting the data into a 2D representation. The rationale behind this fusion strategy is that hardness-related tactile information may appear in both temporal pressure variations and GADF-based spatial patterns. Therefore, fusing the 1D and 2D representations allows the model to exploit complementary features and reduces reliance on a single feature domain for hardness discrimination. An overview of the proposed Hybrid-CNN architecture is illustrated in Fig. 2.

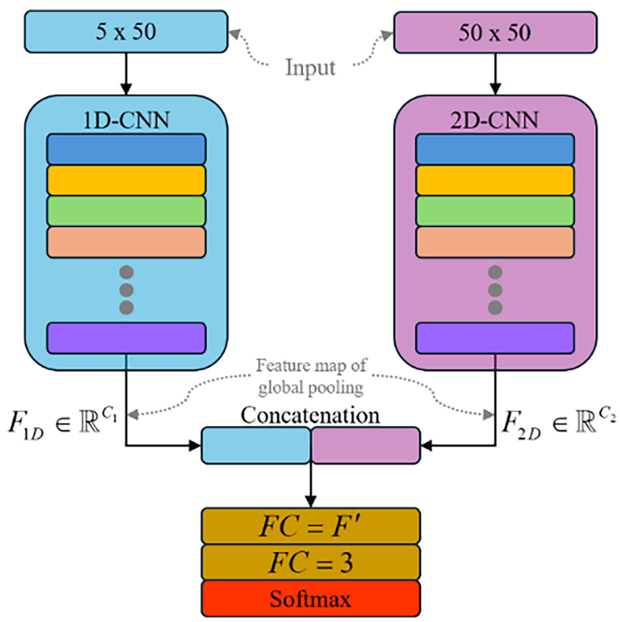


Figure 2: Hybrid-CNN architecture of the proposed research.

After feature extraction, the output feature maps from the 1D-CNN and 2D-CNN branches are processed using Global Average Pooling (GAP), producing one-dimensional feature vectors $F_1 \in \mathbb{R}^{C_1}$ and $F_2 \in \mathbb{R}^{C_2}$, respectively, where C_1 and C_2 denote the number of feature channels determined by the respective network architectures. These vectors represent temporal and spatial features extracted from the two branches. The mathematical formulation of the combined feature vector is defined in Eq. (1):

$$F_{concat} = [F_{1D}; F_{2D}] \in \mathbb{R}^{C_1+C_2} \tag{1}$$

This operation integrates complementary information from both temporal and spatial representations into a unified feature space.

The vector F_{concat}, derived from the preceding step, is in one-dimensional format. Consequently, a Fully Connected (FC) layer is applied to reduce the dimensionality of the fused feature vector. The reduced dimension is set to fifty percent of the concatenated size, as defined in Eq. (2):

$$C' = \frac{C_1 + C_2}{2} \tag{2}$$

This dimensionality reduction helps control model complexity and mitigate feature redundancy, while enabling the model to learn interactions between the fused features from both branches.

The linear transformation in the FC layer is defined in Eq. (3):

$$F' = \sigma\left(WF_{concat} + b\right), \quad F' \in \mathbb{R}^{C'} \tag{3}$$

where $W \in \mathbb{R}^{C'}$ denotes the weight matrix, b is the bias term, and σ represents the activation function (ReLU).

This transformation produces a fused feature representation by combining temporal and spatial features for subsequent classification.

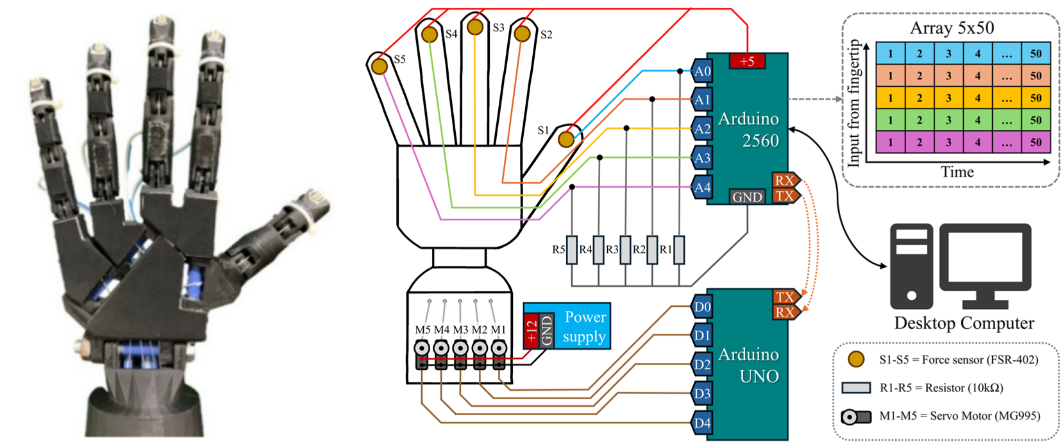
The resultant vector F' will be sent to the final FC layer, which contains three neurons corresponding to the number of hardness levels to be classified. The outcomes are further processed using the SoftMax function to produce probability values for each category.

4 Experiments Setup

4.1 Tactile Sensor on a Robotic Hand

This research sought to improve the tactile sensing capabilities of robotic hands by selecting an economical, reliable sensor suitable for seamless integration into a 3D-printed robotic hand system. Piezoresistive tactile sensors are widely used in robotic tactile sensing because they can capture pressure-related responses during physical interaction while maintaining a simple and compact sensing structure [9]. Therefore, the FRS-402 pressure sensor was selected as a piezoresistive tactile sensor, offering a balance of sensitivity, robustness, and cost-effectiveness. In this work, no calibration was performed to convert sensor readings into physical force units. Instead, raw sensor signals were directly used to represent tactile interactions. Material hardness influences tactile interaction patterns during grasping, reflected in temporal variations in sensor signals. Consequently, hardness recognition was achieved by learning signal patterns rather than directly measuring force values.

