Tactile Texture Classification on Uneven Surfaces Using a Neural Network Soft Voting Ensemble

by

Soheil Khatibi

A thesis submitted to the
Department of Computer Science
in conformity with the requirements for
the degree of Master of Science

Lakehead University
Thunder Bay, Ontario, Canada
August 2025

Abstract

With the growing capabilities of intelligent robots in object recognition and manipulation, the ability to sense and interpret physical contact through touch has become a crucial component to enabling effective interaction with the physical world. Although tactile texture classification on flat surfaces has been broadly studied in recent years, uneven surfaces pose additional challenges due to variations in contact geometry and surface normals. To address these challenges, this study introduces a new tactile texture dataset comprising both flat surfaces and several distinct uneven surfaces, and proposes a soft voting-based classification system built on deep neural networks, which combines predictions from multiple temporal window sizes to improve robustness.

The dataset is collected using a compliant tactile sensor mounted on the end effector of a UFactory Lite6 robotic arm that combines MARG and barometric data for capturing dynamic contact interactions. The dataset includes six types of uneven surfaces, each including a variety of textures to create diverse and challenging contact conditions. To improve classification robustness and enable multi-scale analysis, the time-series data are segmented using a sliding window approach with varying window sizes. Multiple model architectures are trained on the windowed segments, including 1D Convolutional Neural Networks (1D-CNNs), Bidirectional Long Short-Term Memory (BiLSTM) networks, hybrid 1D-CNN–BiLSTM models, self-attention-based

networks, and hybrid 1D-CNN–self-attention models. Their predictions are combined using a soft voting strategy to enhance overall classification accuracy.

Experimental results based on 5-fold cross-validation demonstrate that self-attention-based models consistently outperform other individual architectures across all window sizes. Moreover, the proposed voting system, which combines predictions from different window sizes, further improves classification performance for all model types by leveraging complementary temporal features.

This study demonstrates that combining deep neural networks with a soft voting mechanism across multiple window sizes enables accurate tactile texture classification on various types of uneven surfaces, contributing toward more robust and adaptable robotic perception in complex environments.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Thiago E. Alves de Oliveira, for his unwavering support, guidance, and encouragement throughout the course of my Master's research. His exceptional mentorship, insightful feedback, and dedication to fostering my academic growth have been truly invaluable. The level of care and attention he has given to my work has shaped not only this thesis but also my approach to research and problem-solving.

I would also like to extend heartfelt thanks to Amirmasoud Azadfar, whose friendship, presence, and unwavering support have been a constant source of motivation and comfort throughout this journey. His encouragement and companionship have meant a great deal to me.

Special thanks to Frederico L. M. de Sousa; the time we knew and saw each other was not long, but he is truly a great person and a genuine friend.

I am also grateful to the faculty and staff at the Department of Computer Science, Lakehead University, for providing a supportive research environment and valuable learning opportunities.

Thanks as well to my collaborators and labmates for their feedback, ideas, and camaraderie during our discussions and projects. Finally, I am indebted to my friends and family for their constant encouragement, especially during challenging times.

This work would not have been possible without the support and inspiration I received from all those around me.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent decades, robotic systems have made significant progress in perceiving and interacting with the physical world, primarily driven by advancements in computer vision, manipulation, and artificial intelligence [3, 4]. These developments have enabled robots to perform increasingly complex tasks in structured environments such as manufacturing, logistics, and medical applications. However, true physical interaction remains an open challenge, particularly in unstructured or unpredictable environments. One critical yet underdeveloped component enabling such interaction is *tactile perception*, the ability of robots to sense and interpret physical contact with objects or surfaces [5].

Biological systems, particularly humans, rely extensively on tactile information for object recognition, surface exploration, and manipulation [6]. This sense provides complementary feedback to vision, especially in scenarios where visual cues are occluded, ambiguous, or absent. In robotics, tactile perception bridges the gap between sensing and action, playing an essential role in manipulation tasks such as grasping, insertion, and surface following. Without touch, robotic systems are essentially blind to the details of contact dynamics, force distribution, and material properties.

1.1 From Sensing to Perception: The Role of Tactile Sensors

Tactile sensing in robotics involves transduction mechanisms that convert physical interactions—such as pressure, vibration, or shear—into electrical signals. Numerous tactile sensors have been developed, including capacitive, piezoresistive, optical, and barometric types, each with varying resolution, compliance, sensitivity, and integration feasibility [3, 1]. These technologies form the foundation for tactile perception by providing the raw measurements needed for higher-level interpretation.

In this work, a compliant tactile sensor is employed that combines barometric pressure sensing and inertial measurement (IMU) data. This setup is lightweight, low-cost, and easily integrated into the end-effector of a robotic manipulator. The barometric sensor captures static and dynamic pressure variations during contact, while the IMU provides six-axis motion data, including linear acceleration and angular velocity. Together, they offer a rich, multi-modal representation of dynamic tactile interactions.

This configuration builds upon prior work by Marzani et al., who used the same sensing setup to classify textures on flat surfaces [7]. The present approach extends this to more challenging uneven surfaces, which more closely resemble real-world contact conditions where surface normals, contact forces, and motion patterns are irregular.

1.2 Point Collection and Interaction Strategy

Tactile data collection in robotic systems must be controlled and reproducible. In our setup, a robotic manipulator equipped with the tactile sensor follows a predefined trajectory to probe the surface of interest. A surface reconstruction pipeline

generates probing points—each associated with a 3D position and local surface normal—ensuring consistent contact across the texture.

Each probing point serves as a local site for dynamic contact. The robot first moves to a hover position above the target, then descends along the surface normal until contact is made. During this descent and subsequent contact phase, time-series data from the barometric and inertial sensors are recorded at approximately 130 Hz. The robot then retracts and moves to the next point, repeating this process across the surface.

This point-wise probing strategy provides spatial coverage while preserving localized temporal patterns of tactile interaction, enabling downstream segmentation, analysis, and learning.

1.3 Tactile Data Collection on Uneven Surfaces

Unlike flat surfaces, uneven surfaces introduce significant variability in contact dynamics. Changes in curvature, slope, and material compliance result in varying pressure distributions and motion signatures [8]. To address this complexity, a dataset was designed comprising one flat and five custom uneven surfaces, each covered with twelve distinct textures ranging from fabrics and polymers to woods and leathers.

For each texture–surface combination, the robot performs multiple probing sequences, recording barometric and IMU data at each point. These recordings are synchronized and stored as ROS 2 bag files, later processed to extract normalized time-series data. The resulting dataset is one of the first to systematically capture dynamic tactile interactions over irregular geometries with labeled texture annotations.

1.4 From Tactile Signals to Perception and Classification

Raw sensor signals alone do not constitute perception. The challenge lies in interpreting these time-series data streams to infer meaningful semantic labels—such as texture class—across a range of surfaces. Traditional texture classification approaches have relied on hand-crafted features, such as statistical descriptors or frequency-domain analyses, which often lack robustness to complex, real-world contact scenarios [4]. Deep learning methods, by contrast, can automatically learn discriminative representations from raw data, enabling improved generalization.

In this thesis, deep neural networks are trained on fixed-length windows of tactile data, capturing temporal segments of the contact interaction. To improve robustness, multiple window sizes—short (128 samples), medium (256), and long (512)—are used, each feeding into a separate neural network. Predictions from these models are combined via a soft voting strategy, yielding a more reliable classification outcome.

1.5 Challenges in Texture Recognition on Uneven Surfaces

Texture classification on flat surfaces has been extensively studied [9], but uneven surfaces pose additional challenges. Varying surface normals alter how the sensor engages with the material, while differences in compliance and local geometry influence signal evolution. Barometric readings may exhibit nonlinear or multi-peaked profiles, and IMU signals may contain oscillatory patterns or noise. These factors make signal interpretation more difficult, reducing classification accuracy if not properly addressed.

The proposed method mitigates these challenges by capturing diverse signal patterns at multiple temporal resolutions and combining model outputs through a voting

scheme.

1.6 Research Objectives

The primary objective of this research is to develop a robust tactile texture classification system that generalizes across flat and uneven surfaces. This involves designing a comprehensive data collection protocol, developing deep learning models capable of learning temporal patterns, and implementing a voting mechanism to fuse multiple temporal resolutions. The approach is validated through a comparative analysis of single-window and multi-window systems across all surface types in the dataset.

1.7 Contributions

The main contributions of this thesis are as follows:

- **Tactile dataset:** Development of a novel tactile dataset collected using the BioIn-Tacto sensing module, which combines barometric pressure and inertial measurement data. The dataset covers six different surfaces (one flat and five uneven) and twelve different textures.
- **Single-window classifier:** Design and implementation of a neural network classifier based on an attention-based architecture, trained on fixed-length time-series windows extracted from tactile interaction sequences.
- **Multi-resolution voting framework:** Introduction of a soft voting system that integrates predictions from multiple temporal resolutions (128, 256, and 512 samples) to improve classification robustness, particularly on uneven surfaces.

- **Data acquisition and processing pipeline:** Development of a reproducible ROS2-based pipeline for tactile data acquisition, surface probing, signal normalization, and dataset preparation.
- **Experimental evaluation:** Comparative analysis of single-window models and the proposed voting system, demonstrating improved accuracy and robustness in texture classification across both flat and uneven surfaces.

1.8 Thesis Organization

Chapter 2 reviews tactile sensing, texture classification, and relevant machine learning methods. Chapter 3 describes the hardware, surface design, data collection protocol, and model training process. Chapter 4 presents experimental results and comparative analyses. Chapter 5 summarizes findings and outlines future research directions in robotic tactile perception.

Chapter 2

Literature Review

Tactile texture recognition is a fundamental component of robotic perception, enabling robots to interpret and interact with their surroundings through direct physical contact. This chapter surveys the literature on tactile sensing and perception, highlighting the importance of touch in robotics, sensor technologies, tactile feedback in object recognition and manipulation, and algorithmic approaches to surface and texture understanding. We review prior research on object recognition through touch, surface characterization, and dynamic contact profile classification, noting both the potential and limitations of these methods for robust texture recognition. Particular attention is given to works employing the BioIn-Tacto sensing module, including its use in tactile texture classification on flat surfaces and other applications such as object recognition and manipulation. Finally, we outline the progression from early tactile sensing systems to modern deep learning and ensemble-based methods, identifying key challenges such as handling uneven surfaces and variable contact conditions that motivate the contributions of this thesis.

2.1 Tactile Sensing and Perception in Robotics

Tactile sensing provides robots with physical awareness akin to human touch, playing a vital role in object recognition, manipulation, and interaction in unstructured environments. Compared to vision-based systems, tactile sensors offer reliable, contact-based information, particularly in scenarios involving occlusions, poor lighting, or fragile object handling [6, 5, 3, 4]. Foundational studies have examined tactile sensing as a multi-dimensional process involving pressure, vibration, deformation, and proprioception.

These early efforts informed the development of tactile-enabled prosthetic systems [10], haptic teleoperation hands [11], and robotic systems capable of adjusting manipulation in real time based on tactile feedback [12]. More recent works emphasize tactile perception as a critical requirement in healthcare, manufacturing, and assistive robotics.

Touch sensing for humanoid robots [13] further emphasizes the role of tactile perception in humanoid platforms. The past decade has seen rapid advances in tactile sensor design. Multimodal and bioinspired sensors such as BioIn-Tacto [14, 1] provide high-fidelity pressure and motion data via MARG and barometric sensing. Leveraging compliant tactile perception for haptic blind surface reconstruction [15] illustrates the benefits of end-effector flexibilization for enhanced tactile interaction. Other designs include flexible triboelectric systems [16], magnetorheological elastomer-based sensors [17], and self-healing, multi-material sensors. Fingerpad-inspired skins [18] and dual-inductive sensors [19] reflect the trend toward mimicking biological tactile capabilities.

These advances have improved spatial resolution, robustness under deformation,

and multimodal data capture. Recent designs also integrate tactile perception with proprioception and visual-tactile fusion [20, 21].

2.2 Robotic Manipulation and Tactile Feedback

Exploring tactile temporal features for object pose estimation [12], extraction of non-regular pegs using tactile sensing and reinforcement learning [22], grasp approach under positional uncertainty [23], and fuzzy controlled object manipulation [24] demonstrate the importance of tactile sensing in manipulation strategies.

Computational intelligence and mechatronics solutions for robotic tactile object recognition [25] detail methods for object identification using tactile data. Estimating the orientation of objects from tactile sensing data using machine learning methods [26] connects surface geometry analysis to downstream texture classification.

Data-driven analysis of kinaesthetic and tactile information for shape classification [27] and tactile profile classification using a multimodal MEMs-based sensing module [28] illustrate dynamic contact profile analysis, but also highlight that such profiles alone are insufficient for robust texture classification due to inconsistent contact.

2.3 The BioIn-Tacto Sensing Module in Research

Tactile object recognition in early phases of grasping using underactuated robotic hands [29] demonstrates BioIn-Tacto's ability to contribute to early-stage grasp planning.

Dynamic tactile exploration for texture classification [30], classification of textures using a tactile-enabled finger [31], and a multimodal tactile dataset for dynamic texture classification [32] show BioIn-Tacto's utility for planar surface texture recognition. Heart rate detection using a multimodal tactile sensor [33, 34] extends BioIn-Tacto's applications to physiological monitoring.

2.4 From Flat to Uneven Surfaces

While tactile texture recognition on flat, planar surfaces has been extensively studied [30, 31, 32, 35, 36], real-world applications often involve contact with curved or irregular geometries where the contact patch, surface normal, and applied force vary continuously during exploration. These variations introduce additional challenges, including non-uniform deformation of the sensing surface, inconsistent contact forces, and signal distortions caused by local curvature.

Early work with the BioIn-Tacto sensor focused on flat surfaces, establishing baselines for dynamic tactile texture recognition [30, 31]. More recent studies have sought to bridge this gap by extending tactile perception to uneven or non-planar environments. Marzani et al. [37] presented deep learning approaches for texture recognition on concave–convex surfaces, demonstrating the feasibility of classification despite significant contact variability. Cheret et al. [38] further explored this problem by integrating haptic surface reconstruction with reinforcement learning to enhance texture recognition performance, highlighting the benefits of adaptive trajectory planning for complex geometries.

Khatibi et al. [39] developed self-attention-based neural networks tailored for uneven surface data, addressing challenges related to long-range temporal dependencies

and noisy contact patterns. In parallel, Khatibi et al. [40] introduced a multi-scale voting framework, which combines predictions across different temporal resolutions to improve robustness to irregular contact dynamics.

Despite these advancements, texture recognition on uneven surfaces remains relatively underexplored compared to flat-surface studies. In particular, there is a need for larger and more diverse datasets, improved handling of non-uniform contact mechanics, and modeling strategies that can adapt to varying interaction conditions without overfitting to specific geometries. These limitations continue to motivate further research in this area.

2.5 Tactile Data Collection and Benchmarking

Creating high-quality tactile datasets is essential for benchmarking classification systems. Many datasets [41, 35] focus on planar surfaces and rely on fixed-length time-series signals. Recently, researchers have begun addressing the complexities of uneven geometries. Marzani et al. [7] introduced one of the first datasets collected over concave-convex surfaces, incorporating 12 textures and multimodal sensing (pressure + MARG). Similar works explore dynamic texture classification [32] and peg-based tactile benchmarks [42].

To ensure generalizability, researchers increasingly prevent data leakage by assigning entire experiments to training or validation folds. Standard preprocessing now includes sensor alignment, normalization, and contact filtering during sliding data collection.

2.6 Deep Learning and Temporal Modeling for Tactile Data

Deep learning has transformed tactile signal processing in recent years. Early methods relied on 1D CNNs or LSTM architectures [43], but these models often struggled with noise, long-range dependencies, and generalization. More advanced models incorporate attention mechanisms and transformers [44, 45, 46], enabling both local and global temporal pattern extraction from noisy tactile signals. Self-attention-based models, such as the approach in [39], have demonstrated strong performance on tactile datasets collected from uneven surfaces, improving classification robustness against contact variability.

Cao et al. [47] proposed spatio-temporal attention models, while others introduced hybrid CNN–BiLSTM architectures for multi-scale analysis. Informer-based classifiers and self-attention mechanisms have shown improved accuracy on long tactile sequences, particularly for textures with similar microstructures.

2.7 Texture Recognition on Uneven Surfaces

While progress on flat surfaces is significant [35, 36], texture recognition on uneven surfaces remains underexplored. Key challenges include non-uniform contact, sensor noise from varying curvature, and inconsistent force distribution. Some studies [48, 49, 37] attempt hierarchical and multimodal fusion, while others leverage reinforcement learning for enhancing tactile texture recognition from haptic surface reconstruction [38]. However, few works address dynamic texture recognition on irregular geometries with robust modeling approaches.

Recent approaches combine trajectory planning, surface reconstruction, and tactile sensing [14]. Previous works proposed deep learning pipelines for texture classification after adaptive trajectory execution, yet fixed-length models often fail to capture the variability of real tactile interactions.

2.8 Ensemble Learning for Robust Tactile Classification

Multi-resolution learning offers a promising strategy for handling variability in tactile signals. Voting-based systems combine predictions from models trained on different temporal window sizes (e.g., 128, 256, 512 samples), capturing both short- and long-term features. The multi-scale voting system proposed in [40] soft-averages the outputs from these models, mitigating the limitations of single-scale approaches and improving classification performance on uneven textures. Such ensembles have been shown to outperform the best single model by more than 2% accuracy. The weighting of each temporal resolution can be manually assigned or tuned for optimal balance.

2.9 Summary and Research Gaps

The reviewed literature highlights:

- Major progress in tactile sensor design, dataset creation, and modeling techniques.
- Limited exploration of texture recognition on uneven surfaces and the use of multi-scale learning.
- Few implementations of ensemble methods for combining predictions across temporal resolutions.

The present study addresses these gaps by:

- Extending tactile datasets to include diverse and complex surface geometries.
- Introducing a multi-scale voting-based ensemble that surpasses fixed-window classifiers in accuracy.
- Combining transformer-based and CNN architectures to capture both local and global temporal features.

This work lies at the intersection of tactile sensing innovation, advanced time-series modeling, and ensemble classification, aiming to overcome key limitations in current tactile perception systems.

Chapter 3

Methodology

This chapter outlines the methodology used to collect, process, and analyze tactile data for texture classification on uneven surfaces.

3.1 Robotic Setup and Tactile Sensor

This section presents the hardware configuration used for probing textured surfaces and collecting tactile data. It comprises a robotic manipulator for executing contact motions and a multimodal sensor mounted on its end-effector.

3.1.1 Robotic Arm

The UFactory Lite6, a 6-degree-of-freedom (6-DOF) collaborative robotic arm, was used for executing surface probing and contact trajectories. The 6-DOF configuration provides full spatial positioning and orientation control of the end-effector, which is critical for maintaining a consistent contact normal when scanning surfaces with varying curvature. The arm offers a repeatability of ± 0.1 mm, a maximum payload of 600 g, and a reach of 440 mm, delivering both dexterity and precision for the controlled tactile exploration required in this study. It is fully programmable through ROS 2,

enabling precise motion control along Cartesian paths and seamless integration with sensing and data acquisition modules.