A prototype robotic hand, inspired by the open-source Inmoov hand, was fabricated via 3D printing, as shown in Fig. 3a. Also, the data acquisition and control system is shown in Fig. 3b. The robotic hand consists of five anthropomorphic digits, each with three joints, controlled by servo motors via tendon-driven actuation. FRS-402 pressure sensors were positioned at the center of each fingertip to capture pressure variations during object interaction. Although sensors were installed on all fingertips, contact conditions naturally varied among fingers because of differences in finger geometry and grasping orientation, particularly for the thumb and little finger. Similar to human grasping behavior, tactile perception was obtained from the combined responses of multiple sensors rather than a single contact point, allowing the robotic hand to capture distributed tactile information associated with object hardness.



(**a**) (**b**)

Figure 3: Robotic hand and data acquisition setup used in this study: (**a**) 3D-printed Inmoov-inspired robotic hand with fingertip pressure sensors; (**b**) structured data acquisition and control process for tactile signal collection.

4.2 Data Acquisition and Pre-Processing

The silicone models were made in the shape of cylinders measuring 5 cm wide and 8 cm tall. The silicone used is rated on the Shore A hardness scale [37]. There are three levels: Shore A 00, which is the softest; Shore A 10, which is of moderate hardness; and Shore A 20, which is the toughest. We chose these three levels of hardness to provide different, controllable kinds of tactile information. Choosing just three levels makes it easier to evaluate the model's performance under consistent conditions, reducing the impact of external factors, such as changes in surface roughness or material morphology, that might otherwise affect the model's results. As a result, this choice is suitable for early tests of the model's performance and may be systematically extended in the future to cover more complex real-world scenarios.

Tactile data were collected using a robotic hand that grasped silicone models of each hardness level 500 times. The silicone models were placed at the center of the robot's palm in a horizontal orientation, with consistent positioning maintained across trials. A structured data acquisition process ensures accurate data collection and reliable analysis. The approach used to gather information for this investigation is shown in Fig. 3b. The process begins with the computer sending instructions to the Arduino Mega 2560 microcontroller board over a serial connection. It is the first step in the operation. The microcontroller will collect data from the pressure sensor every 0.1 for 5 s. The grasping motion is controlled by the Arduino Uno microcontroller, which regulates the servo motors to predefined positions to ensure the same grasping motion is applied in each trial. The collected data is stored in a 5×50 array, where 5 rows correspond to different sensor readings and 50 columns represent time steps. After data acquisition, the Arduino Mega 2560 microcontroller board will send the data back to the computer via a serial connection and save it as CSV files for further analysis. The dataset consists of 1500 samples and was evaluated using a five-fold cross-validation scheme. Examples of tactile signal responses for the three hardness levels are shown in Fig. 4.

In the pre-processing stage, data is restructured to align with the architecture of each CNN type. For 1D-CNN, raw sensor data is used directly in a 5×50 format, preserving the temporal characteristics of tactile interactions and retaining signal variations relevant to hardness discrimination. The X-axis (rows) represents features, and the Y-axis (columns) represents the time sequence. This representation preserves key signal characteristics for subsequent learning. In contrast, for 2D-CNN, the data format is transformed from 5×50 to 1×250 via row-wise flattening, unifying multichannel temporal signals into a single continuous sequence suitable for angle-based encoding. After flattening, the data is rescaled to the range of [−1, 1] to ensure numerical validity for angular operations. The normalized sequence is then encoded into a 2D image using the Gramian Angular Difference Field (GADF), which transforms the temporal sequence into a spatial representation by computing angular differences. The GADF transformation is configured to produce a 50×50 image, ensuring compatibility with the 2D-CNN model's input dimension requirements. To enhance visual contrast and make pressure variations more distinguishable, the resulting GADF image is further converted to a color image using a colormap. The overall transformation process is illustrated in Fig. 5, which outlines the step-by-step conversion from raw sensor data to a GADF image. This transformation enables the 2D-CNN to capture underlying temporal dependencies through spatial feature extraction effectively. GADF was selected based on its previously demonstrated superior performance in similar time-series classification tasks [27].

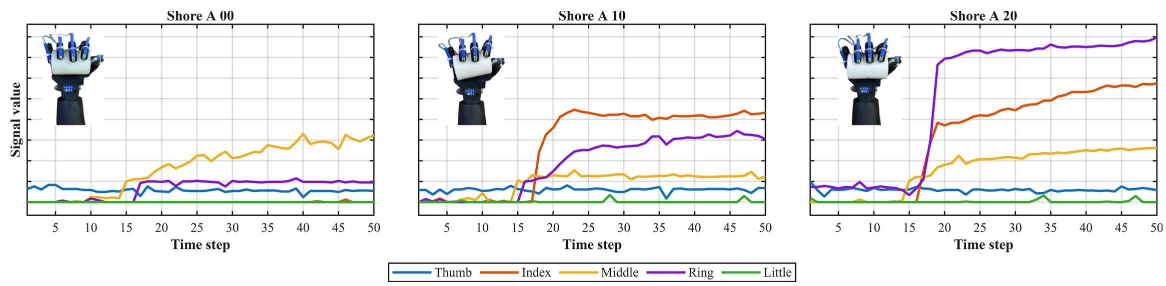


Figure 4: Tactile signal responses from five fingertip sensors under three shore a hardness levels.

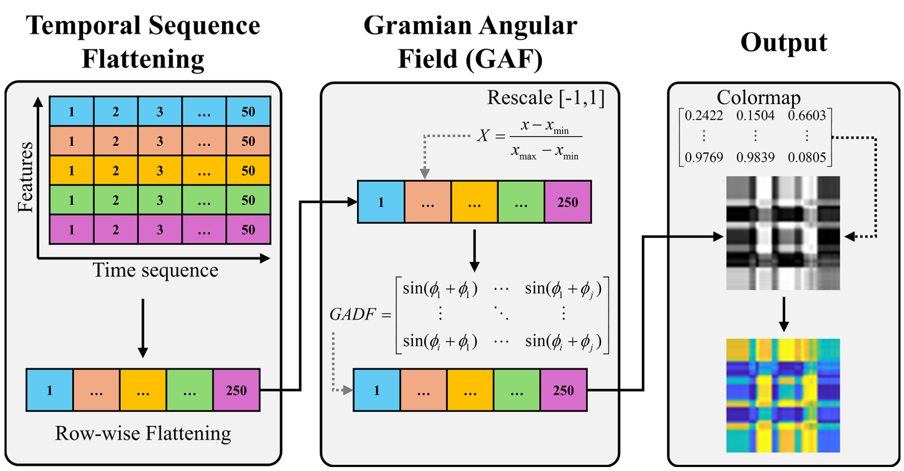


Figure 5: Data pre-processing using GADF for 2D-CNN input.

4.3 Training Parameter

A computer-based training approach was used to evaluate CNNs' ability to accurately recognize hardness. The system has an Intel Core i7-8750H processor and an NVIDIA RTX 3080 graphics card. To ensure model equivalence, the training parameters are adjusted based on the information in Table 1. For the batch size, a value of 8 was used for VGG16, ResNet18, and DenseNet201 due to their high computational complexity and GPU memory constraints, whereas a larger batch size of 64 was adopted for TactNet [38], MLP [39], LSTM [40], 1D-CNN, 2D-CNN, and Hybrid-CNN, as these models generally require lower computational resources and allow more efficient batch processing. Similarly, different learning rates were applied to VGG16, ResNet18, and DenseNet201 to account for variations in network depth and convergence behavior, while a fixed learning rate of 0.01 was used for TactNet, MLP, LSTM, 1D-CNN, 2D-CNN, and Hybrid-CNN to maintain consistent training conditions across the proposed and comparison models.

Table 1: Training parameters for the different models.

Models	Input	Batch Size	Learn Rate
VGG16	(224 × 224)	8	0.0001
ResNet18	(224 × 224)	8	0.001
DenseNet201	(224 × 224)	8	0.001
TactNet	(50 × 50)	64	0.01
MLP	(1 × 250)	64	0.01
LSTM	(5 × 50)	64	0.01
1D-CNN	(5 × 50)	64	0.01
2D-CNN	(50 × 50)	64	0.01
Hybrid-CNN	(5 × 50) & (50 × 50)	64	0.01

Additionally, all models were trained using the Stochastic Gradient Descent with Momentum (SGDM) optimizer with a momentum factor of 0.9, a commonly adopted setting for stable convergence in deep learning. The training was conducted over 100 epochs to improve the model's feature-extraction performance.

5 Results and Discussion

It is essential to evaluate the effectiveness of the proposed methods for assessing object hardness. In this study, we evaluated performance using accuracy, precision, recall, F1 score, and inference time per sample as our designated performance metrics, as shown in Eqs. (4)–(8):

$$Accuracy = \left(\left(\sum\nolimits_N \frac{tp + tn}{tp + tn + fp + fn}\right)/N\right) \times 100 \tag{4}$$

$$Precision = \left(\left(\sum\nolimits_N \frac{tp}{tp + fp}\right)/N\right) \times 100 \tag{5}$$

$$Recall = \left(\left(\sum\nolimits_N \frac{tp}{tp + fn}\right)/N\right) \times 100 \tag{6}$$

$$F1 = \left(\left(\sum\nolimits_N \frac{2 \times precision \times recall}{precision \times recall}\right)/N\right) \times 100 \tag{7}$$

$$Time = \frac{1}{N}\sum_{i=1}^{N}\left(T_{end}^{(i)} - T_{start}^{(i)}\right) \tag{8}$$

5.1 Performance of Lightweight CNN Models

This research evaluates the effectiveness of the fundamental architecture of Convolutional Neural Networks (CNNs) for hardness detection from tactile data by comparing 1D-CNN and 2D-CNN models across three augmented architectures: VGG-6-Lite, ResNet-8-Lite, and DenseNet-6-Lite. The evaluation is conducted using multiple performance metrics, including Accuracy, Precision, Recall, and F1-score, and the results are reported as mean ± standard deviation (SD) from five-fold cross-validation, as summarized in Table 2.

Table 2: Performance comparison of lightweight 1D-CNN and 2D-CNN models across different architectures (mean ± SD from five-fold cross-validation).

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1D-Vgg6-Lite	86.67 ± 1.03	86.60 ± 1.01	86.67 ± 1.03	86.60 ± 1.02
1D-ResNet8-Lite	87.94 ± 1.42	87.97 ± 1.44	87.93 ± 1.42	87.90 ± 1.42
1D-DenseNet6-Lite	86.20 ± 0.80	86.41 ± 0.88	86.20 ± 0.80	86.15 ± 0.87
2D-Vgg6-Lite	86.27 ± 2.67	86.15 ± 2.71	86.27 ± 2.67	86.17 ± 2.72
2D-ResNet8-Lite	84.53 ± 2.12	84.56 ± 2.18	84.53 ± 2.12	84.53 ± 2.15
2D-DenseNet6-Lite	83.00 ± 1.56	83.05 ± 1.53	83.00 ± 1.56	83.00 ± 1.53

All designs show that 1D-CNN models outperform 2D-CNN models. In all models, 1D-ResNet8-Lite performed best, suggesting that residual connections may help in learning temporal tactile patterns by preserving discriminative information across layers. The 1D-Vgg6-Lite model also performed well, demonstrating that sequential convolutional structures can extract temporal characteristics from tactile inputs. Although 1D-DenseNet6-Lite had slightly lower accuracy than the other 1D architectures, it still performed comparably, suggesting that dense feature reuse can yield robust feature representations on tiny tactile datasets. The best 2D-CNN model was VGG-6-Lite, followed by ResNet-8-Lite and DenseNet-6-Lite. This shows that more complex connection pathways may not fully benefit altered tactile signal image representations. Since tactile images have simple spatial structures and low inter-class variance, straightforward convolutional architectures may be better suited to this work than those with more sophisticated feature propagation processes. Similarly to classification performance, the five-fold cross-validation standard deviation indicates model stability. Although 1D-ResNet8-Lite had the best accuracy, both ResNet8-Lite and DenseNet6-Lite had lower standard deviation than Vgg6-Lite, suggesting more consistent performance across data splits. DenseNet6-Lite was the most stable model, demonstrating that dense feature propagation and feature reuse techniques may increase resilience and reduce sensitivity to training data. These data suggest that both model stability and classification accuracy should be evaluated for tactile hardness recognition performance.