During experimentation, the robot executed structured movement sequences to scan or probe textured regions across flat, convex, concave, and mixed geometries. The contact direction was maintained predominantly orthogonal to the surfaces, achieved through calibration and the use of computed surface normals during path planning.

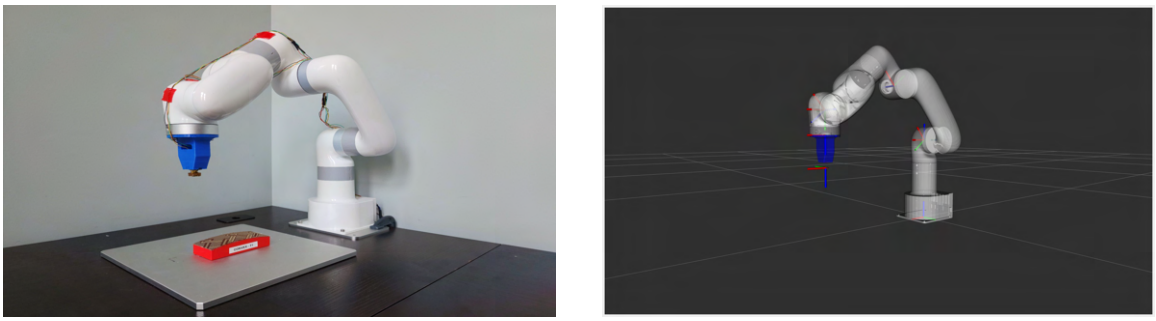

Figure 3.1: Experimental setup showing the UFactory Lite6 robotic arm with the tactile sensor mounted on the end-effector from two different angles.

3.1.2 Tactile Sensing Module

The tactile sensing module used in this work is the BioIn-Tacto sensor [1], a compliant and flexible multi-modal tactile sensing device designed for robotic applications. It combines a deformable contact surface with two sensing elements:

- MARG (Magnetic, Angular Rate, and Gravity) System: A 9-axis IMU that provides 3-axis acceleration, 3-axis gyroscopic angular velocity, and 3-axis orientation (quaternion).
- Barometric Pressure Sensor: A sealed pressure sensor housed within a soft dome, used to capture force-related pressure changes during contact.

This configuration offers a unique package of capabilities: a compliant and flexible sensing interface, high-fidelity orientation measurement, and direct pressure sensing from barometric data. The fusion of barometric pressure with IMU-derived orientation enables accurate, real-time computation of local surface normals during contact. Such combined functionality is particularly advantageous for exploring uneven surfaces while maintaining controlled approach angles and contact conditions.

In contrast, existing tactile sensing technologies—such as capacitive arrays, resistive sheets, optical tactile sensors, and vision-based tactile systems—can provide detailed pressure or deformation patterns, but none currently offer this exact combination of compliance, orientation sensing, and direct barometric pressure measurement in a single integrated module. These alternative sensors typically require additional estimation algorithms or external sensing modalities to approximate surface normals, introducing extra complexity and potential sources of error. All sensor streams are recorded through ROS 2 topics at an average rate of approximately 145 Hz.

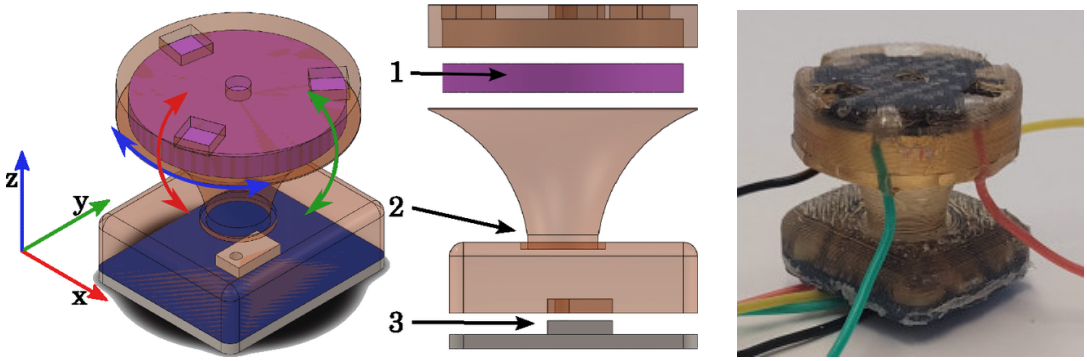

Figure 3.2: The **BioIn-Tacto** tactile sensing module comprises three primary components: (1) a MARG system, (2) a compliant structure, and (3) a barometer [1]. The entire module is encapsulated within a flexible structure [2].

3.1.3 Sensor Mounting and Configuration

The sensor is mounted using a rigid coupler aligned with the robot's end-effector tool frame. The orientation ensures that the sensing surface remains perpendicular to the contact direction. This configuration simplifies robot-sensor calibration and enables consistent probing geometry across all experiments.

Sensor wiring is secured along the robotic links to prevent disturbances, and data acquisition is fully synchronized within the ROS 2 data pipeline.

3.2 Surface Reconstruction

To facilitate consistent tactile data collection across complex and curved surfaces, a surface reconstruction phase is conducted prior to texture probing. This step enables the robot to build a sparse but structured map of the surface geometry using contact-based probing, which is later used to guide the data collection process in a consistent and controlled manner.

3.2.1 Contact Point Generation

For each surface-texture combination, the robot executes a predefined probing trajectory arranged in a 2D grid across the surface. At each probing location, the end-effector descends vertically until contact is detected based on a significant drop in the barometric pressure signal. The robot then records the 3D pose of the end-effector at the time of contact as a discrete surface point.

As shown in Figure 3.3(a), the robot approaches the surface with its end-effector aligned vertically. This consistent approach direction simplifies contact geometry and ensures the resulting contact point reflects a meaningful interaction between the

sensor and the surface.

3.2.2 Surface Normal Estimation

Once a sufficient number of contact points are collected, local surface normals are estimated through planar fitting. For each point, a neighborhood is selected, and a best-fit plane is computed using least-squares regression. The normal vector to this plane is stored as the estimated normal for that contact point.

As illustrated in Figure 3.3(b), when the sensor contacts the surface, slight deformation occurs in the compliant structure. This deformation confirms the sensor has made physical contact, and the associated normal is used to orient the sensor appropriately in the texture probing phase.

This strategy enables adaptive motion planning where the robot can align the sensing surface orthogonally to the local geometry, improving consistency and tactile signal quality.

3.2.3 Reconstruction Workflow

The reconstruction procedure is deterministic and repeatable. Each probing motion is executed at a fixed vertical velocity, with real-time monitoring of pressure data for contact detection. Once contact is detected, the robot holds briefly, retracts upward, and moves to the next planned point in the grid.

Figure 3.3(c) shows this full probing sequence in operation. By following this approach, a structured and repeatable point cloud with associated normals is obtained, which is then used to generate the desired poses for tactile data collection.

The computed surface normals, as illustrated in Figure 3.4, play a crucial role

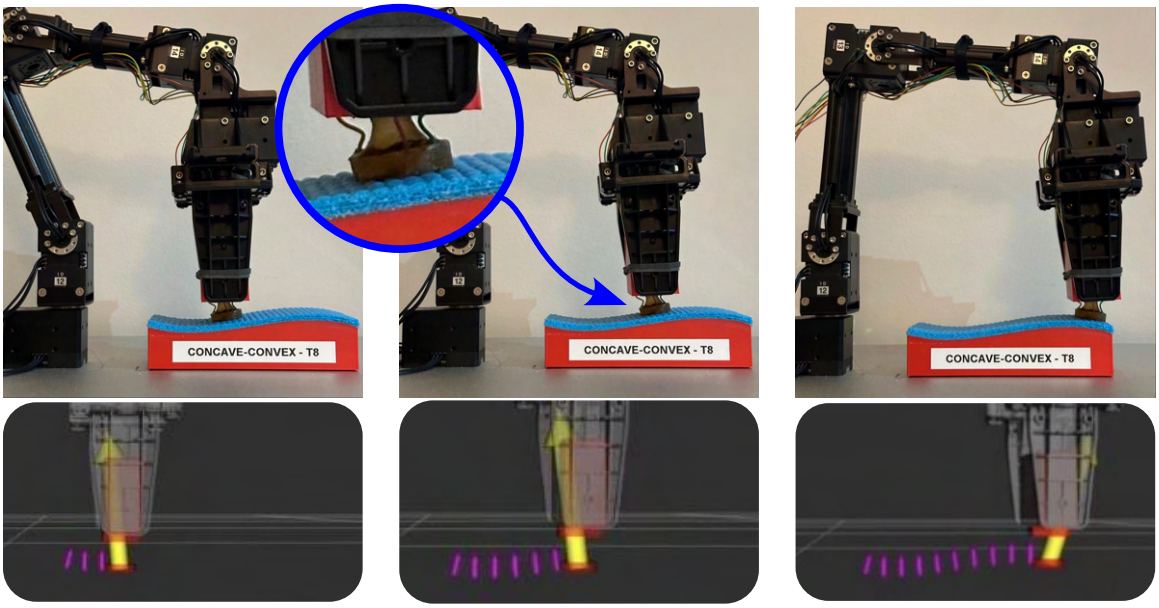

Figure 3.3: a–c) illustrate a robot equipped with a tactile sensing module probing a surface while maintaining the end-effector's approach angle normal to the horizontal plane. (a) shows the initial descent; (b) highlights the slight deformation upon contact; (c) depicts the sequence of probing movements over the surface.

in ensuring that the tactile sensing module maintains the desired contact orientation during data collection. By aligning the end-effector with the extracted normals, the robot can consistently achieve orthogonal contact across varied surface geometries, improving repeatability and signal quality. This information is also used in the trajectory planning stage to generate poses for subsequent tactile data acquisition.

3.3 Data Collection Procedure

After surface reconstruction, the robot proceeds to collect tactile data for texture classification. The objective is to capture consistent and high-resolution multimodal signals that reflect the physical interaction between the compliant sensor and a variety

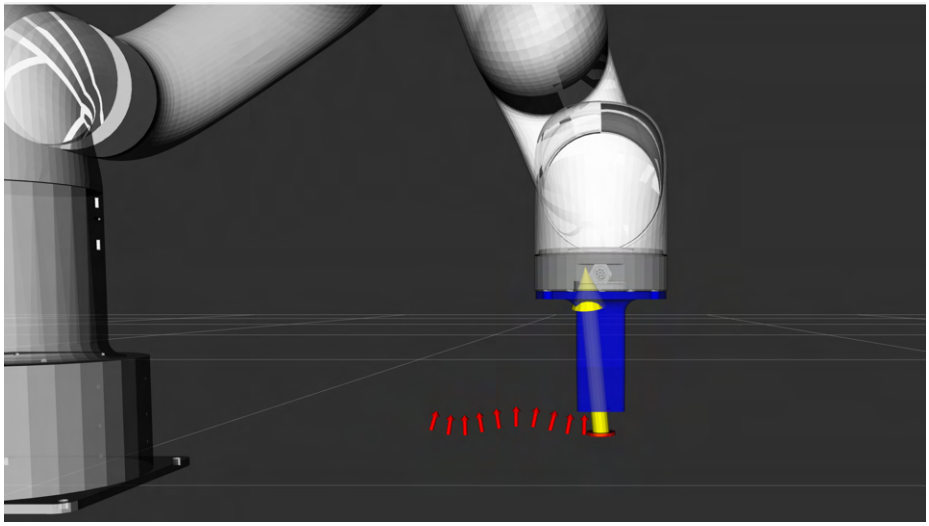

Figure 3.4: Visualization of surface normals extracted by the UFactory Lite6 during the reconstruction phase. Each arrow represents the computed local surface normal at a probing point, derived from the reconstructed contact geometry.

of textured surfaces.

3.3.1 Surface and Texture Set

The dataset includes six distinct base surfaces: five with varying degrees of curvature and one flat control surface. Each surface is covered with 12 different textures, including a range of materials such as fabrics (e.g., brocade, mesh cotton), leathers, polymers, and rigid textures like embossed plastics.

These textures were chosen to ensure diversity in key tactile properties including stiffness, roughness, pattern regularity, and friction. The aim was to generate a dataset representative of real-world variability encountered in robotic tactile sensing.

3.3.2 Contact Strategy and Probing Motions

For each surface-texture pair, the robot performs 25 separate exploration sequences. Each sequence involves a horizontal trajectory subdivided into multiple vertical probing contacts. At each probing point, the robot:

1. Moves to a predefined hover pose above the surface.
2. Descends vertically until a contact is detected via pressure signal drop.
3. Applies a small over-travel to ensure sensor deformation.
4. Retracts upward and moves to the next contact location.

This probing pattern is guided by the reconstructed contact points and surface normals obtained during the reconstruction phase. The consistency of approach angle and contact depth improves repeatability across textures and surfaces.

Figure 3.5 shows the robotic system in different phases of a typical exploration path—beginning, midpoint, and end—demonstrating the systematic motion strategy over uneven terrain.

3.3.3 Recorded Data Streams and Synchronization

During each probing contact, two synchronized data streams are recorded via ROS 2:

- Barometric Pressure: Measures internal air pressure variation inside the sensor's soft chamber, reflecting the force applied during surface contact.
- IMU Signals: Includes 3-axis accelerometer, 3-axis gyroscope, and 3-axis orientation (quaternion), enabling the system to capture the dynamic response of the sensor during contact and release.

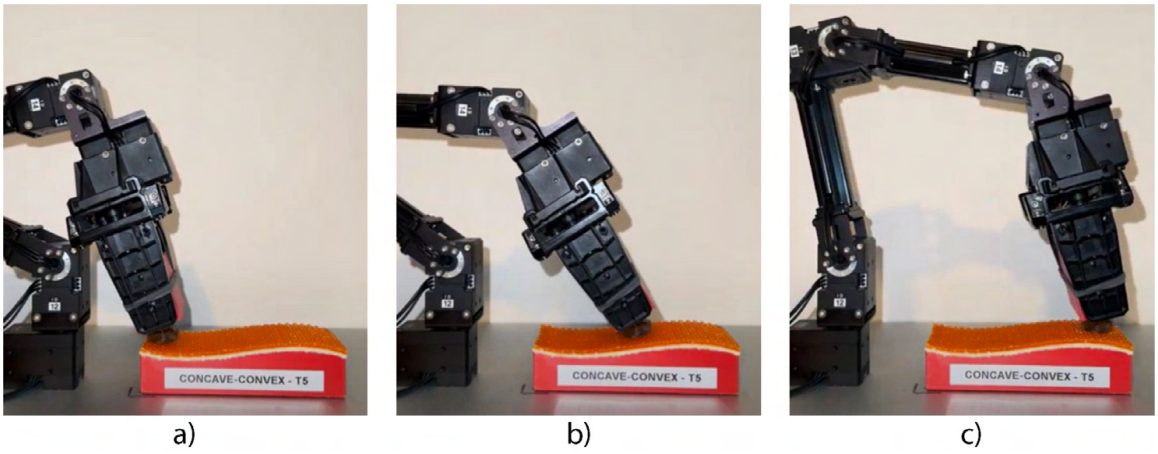

Figure 3.5: Overview of the OpenManipulatorX robotic arm with the sensing module mounted, collecting data on an uneven surface with a textured material: (a) the robot at the beginning of the exploration path; (b) the robot at the midpoint of the path; (c) the robot at the end of the exploration path.

All signals are recorded at approximately 145 Hz and are timestamped for synchronization. Each trial produces a time-series recording of multi-modal sensor data, stored in a structured directory hierarchy organized by surface, texture, and trial index. These recordings form the raw input for downstream preprocessing and model training.

3.4 Data Preprocessing Pipeline

Raw sensor data collected during tactile probing is preprocessed through a structured multi-step pipeline to prepare it for classification. The goal is to extract clean, normalized, and properly segmented time-series windows suitable for training deep learning models. Each step in the pipeline is implemented using modular Python scripts and applied consistently across all experimental trials.

3.4.1 Trimming and Cropping

Due to slight delays in robot motion or pauses before contact, the beginning and end of each recorded time series often contain idle data with no meaningful contact. These segments are trimmed using a threshold-based method applied to the barometric pressure signal. A JSON file stores the detected start and end indices for each trial, ensuring consistent and reproducible cropping across the dataset.

3.4.2 Sensor Data Merging

Barometric and IMU data are initially recorded as separate ROS 2 topics. These streams are synchronized and merged into a unified CSV format based on timestamps. The merged data includes all nine IMU features (acceleration, gyroscope, and orientation) and the barometric pressure value, resulting in a 10-dimensional time series per sample.

To avoid introducing artificial correlations, no forward or backward filling is applied; only timestamps common to both streams are preserved.

3.4.3 Normalization

After merging, each sensor signal is normalized individually using min-max scaling. The minimum and maximum values are computed per channel across all training samples and stored for consistent normalization of validation and test data. This scaling ensures balanced feature ranges and improves model convergence during training.

Normalization is applied only after trimming and merging to ensure clean and aligned data.

Figure 3.6 shows an example of a normalized IMU time series segment after trimming and alignment. Figure 3.7 presents a normalized barometric pressure signal from a typical probing interaction.

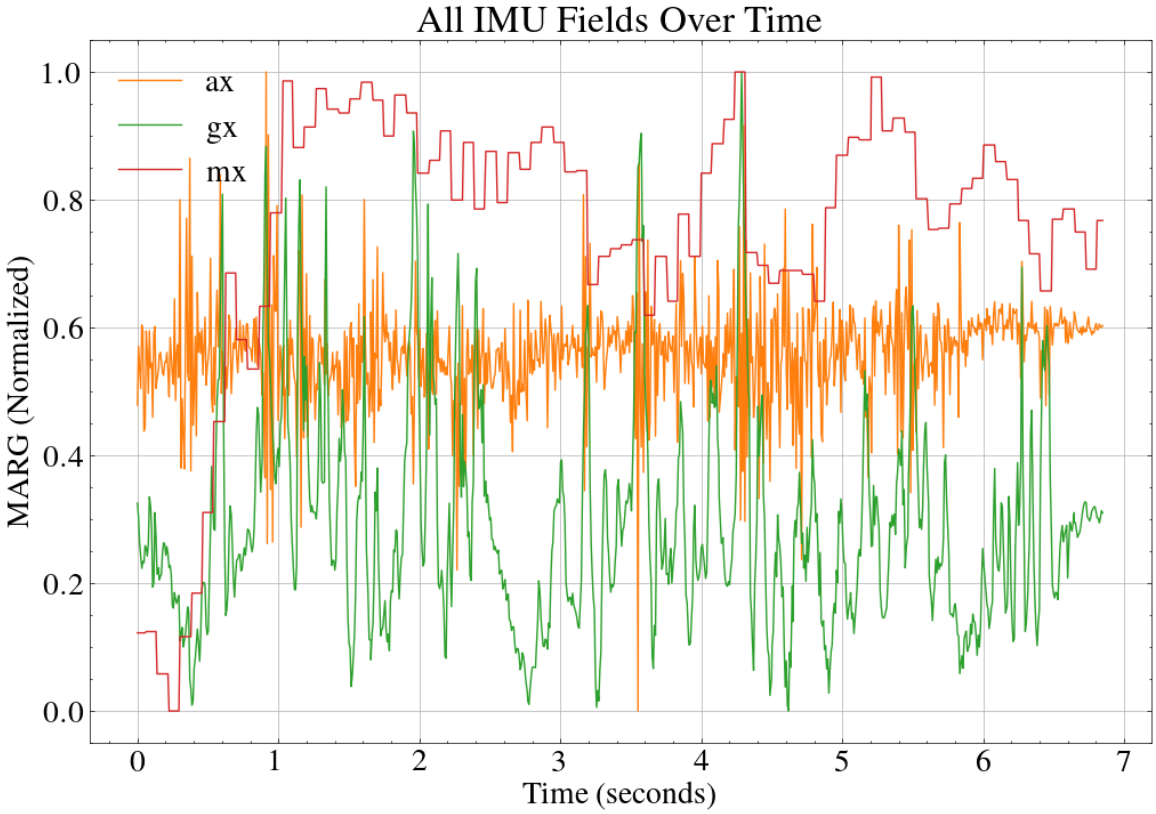

Figure 3.6: Preprocessed IMU data after trimming, merging, and normalization. The plot shows acceleration, gyroscope, and orientation (quaternion) values across time.