Fig. 6 shows confusion matrices for the best-performing models in each domain, 1D-ResNet8-Lite and 2D-Vgg6-Lite. Both models correctly classified Shore A 00 with negligible misclassification across all folds, demonstrating that the lightest hardness level produces unique tactile signal features. However, the Shore A 10 and Shore A 20 classes regularly exhibit greater uncertainty, suggesting that surrounding hardness levels elicit overlapping tactile responses. We also found that Shore A 20 samples are more often misclassified as Shore A 10 than *vice versa*. This shows that the models anticipate some higher-hardness samples as intermediate hardness, indicating lower separability across nearby classes. In addition, 1D-CNN and 2D-CNN models exhibit comparable misclassification trends. This consistency suggests that similarities in tactile signals, rather than restrictions on model design or input representation, are responsible for the difficulty of identifying Shore A 10 from Shore A 20.

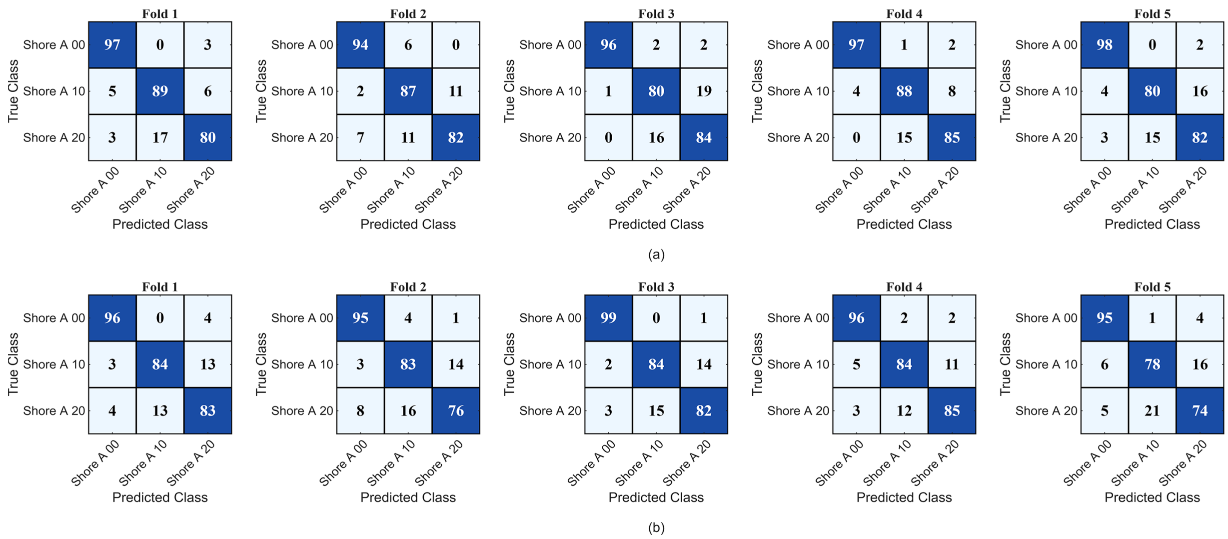


Figure 6: Confusion matrices across five folds for the best-performing models in each domain: (**a**) 1D-ResNet8-Lite; (**b**) 2D-Vgg6-Lite.

5.2 Performance of Hybrid-CNN

To evaluate whether temporal and spatial tactile representations provide complementary information for hardness recognition, this study developed and assessed four Hybrid-CNN models by integrating 1D-CNN and 2D-CNN branches. Three models were constructed using the same architecture across both domains, namely Hybrid-CNN-Vgg, Hybrid-CNN-ResNet, and Hybrid-CNN-DenseNet. In addition, Hybrid-CNN-ResVgg was designed by combining the best-performing 1D-CNN model from Section 5.1, 1D-ResNet8-Lite, with the best-performing 2D-CNN model, 2D-Vgg6-Lite. The performance results are summarized in Table 3.

Table 3: Performance comparison of Hybrid-CNN architectures for hardness recognition.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Hybrid-CNN-Vgg	88.53 ± 3.12	88.46 ± 3.17	88.53 ± 3.12	88.47 ± 3.17
Hybrid-CNN-ResNet	88.73 ± 2.47	88.70 ± 2.49	88.73 ± 2.47	88.66 ± 2.48
Hybrid-CNN-DenseNet	86.80 ± 2.39	86.79 ± 2.51	86.80 ± 2.39	86.71 ± 2.47
Hybrid-CNN-ResVgg	89.47 ± 3.10	89.46 ± 3.15	89.47 ± 3.10	89.43 ± 3.12

Hybrid-CNN-ResVgg had the greatest mean classification accuracy. This shows that hardness recognition may benefit from complementing temporal and spatial feature representations from the highest-accuracy 1D and 2D branches. While 1D-ResNet8-Lite scored 87.94 ± 1.42% accuracy, Hybrid-CNN-ResVgg outperformed it by 1.53%. This improvement suggests that temporal–spatial feature fusion provides more discriminative information than temporal tactile signals alone, but the larger standard deviation indicates greater fold-to-fold variability. While Hybrid-CNN-DenseNet had poorer accuracy, Hybrid-CNN-ResNet and Vgg performed similarly for the same architectural hybrid setups. Same-architecture fusion performed competitively, but Hybrid-CNN-ResVgg's best-domain pairing achieved the highest average accuracy. This shows that merging branches with complementary architectural traits may be better than matching identical structures across temporal and geographical domains. Ablation research was performed on the best Hybrid-CNN configuration, Hybrid-CNN-ResVgg, to assess the contribution of the fusion head. Table 4 shows that the model with the fully connected (FC) fusion layer achieved a higher average classification accuracy. The FC layer increased classification accuracy by 0.80%, demonstrating that the FC-based fusion head refines concatenated temporal–spatial data before classification. The greater standard deviation suggests that this gain is mostly in average classification performance rather than fold stability.

Table 4: Ablation study of the FC-based fusion head in Hybrid-CNN-ResVgg.

Hybrid-CNN-ResVgg	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
No FC	88.67 ± 2.66	88.62 ± 2.71	88.67 ± 2.66	88.60 ± 2.69
With FC	89.47 ± 3.10	89.46 ± 3.15	89.47 ± 3.10	89.43 ± 3.12

These findings demonstrate that Hybrid-CNN fusion can improve hardness recognition performance by integrating temporal and spatial tactile representations. Nevertheless, the improvement over the best single-domain model remains moderate, indicating that the similarity between adjacent hardness classes and the limited complexity of the tactile dataset may influence the benefit of fusion. Therefore, the proposed Hybrid-CNN-ResVgg provides the best overall classification performance among the evaluated lightweight configurations, while maintaining a relatively compact architecture suitable for further comparison with standard and tactile perception models in the next section.

5.3 Comparative Evaluation of Accuracy and Computational Efficiency

To provide a broader comparison, the selected lightweight CNN and Hybrid-CNN models were evaluated against non-CNN baseline models, standard CNN architectures, and tactile perception models. The non-CNN baseline models include MLP and LSTM, representing flattened-feature learning and sequence-based learning, respectively. The standard CNN architectures include VGG16, ResNet18, and DenseNet201, while the tactile perception models include TactNet4, TactNet6, and TactResNet. The selected models from this study include 1D-ResNet8-Lite, 2D-Vgg6-Lite, and Hybrid-CNN-ResVgg. The comparison was conducted on classification performance, computational efficiency, and inference time to evaluate both the predictive capability and the practical feasibility of lightweight tactile hardness recognition. As in the previous evaluations, the classification metrics were reported using the same five-fold cross-validation protocol. The results are summarized in Table 5. Integrating temporal and spatial tactile representations yielded the highest mean classification accuracy for Hybrid-CNN-ResVgg. The CNN-based models in this study—1D-ResNet8-Lite, 2D-Vgg6-Lite, and Hybrid-CNN-ResVgg—performed better than the non-CNN baseline models. This shows that convolutional feature extraction captures discriminative tactile patterns more effectively than flattened feature learning. LSTM, which models sequential dependencies, performed worse than these lightweight CNN and Hybrid-CNN models, suggesting that local temporal feature extraction using convolutional operations may be better for tactile hardness recognition than recurrent sequence modeling alone. The comparison with typical CNN architectures demonstrates that larger image-based CNN models may not perform better with an altered tactile image representation. VGG16, ResNet18, and DenseNet201 are commonly used in image classification, yet they performed similarly to the lightweight CNN and Hybrid-CNN models in our investigation. The basic spatial organization of the generated tactile images, along with their low inter-class variance across neighboring hardness levels, may explain this. In this case, large-scale conventional CNN models may increase architectural complexity without improving classification performance. Because they were designed for tactile learning, tactile perception models are another important similarity. They still perform worse than Hy-brid-CNN-ResVgg in this research. The Hybrid-CNN configuration, which combines temporal information from the original tactile sequence with spatial representations derived from altered tactile signals, appears better suited to identifying tactile hardness. Thus, Hybrid-CNN-ResVgg's ability to incorporate complementary representations, rather than relying on a single input representation or a generic CNN architecture, seems to be its advantage. Table 5 shows that model size and GFLOPs were used to assess computing efficiency in addition to classification performance. Traditional CNN models require more storage and processing resources than the lightweight models developed in this work, potentially limiting their usefulness for resource-constrained tactile sensing applications. In contrast, 1D-ResNet8-Lite, 2D-Vgg6-Lite, and Hybrid-CNN-ResVgg achieve competitive classification performance with small model sizes and low GFLOPs. Hybrid-CNN-ResVgg has two branches, but it is a smaller model with lower computational cost than traditional CNN systems. The dual-branch architecture does not always increase computational load when lightweight component networks are employed. For practical tactile perception systems on resource-constrained platforms, memory utilization, computational cost, and reaction time are crucial. Hybrid-CNN-ResVgg's compactness is crucial.

Table 5: Comparative evaluation of classification performance and computational efficiency across the evaluated models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Model Size (MB)	GFLOPs
MLP	83.13 ± 2.39	83.32 ± 2.24	83.13 ± 2.39	82.87 ± 2.51	0.12	0.00006
LSTM	80.73 ± 1.79	80.98 ± 2.24	80.73 ± 1.79	80.49 ± 1.70	0.27	0.0068
VGG16	81.27 ± 2.09	81.64 ± 1.39	81.27 ± 2.09	81.22 ± 2.08	512.2	30.9459
ResNet18	83.67 ± 2.07	83.59 ± 2.02	83.67 ± 2.07	83.58 ± 2.03	42.7	1.7965
DenseNet201	83.13 ± 2.57	83.08 ± 2.52	83.13 ± 2.57	83.07 ± 2.55	70	3.5720
TactNet4	83.67 ± 2.37	83.69 ± 2.18	83.67 ± 2.37	83.63 ± 2.29	0.04	0.0060
TactNet6	85.20 ± 3.44	85.20 ± 3.39	85.20 ± 3.44	85.18 ± 3.40	0.40	0.0102
TactResNet	78.13 ± 1.91	78.47 ± 2.12	78.13 ± 1.91	78.10 ± 1.96	2.9	0.2398
1D-ResNet8-Lite	87.94 ± 1.42	87.97 ± 1.44	87.93 ± 1.42	87.90 ± 1.42	0.06	0.0002
2D-Vgg6-Lite	86.27 ± 2.67	86.15 ± 2.71	86.27 ± 2.67	86.17 ± 2.72	0.39	0.0392
Hybrid-CNN-ResVgg	89.47 ± 3.10	89.46 ± 3.15	89.47 ± 3.10	89.43 ± 3.12	0.46	0.0394