3.4.4 Windowing and Labeling

To enable training of temporal models, the continuous time-series data is segmented into fixed-length overlapping windows. Three different window sizes are used in this work: 128, 256, and 512 time steps. Each window is saved as an independent training example, with its corresponding texture label inherited from the trial folder.

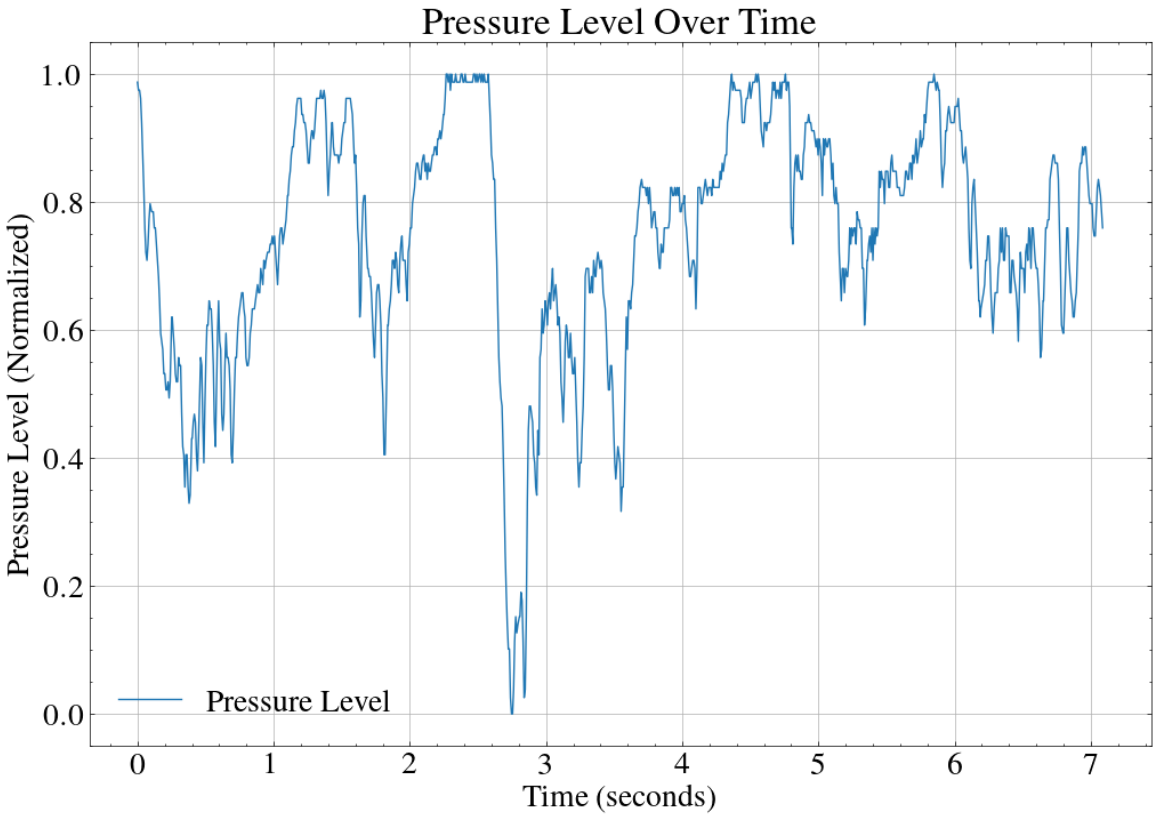

Figure 3.7: Preprocessed barometric pressure data from a single trial. The signal reflects surface contact and deformation dynamics after normalization.

Overlapping windows (e.g., with 50% overlap) increase dataset size and improve model robustness to minor temporal shifts. Each window is saved as a separate '.csv' file and then converted to '.npy' format for faster loading during model training.

3.4.5 Directory Organization

The entire preprocessing pipeline maintains a consistent directory structure. For each surface and texture, preprocessed files are saved under folders named by texture ID, window size, and processing stage (e.g., '4-Normalized', '5-Windowed', '6-Reindexed', '7-NPY'). This hierarchy enables efficient access to data subsets for experiments and

reproducibility across runs.

3.5 Classification Models

To classify tactile textures from preprocessed sensor signals, five deep learning models are designed and trained. Each model operates on a specific window size (128, 256, or 512 samples) and takes as input a 10-dimensional multivariate time series. The models fall into two main categories: baseline architectures and self-attention-based architectures.

3.5.1 Baseline Architectures

1D CNN model: Convolutional layers are effective at capturing local temporal patterns in sequential tactile data by applying learnable filters over short time spans. These filters act as feature extractors, identifying repetitive structures such as vibration signatures, micro-texture patterns, or transient force variations. Pooling layers reduce temporal resolution while retaining dominant features, and batch normalization improves training stability. The use of ReLU activation ensures non-linear feature learning, and fully connected layers at the end combine extracted patterns for final classification. Figure 3.8 shows the architecture of the 1D CNN model.

BiLSTM model: Long Short-Term Memory (LSTM) units are designed to capture temporal dependencies by selectively remembering and forgetting past information. Using a bidirectional configuration allows the model to learn from both past and future contexts within each window, which is important in tactile sequences where motion direction or contact history influences the signal. This is especially useful for capturing the progression of texture-induced patterns over time. The BiLSTM

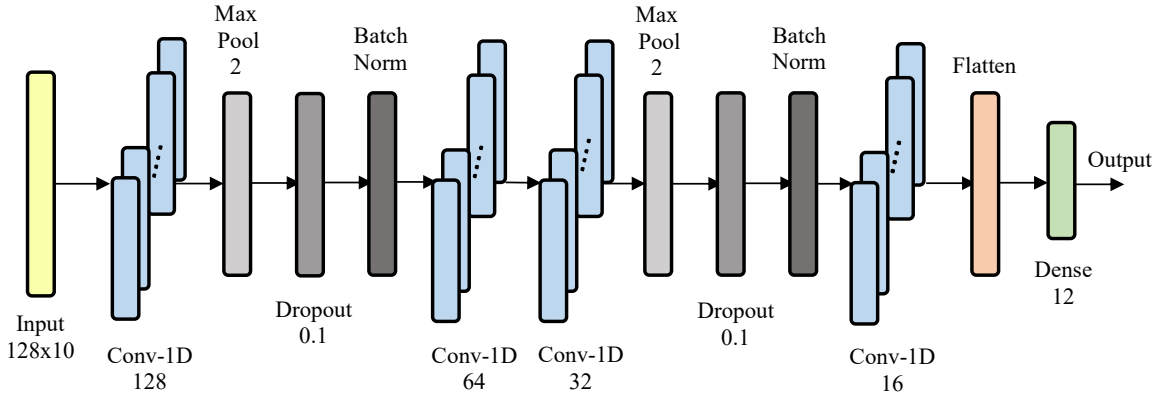

Figure 3.8: Architecture of the 1D CNN model used for tactile texture classification.

architecture is shown in Figure 3.9.

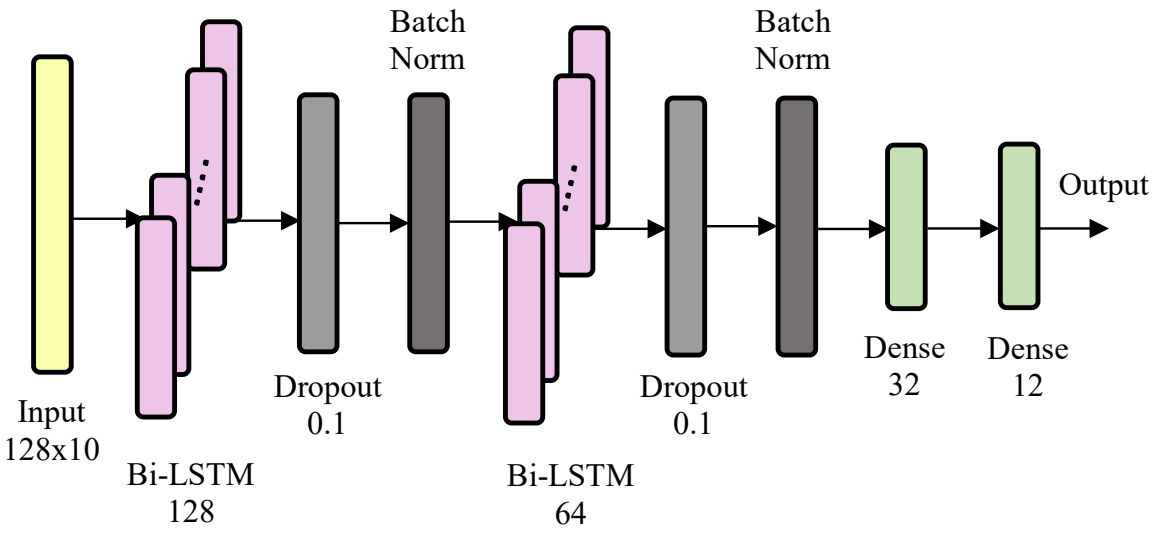

Figure 3.9: Architecture of the BiLSTM model, which models temporal dependencies bidirectionally.