Fig. 7a shows CPU and GPU inference time per sample for assessed models. Due to their larger model size and computational complexity, inference times for traditional CNN architectures are longer, particularly on CPU. Inference times are shorter for non-CNN baselines and lightweight models. Because it operates on flattened feature vectors with a basic fully connected structure, the MLP achieved the shortest inference time among the evaluated models, whereas the LSTM required a longer inference time owing to its sequential computation over temporal inputs. The selected lightweight CNN and Hybrid-CNN models also show reasonable inference times, supporting their potential for future real-time tactile hardness identification. Classification accuracy and model size are further contrasted in Fig. 7b. Standard CNN models are larger but have lower classification accuracy. Tactile perception models are smaller than CNNs but less accurate than the Hybrid-CNN model. In contrast, this study's models lie in a positive trade-off space, achieving excellent classification accuracy with minimal model size. The most balanced is Hybrid-CNN-ResVgg, which has the best mean classification accuracy and the lowest computational complexity. These findings show that Hybrid-CNN-ResVgg balances classification performance and computational economy. Unlike typical CNN designs, which require large model sizes and high computational cost, the hybrid configuration delivers strong recognition performance with a compact structure. This shows that lightweight temporal and spatial feature extractors might improve tactile hardness detection, especially in applications that need precision and real-time processing efficiency.

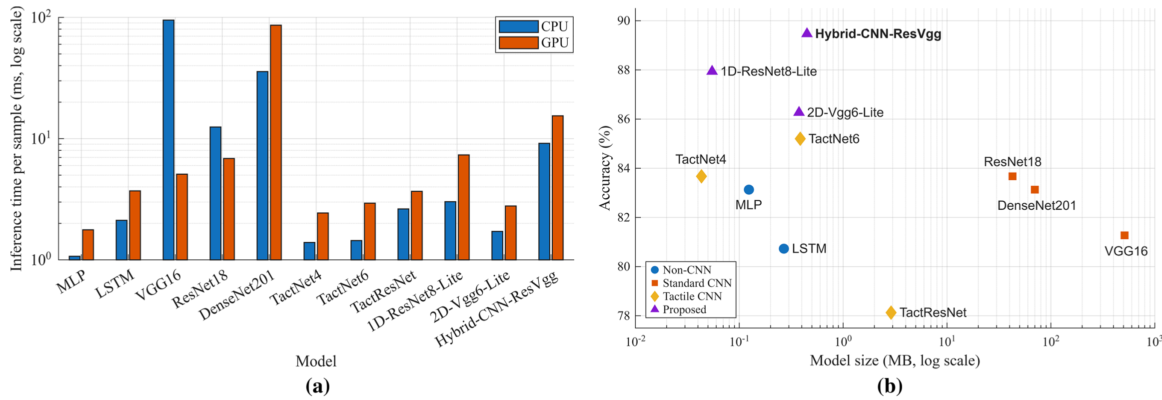


Figure 7: Computational efficiency of the evaluated models: (**a**) CPU and GPU inference time per sample; (**b**) accuracy–model size trade-off.

5.4 Discussion

The experiments illustrate how temporal, spatial, and hybrid feature representations affect tactile hardness identification. The findings show that temporal tactile information dominates hardness differentiation. Learning directly from sequential tactile sensor data better preserves discriminative patterns than using modified image representations, as 1D-CNN models consistently outperformed 2D models. In tactile hardness perception, the sensor response varies over time when touching materials of differing hardnesses. The poorer performance of 2D-CNN models does not mean modified tactile p images are useless. Instead of replacing the temporal signal, the 2D representation appears to provide complementary information. Since tactile images are derived from pressure-based sensor data, they may include spatially ordered patterns representing sensor responses. The modification may also eliminate small temporal differences that distinguish neighboring hardness levels. This is why 2D-CNN models were less effective than 1D-CNN models but remained beneficial when incorporated into the Hybrid-CNN architecture. The performance enhancement by Hybrid-CNN-ResVgg suggests that temporal and spatial representations may work together for hardness recognition. The 1D branch captures local temporal fluctuations from tactile signals, whereas the 2D branch learns transformed spatial patterns from interaction data. Combining these two representations allows the Hybrid-CNN model to catch information that neither branch alone can. Improvement over the best single-domain model is modest, suggesting that the converted 2D representation contains additional, albeit limited, discriminative information. The confusion matrix analysis sometimes confused neighboring hardness classes, such as Shore A 10 and Shore A 20, due to their fundamental similarities. All fingers provide tactile impulses, reflecting the distribution of grasp-based tactile perception. Under controlled settings, the robotic hand touched hardness samples, finger geometry, gripping position, and contact location may affect each finger's contribution. Instead of relying on a single contact point, using all fingertip sensors enables the model to capture a more comprehensive tactile response pattern. This design mimics human clutching, in which tactile input from multiple fingers infers hardness. These findings suggest that sensor-channel optimization may be feasible, as certain central fingers may contribute more consistently in specific gripping situations. The contrast with ordinary CNN architectures underscores the need for tactile-data-specific models. VGG16, ResNet18, and DenseNet201 are large CNN models that learn complicated visual patterns from real images. In contrast, this work employed tactile image representations built from low-dimensional sensor inputs with simplified spatial structures. Thus, model size and depth may not always improve classification. The lightweight CNN and Hybrid-CNN models developed in this work better balance representation capacity and computational cost for tactile hardness identification. In practice, the lightweight and Hybrid-CNN models' small model size, low GFLOPs, and fast inference time make tactile hardness identification computationally efficient. These results indicate that lightweight temporal–spatial feature fusion may balance recognition performance and computational cost.