Hybrid 1D CNN–BiLSTM model: This architecture leverages the strengths of both components: convolutional layers first extract short-range temporal features from raw tactile data, reducing noise and emphasizing local patterns, while BiLSTM layers model the longer-range temporal structure that emerges over a full contact sequence. This combination allows the network to learn both fine-grained textural cues

and global sequence-level relationships, which is especially valuable for uneven surfaces where local and global dynamics interact. The hybrid architecture is illustrated in Figure 3.10.

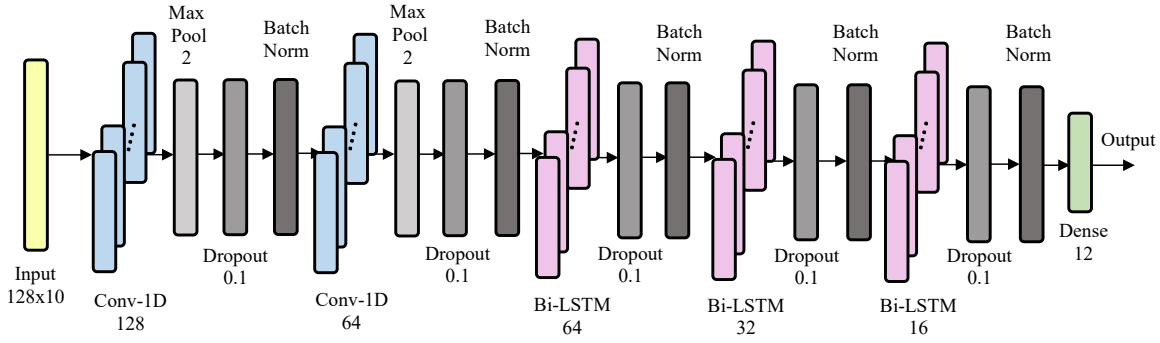

Figure 3.10: Architecture of the hybrid model combining 1D CNN and BiLSTM layers.

3.5.2 Self-Attention-Based Architectures

Self-attention mechanism: Self-attention enables the model to learn non-local relationships by allowing each time step in the tactile sequence to selectively attend to other relevant time steps. This is particularly useful when texture characteristics manifest over long temporal spans or when uneven surfaces introduce irregular spacing of distinctive events. Unlike recurrence, attention mechanisms compute dependencies in parallel, improving efficiency and allowing flexible context modeling.

Hybrid 1D CNN–self-attention model: This approach combines convolutional layers for initial local feature extraction with multi-head self-attention layers for modeling long-range dependencies. The CNN front-end filters and condenses the raw signal into high-level feature maps, while the attention block identifies relationships between distant points in time, enabling the model to reason about distributed patterns that may occur in different parts of the contact sequence. This balance of local sensitivity

and global reasoning makes it effective for challenging, geometry-dependent tactile recognition tasks. The architecture is shown in Figure 3.11.

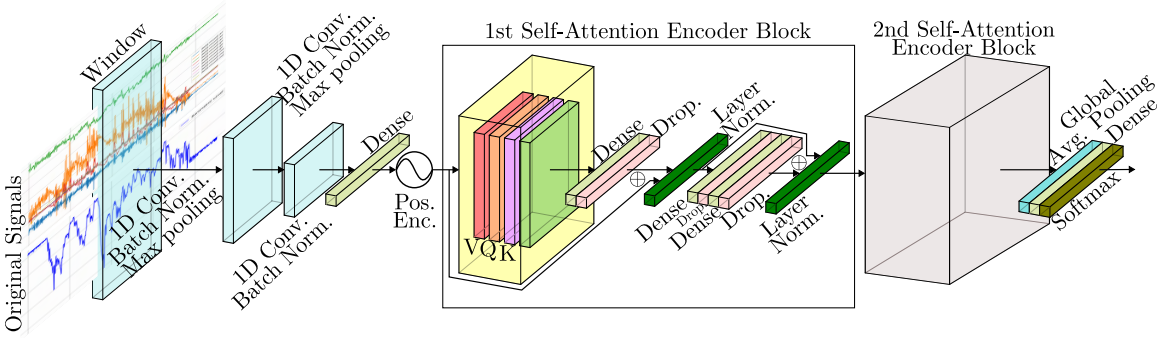

Figure 3.11: Architecture of the hybrid model combining 1D CNN and self-attention layers.

3.5.3 Training Setup

All models are trained using categorical cross-entropy loss with the Adam optimizer. The initial learning rate is set to 1e−3 and decayed based on validation performance. Training is conducted for up to 50 epochs with early stopping to prevent overfitting.

Batch sizes range from 32 to 64 depending on the model and input length. Instead of a fixed train/validation/test split, we use five-fold cross-validation to evaluate model performance. The dataset is divided into five folds, and each model is trained and validated five times using a different held-out fold for testing in each round. This strategy ensures a more reliable and unbiased estimate of performance across textures and surfaces. The implementation is done in TensorFlow, and all training is performed on GPU-accelerated systems. Detailed results and fold-wise evaluation metrics will be presented in the next chapter.

In total, these distinct model architectures are used: 1D CNN, BiLSTM, hybrid 1D CNN–BiLSTM, and hybrid 1D CNN–self-attention.

3.6 Voting-Based Ensemble System

3.6.1 Motivation for Model Combination

Each input window size (128, 256, or 512 samples) captures a different temporal resolution of the tactile signal. Shorter windows are more sensitive to transient contact features and local variations, while longer windows are better at modeling global temporal patterns and contact evolution over time.

Instead of choosing a single window size arbitrarily, the proposed approach combines the predictive power of all three. This allows the model to leverage fine-grained, mid-range, and long-term features simultaneously, increasing robustness to surface shape, contact duration, and signal shifts.

3.6.2 Soft Voting Mechanism

The ensemble system uses a weighted soft voting strategy to combine the predictions from three independently trained models. Each model is trained using the same architecture (for example, CNN or self-attention-based) but on a different window size (128, 256, or 512). During inference, the same input signal is segmented into all three window lengths and passed to its respective model.

Each model outputs a probability distribution over the 12 texture classes. These three vectors are then combined using a weighted average, where the weights determine the contribution of each window size to the final decision. The class with the highest combined probability is selected as the predicted label.

3.6.3 Weight Selection via Grid Search

The weights for the 128, 256, and 512 models are determined through a grid search procedure. Candidate weight combinations are evaluated using five-fold cross-validation on the training set. The combination that achieves the highest mean validation accuracy is selected for the final evaluation. This ensures that the voting system balances contributions from different temporal resolutions in a data-driven manner rather than using equal weights by default.

3.6.4 Comparison Strategy

To ensure fair evaluation, the voting-based ensemble is constructed separately for each architecture type. For example, three CNNs trained on 128, 256, and 512 samples form one ensemble, and the same holds for BiLSTM and attention-based models.

This allows for direct comparison between ensemble models and their individual single-window counterparts. Performance metrics such as classification accuracy are averaged over five-fold cross-validation, as will be detailed in the next chapter. This setup also helps assess how ensembling affects robustness across flat and uneven surfaces.

3.7 Dataset Description

To support the development and evaluation of tactile texture classification models, we collected a comprehensive dataset using the robotic system and tactile sensing module described earlier. The dataset captures sensor responses across a wide range of surface geometries and material textures, enabling controlled experiments on generalization, invariance, and classification performance.

3.7.1 Surface Types

The dataset includes six distinct surface configurations, each designed to introduce specific contact geometries.

The flat surface serves as a baseline with no curvature, providing a uniform contact plane ideal for isolating texture effects.

The concave surface presents a uniformly inward-curving geometry.

Figure 3.12: Concave surface configuration.

The convex surface features a uniformly outward curvature.

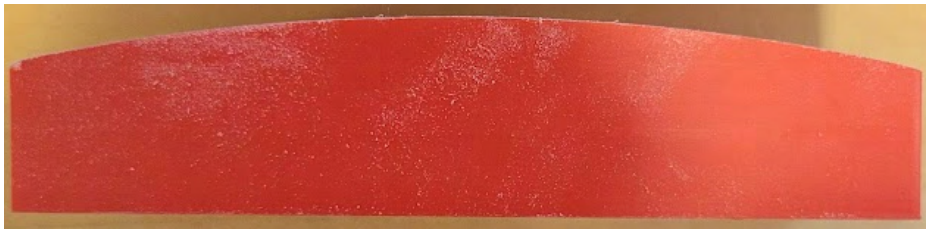

Figure 3.13: Convex surface configuration.

The concave-convex surface combines both inward and outward curvature in a continuous profile.

The concave-convex-concave surface has concave regions enclosing a central convex section, creating a more dynamic interaction profile.

The convex-concave-convex surface is the reverse, with convex outer sections surrounding a concave center.

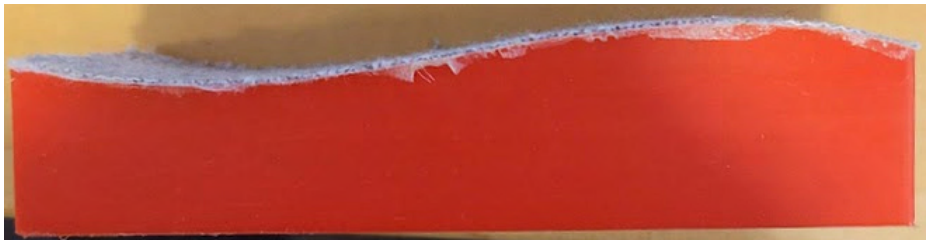

Figure 3.14: Concave-Convex composite surface.

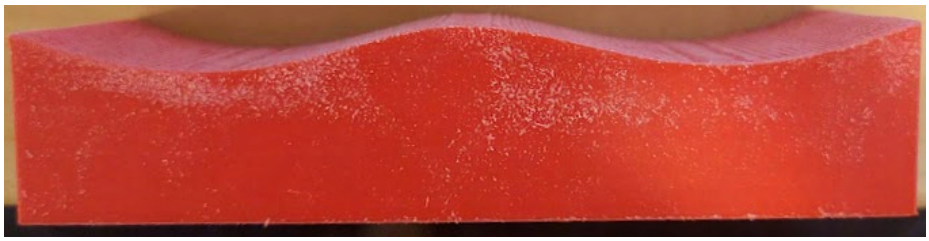

Figure 3.15: ConcaveConvexConcave surface pattern.

3.7.2 Texture Set

Each surface was overlaid with 12 distinct textures selected to span a broad range of material types, compliance, surface roughness, and pattern complexity. The textures include materials such as brocade fabric, honeycomb fabric, silicone mesh, wood, and several leather and polymer-based textures. This variety ensures meaningful signal variation across textures under different contact conditions.

Figure 3.18 provides a visual example of how a single texture is adapted across all six surface geometries in the dataset. This highlights the diversity of contact

Figure 3.16: ConvexConcaveConvex surface pattern.

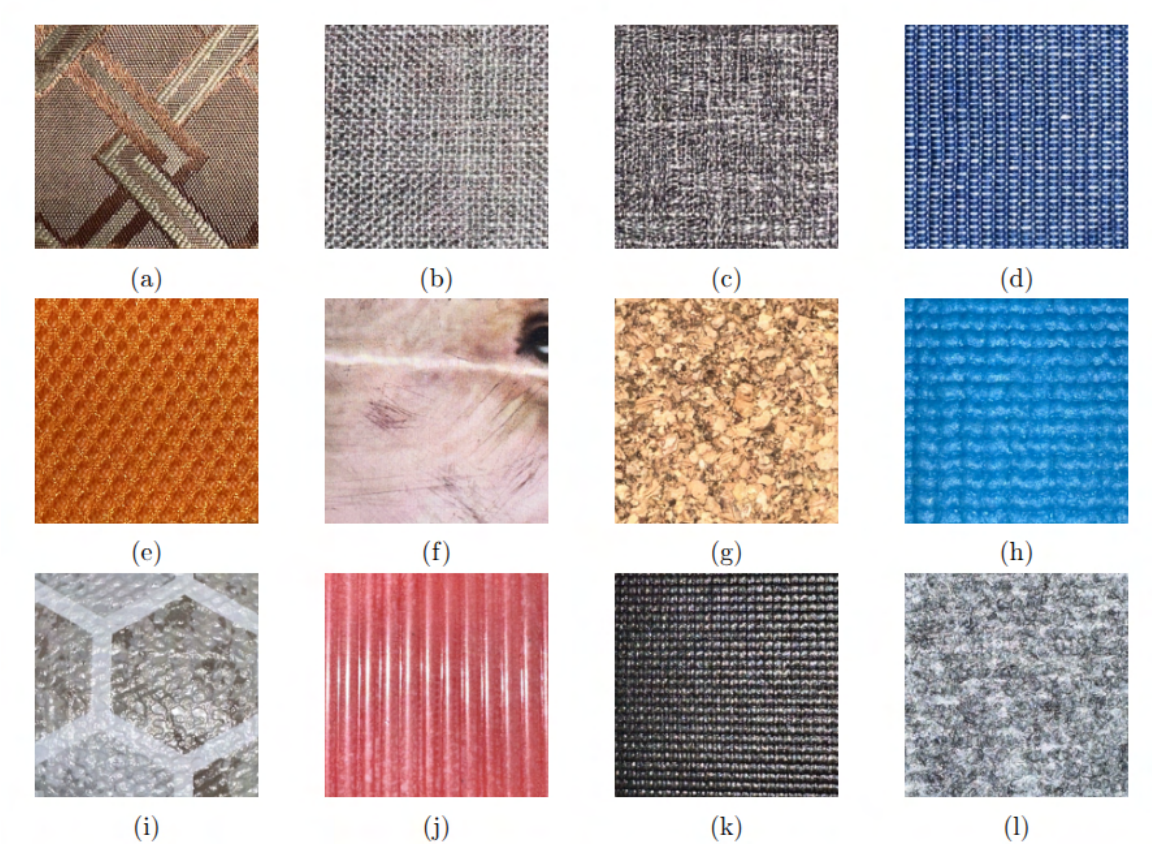

Figure 3.17: The 12 textures used in the experiments. a) brocade fabric; b) open weave cotton; c) tight weave cotton; d) mesh cotton; e) honeycomb fabric; f) embossed plastic; g) wood; h) silicone mesh; i) reptile-patterned leather; j) ridged polymer; k) mesh leather; l) carpet wool.

conditions encountered, even when the surface material remains the same. Such variation is essential for evaluating a model's ability to generalize across changes in surface shape while maintaining consistent texture recognition.

3.7.3 Dataset Variants

To assess different generalization scenarios, the dataset was organized into six experimental variants.

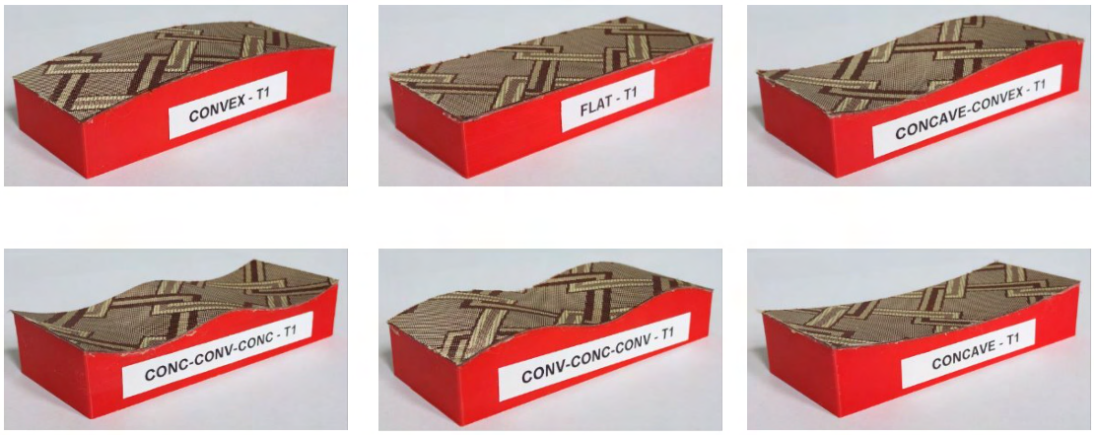

Figure 3.18: Illustration of a single example texture applied to all six surface geometries: flat, concave, convex, concave–convex, concave–convex–concave, and convex–concave–convex. This demonstrates how the same texture interacts with different contact geometries during data collection.

The Different Textures, Different Shapes (DTDS) configuration pairs each surface with a unique subset of textures. This setup simulates varied and unstructured real-world environments. For example, the concave surface was used with textures T1, T6, and T11; the convex surface with T2, T7, and T12; and so on.

The Different Textures, Same Shape (DTSS) variant applies all 12 textures to a single surface shape. This isolates the influence of texture under fixed geometric conditions.

The Same Textures, Different Shapes (STDS) variant uses a fixed subset of textures (T1 to T5) across all six surface types, enabling evaluation of shape generalization with consistent tactile content.

The Flat All Textures (FAT) variant applies all textures on the flat surface. This baseline configuration is particularly useful for comparing performance under minimal geometric variation.

The Flat and Concave/Convex All Textures (FCCAT) variant involves the flat,

Table 3.1: Surfaces and textures used in each dataset variant.

Dataset Variant	Surfaces Used	Textures Used
DTDS	All six surfaces	Different subset of 3 textures per surface
DTSS	One surface	All 12 textures
STDS	All six surfaces	Same subset of 5 textures
FAT	Flat surface	All 12 textures
FCCAT	Flat, concave, convex	All 12 textures
MS	All six surfaces	All 12 textures

concave, and convex surfaces each with all 12 textures. This provides an intermediate level of shape complexity.

The Mixed Scenarios (MS) variant aggregates all surface and texture combinations to create a comprehensive and challenging testbed that reflects realistic variability.

A summary of the surfaces and textures used in each dataset variant is provided in Table 3.1. This tabular view clarifies the design of each configuration and will be referenced in the results chapter.

Chapter 4

Results and Discussion

This chapter presents and analyzes the experimental results obtained from the five deep learning models described in Chapter 3. The evaluation is conducted on the tactile dataset using five-fold cross-validation, and performance is measured using classification accuracy, along with additional metrics and visualizations where relevant.

4.1 Evaluation Protocol

All model evaluations are based on five-fold cross-validation. For each model architecture and window size, the dataset is divided into five folds of equal size. In each iteration, one fold is used for testing, and the remaining four folds are used for training and validation. The process is repeated five times so that each fold serves as the test set once.

This procedure is applied consistently across all five model architectures and across the three temporal window sizes (128, 256, 512). For each combination, the average classification accuracy across the five folds is reported.

To ensure comparability, the same fold divisions are used across all models. The

same five-fold protocol is also applied when evaluating the voting-based ensemble systems. In this case, three independently trained models (one per window size) produce predictions, which are combined via soft voting as described in Section 3.6.

In subsequent sections, we present and compare results for single models, voting-based ensembles, and different surface types (flat vs. uneven), and we analyze confusion matrices to better understand model behavior.