5.5 Limitations

Although the proposed lightweight Hybrid-CNN framework demonstrated promising performance for tactile hardness recognition, several limitations should be considered. (1) The experiments were conducted under controlled conditions using only three hardness levels with identical object geometry and material type. While this design supports systematic evaluation by reducing external variability, it also limits the diversity of real-world tactile conditions, where objects may differ in geometry, material composition, surface roughness, and contact behavior. (2) This study used pressure-based tactile signals as the primary input. Although pressure variation is highly relevant to hardness perception, pressure signals alone may not fully capture more complex tactile properties such as shear force, vibration, slip, temperature, and texture-related responses. Therefore, the current sensing configuration should be regarded as an initial pressure-based framework rather than a complete representation of tactile perception. Finally, the proposed approach was evaluated in an offline setting, and online embedded deployment has not yet been validated. Although the model size, GFLOPs, and inference time results indicate computational feasibility for real-time tactile hardness recognition, practical performance under continuous sensing, real-time data acquisition, and robotic operation remains to be examined. Thus, the results should be interpreted as evidence of computational feasibility rather than direct confirmation of real-time embedded implementation.

6 Conclusion

This study presented a lightweight deep learning framework for tactile hardness recognition using low-cost pressure-based tactile sensor signals. Lightweight 1D-CNN, 2D-CNN, and Hybrid-CNN models were designed and evaluated to investigate the effectiveness of temporal, spatial, and fused temporal–spatial feature representations. The results showed that 1D-CNN models were more effective than 2D-CNN models, indicating that temporal variations in tactile sensor responses provide important discriminative information for hardness recognition. The proposed Hybrid-CNN-ResVgg achieved the highest mean classification accuracy among the evaluated lightweight configurations by combining the highest-accuracy 1D and 2D branches. Compared with non-CNN baseline models, standard CNN architectures, and tactile perception models, the selected lightweight and Hybrid-CNN models provided a favorable balance between classification performance, model compactness, computational complexity, and inference time. These findings suggest that lightweight temporal–spatial feature fusion is a promising strategy for tactile hardness recognition, particularly for applications that require efficient, compact tactile perception models. Future work will expand the dataset to include a wider range of hardness levels, materials, geometries, and surface conditions. In addition, multimodal tactile sensing, sensor-channel reduction, and model optimization techniques such as pruning, quantization, and knowledge distillation should be explored to improve generalization and deployment efficiency in real-time robotic systems.

Acknowledgement: This work was conducted in the Signal Processing Research Laboratory, Department of Electronics and Telecommunication Engineering, RMUTT, Thailand.

Funding Statement: This research was supported by the National Science, Research and Innovation Fund, Thailand Science Research and Innovation (TSRI), through Rajamangala University of Technology Thanyaburi (FRB69E0717) (Grant No.: FRB690069/0168).

Author Contributions: Conceptualization: Thossapon Kaewrakmuk and Jakkree Srinonchat; methodology and software: Thossapon Kaewrakmuk and Jakkree Srinonchat; validation: Thossapon Kaewrakmuk and Jakkree Srinonchat; formal analysis: Jakkree Srinonchat; investigation: Thossapon Kaewrakmuk and Jakkree Srinonchat; resources: Thossapon Kaewrakmuk and Jakkree Srinonchat; data curation: Thossapon Kaewrakmuk and Jakkree Srinonchat; writing—original draft preparation: Thossapon Kaewrakmuk; writing—review and editing: Thossapon Kaewrakmuk and Jakkree Srinonchat; visualization: Jakkree Srinonchat; supervision: Jakkree Srinonchat; project administration: Jakkree Srinonchat; project administration: Jakkree Srinonchat; funding acquisition: Jakkree Srinonchat. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets generated and analyzed in the current study are available from the corresponding author upon reasonable request.

Ethics Approval: This study did not require ethical approval as it did not involve human participants or animals.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Su H, Qi W, Chen J, Yang C, Sandoval J, Laribi MA. Recent advancements in multimodal human-robot interaction. Front Neurorobot. 2023;17:1084000. doi:10.3389/fnbot.2023.1084000.
2. Saffre F, Hildmann H, Heikkila E, Malm T, Pakkala D. Dynamic and probabilistic safety zones for autonomous mobile robots operating near humans. Results Eng. 2024;23(1):102731. doi:10.1016/j.rineng.2024.102731.
3. Jassim HS, Akhter Y, Aalwahab DZ, Neamah HA. Recent advances in tactile sensing technologies for human-robot interaction: current trends and future perspectives. Biosens Bioelectron X. 2025;26:100669. doi:10.1016/j.biosx.2025.100669.
4. Saudabayev A, Varol HA. Sensors for robotic hands: a survey of state of the art. IEEE Access. 2015;3:1765–82. doi:10.1109/ACCESS.2015.2482543.
5. Piazza C, Grioli G, Catalano M, Bicchi A. A century of robotic hands. Annu Rev Control Robot Auton Syst. 2019;2(1):1–32. doi:10.1146/annurev-control-060117-105003.
6. Nguyen TT, Duy CV. Grasping moving objects with incomplete information in a low-cost robot production line using contour matching based on Hu moments. Results Eng. 2024;23(5):102414. doi:10.1016/j.rineng.2024.102414.
7. Xi J, Huaiwen Y, Li X, Wei R, Zhang T, Dong L, et al. Recent advances in tactile sensory systems: mechanisms, fabrication, and applications. Nanomaterials. 2024;14(5):465. doi:10.3390/nano14050465.
8. Wang C, Takele NA, Derege O, Rifiki N, Khalil ME, Tiwari V, et al. Tactile sensing technology in bionic skin: a review. Biosens Bioelectron. 2023;220(3):114882. doi:10.1016/j.bios.2022.114882.
9. Meribout M, Takele NA, Derege O, Rifiki N, Khalil ME, Tiwari V, et al. Tactile sensors: a review. Measurement. 2024;238(10):115332. doi:10.1016/j.measurement.2024.115332.
10. Bonci A, Cheng PDC, Indri M, Nabissi G, Sibona F. Human-robot perception in industrial environments: a survey. Sensors. 2021;21(5):1571. doi:10.3390/s21051571.
11. Wang Y, Hao T, Liu Y, Xiao H, Liu S, Zhu H. Anthropomorphic soft hand: dexterity sensing and machine learning. Actuators. 2024;13(3):84. doi:10.3390/act13030084.
12. Jin J, Wang S, Zhang Z, Mei D, Wang Y. Progress on flexible tactile sensors in robotic applications on objects properties recognition, manipulation and human-machine interactions. Soft Sci. 2023;3:8. doi:10.20517/ss.2022.34.

13. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Proceedings of the Advances in Neural Information Processing Systems; 2012 Dec 3–6; Lake Tahoe, NV, USA.
14. Filipowska A, Filipowski W, Raif Pł, Pieniążek M, Bodak J, Ferst P, et al. Machine learning-based gesture recognition glove. Sensors. 2024;24(18):6157. doi:10.3390/s24186157.
15. Zafar MH, Langås EF, Sanfilippo F. Empowering human-robot interaction using sEMG sensor: hybrid deep learning model for accurate hand gesture recognition. Results Eng. 2023;20(2):101639. doi:10.1016/j.rineng.2023.101639.
16. Chansri C, Srinonchat J. Utilizing gramian angular fields and convolution neural networks in flex sensors glove for human-computer interaction. IEEE Trans Hum Mach Syst. 2024;54(4):475–83. doi:10.1109/THMS.2024.3404101.
17. Qian X, Deng W, Wang W, Liu Y, Jiang L. Mining local and global spatiotemporal features for tactile object recognition. Front Neurorobot. 2024;18:1387428. doi:10.3389/fnbot.2024.1387428.
18. Drimus A, Kootstra G, Bilberg A, Kragic D. Design of a flexible tactile sensor for classification of rigid and deformable objects. Robot Auton Syst. 2014;62(1):3–15. doi:10.1016/j.robot.2012.07.021.
19. Deng L, Shen Y, Fan G, He X, Li Z, Yuan Y. Design of a soft gripper with improved microfluidic tactile sensors for classification of deformable objects. IEEE Robot Autom Lett. 2022;7(2):5607–14. doi:10.1109/LRA.2022.3158440.
20. Moreno HA, Moreno LA, Valentín-Coronado LM, Flores G. Soft paw sensor for tactile and force sensing in legged robots. Mechatronics. 2025;111(16):103407. doi:10.1016/j.mechatronics.2025.103407.
21. Lv Y, Ma Z, Duan J, Sun G, Wang P, Qu S, et al. Biomimetic hydrogel-based sensors with dual-mode dynamic-static tactile sensing capability enabling robotic hand for intelligent material property recognition. InfoMat. 2025;7(11):e70041. doi:10.1002/inf2.70041.
22. Pohtongkam S, Srinonchat J. Object recognition for humanoid robots using full hand tactile sensor. IEEE Access. 2023;11(2):20284–97. doi:10.1109/ACCESS.2023.3249573.
23. Funabashi S, Yan G, Hongyi F, Schmitz A, Jamone L, Ogata T, et al. Tactile transfer learning and object recognition with a multifingered hand using morphology specific convolutional neural networks. IEEE Trans Neural Netw Learn Syst. 2024;35(6):7587–601. doi:10.1109/TNNLS.2022.3215723.
24. Sharma Y, Guo C, Beatty M, Justham L, Ferreira P. Mechanoreceptor-inspired tactile sensor topological configurations for hardness classification in robotic grippers. Electronics. 2025;14(4):674. doi:10.3390/electronics14040674.
25. Liu X, Yang W, Meng F, Sun T. Material recognition using robotic hand with capacitive tactile sensor array and machine learning. IEEE Trans Instrum Meas. 2025;73:9508309. doi:10.1109/TIM.2024.3383886.
26. Song Y, Lv S, Wang F, Li M. Hardness-and-type recognition of different objects based on a novel porous graphene flexible tactile sensor array. Micromachines. 2023;14(1):217. doi:10.3390/mi14010217.
27. Kaewrakmuk T, Srinonchat J. Multisensor data fusion and time series to image encoding for hardness recognition. IEEE Sens J. 2024;24(16):26463–71. doi:10.1109/JSEN.2024.3426045.
28. Amin Y, Gianoglio C, Valle M. Embedded real-time objects' hardness classification for robotic grippers. Future Gener Comput Syst. 2023;148(22):211–24. doi:10.1016/j.future.2023.06.002.
29. Gao Y, Hendricks LA, Kuchenbecker KJ, Darrell T. Deep learning for tactile understanding from visual and haptic data. In: Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA); 2016 May 16–21; Stockholm, Sweden. p. 536–43.
30. Zhang P, Zhou M, Shan D, Chen Z, Wang X. Object description using visual and tactile data. IEEE Access. 2022;10:54525–36. doi:10.1109/ACCESS.2022.3174874.
31. Pastor F, García-González J, Gandarias JM, Medina D, Closas P, García-Cerezo AJ, et al. Bayesian and neural inference on lstm-based object recognition from tactile and kinesthetic information. IEEE Robot Autom Lett. 2021;6(1):231–38. doi:10.1109/LRA.2020.3038377.
32. Lu X, Sun D, Yin H, Xu H, Yan Y, Wu C, et al. 3-D tactile-based object recognition for robot hands using force-sensitive and bend sensor arrays. IEEE Trans Cogn Dev Syst. 2023;15(4):1645–55. doi:10.1109/TCDS.2022.3215021.
33. Li B, Li L, Wang H, Chen G, Wang B, Qiu S. TVT-transformer: a tactile-visual-textual fusion network for object recognition. Inf Fusion. 2025;118(4):102943. doi:10.1016/j.inffus.2025.102943.
34. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. 2014.

35. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NA, USA. p. 770–8.
36. Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21–26; Honolulu, HI, USA. p. 2261–69.
37. Chatzistergos PE, Allan D, Chockalingam N, Naemi R. Shore hardness is a more representative measurement of bulk tissue biomechanics than of skin biomechanics. Med Eng Phys. 2022;105(1):103816. doi:10.1016/j.medengphy.2022.103816.
38. Gandarias JM, García-Cerezo AJ, Gómez-de-Gabriel JM. CNN-based methods for object recognition with high-resolution tactile sensors. IEEE Sens J. 2019;19(16):6872–82. doi:10.1109/JSEN.2019.2912968.
39. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323(6088):533–6. doi:10.1038/323533a0.
40. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi:10.1162/neco.1997.9.8.1735.