4.2 Single Model Performance

To evaluate the classification capability of each model, we trained all five architectures described in Chapter 3 on six datasets, each corresponding to a different surface type. For each dataset-model combination, three different temporal window sizes (128, 256, and 512) were used, resulting in a total of 90 trained models. Every model was evaluated using five-fold cross-validation, and a confusion matrix was generated based on the predictions across all folds.

Each confusion matrix visualizes the distribution of predicted versus true texture labels, offering insight into common misclassifications, class separability, and overall performance.

In the subsections below, we present and analyze the results for each dataset individually. For every dataset, we show the confusion matrices of all five model architectures across three window sizes.

4.3 Single Model Performance Evaluation

This section presents the performance of all five model architectures across six datasets. Each model was trained and evaluated using 5-fold cross-validation and three different temporal window sizes: 128, 256, and 512. The evaluation is visualized using confusion matrices, which show classification accuracy and error distribution across the 12 texture classes.

We begin with the first dataset configuration: Different Textures, Different Shapes.

4.3.1 Results on Dataset: Different Textures, Different Shapes

This configuration evaluates each model's ability to distinguish textures when both the shape and texture vary across samples. The presence of geometric complexity makes this dataset one of the most challenging cases.

1D CNN The 1D CNN model performs moderately across window sizes, with noticeable improvement as the temporal window increases. At window size 128, several classes are misclassified due to insufficient temporal context. At 256 and especially at 512, some of these confusions reduce, indicating the importance of longer sequences for capturing distinctive features.

BiLSTM The BiLSTM model shows improved accuracy compared to 1D CNN, particularly at 256 and 512 window sizes. The recurrent structure helps model the sequence better, reducing confusion in several texture groups. However, performance is still affected by surface variation, and some minor misclassifications persist.

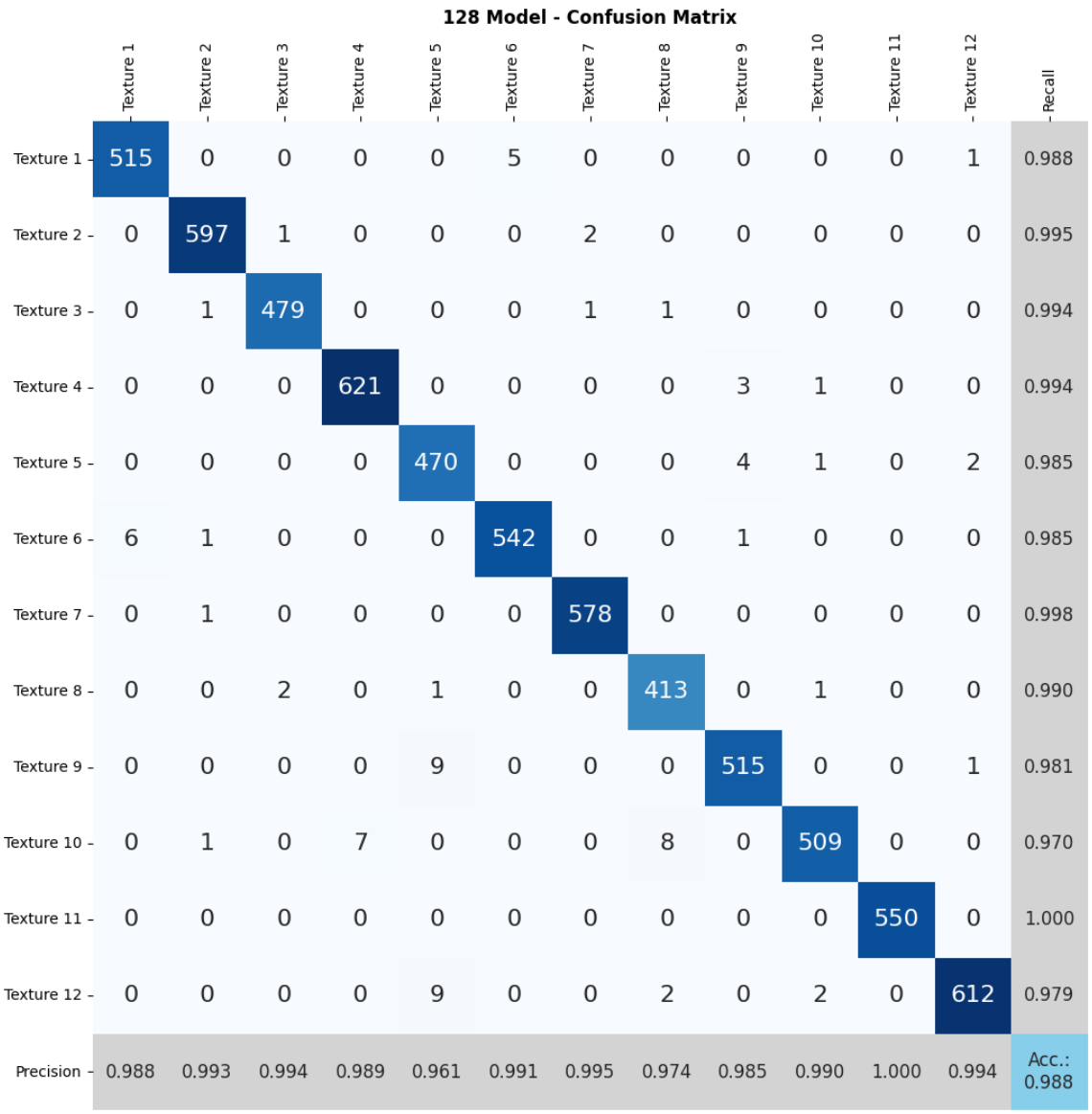

Figure 4.1: Confusion matrix for 1D CNN on Different Textures, Different Shapes with window size 128.

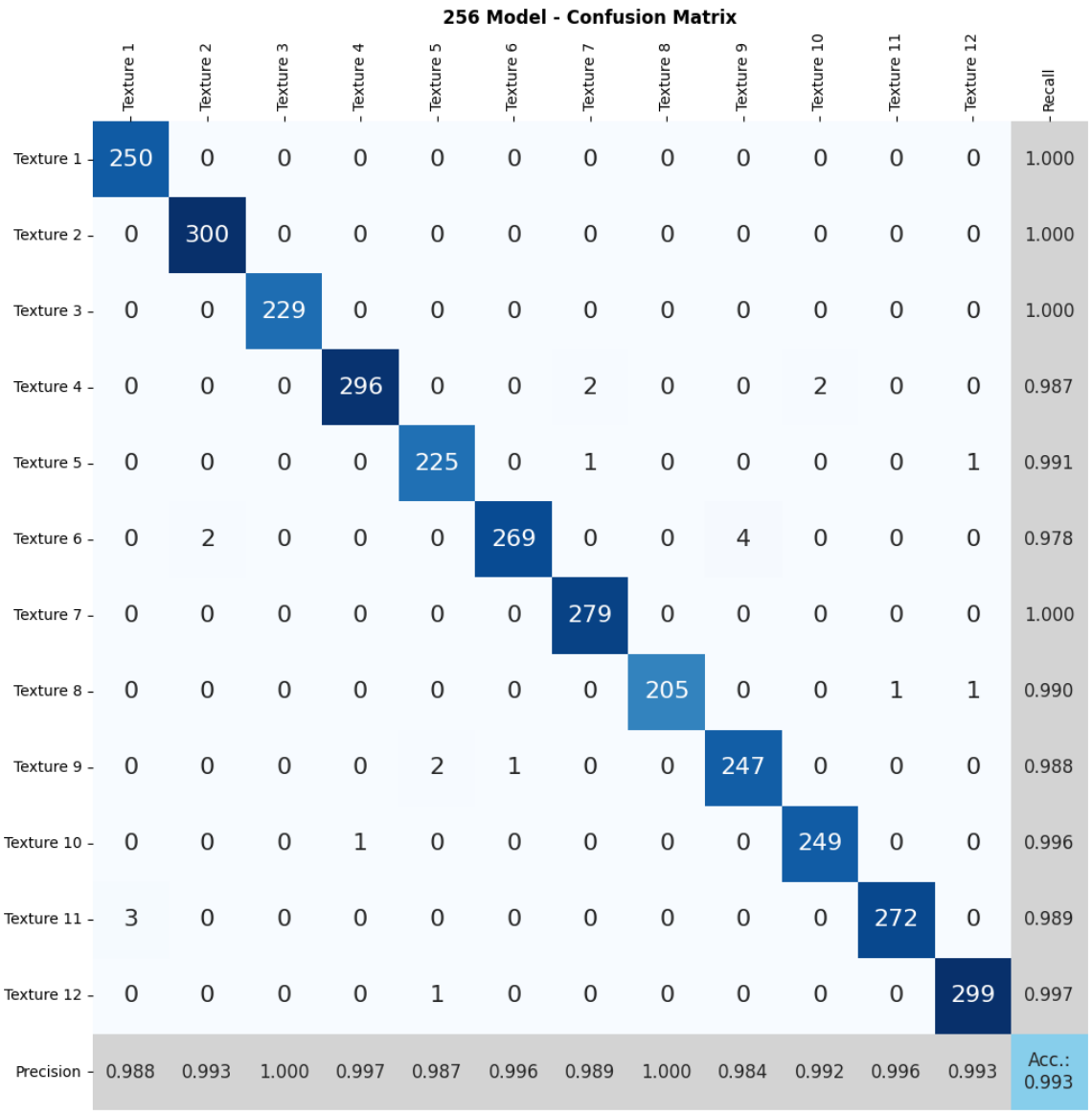

Figure 4.2: Confusion matrix for 1D CNN on Different Textures, Different Shapes with window size 256.

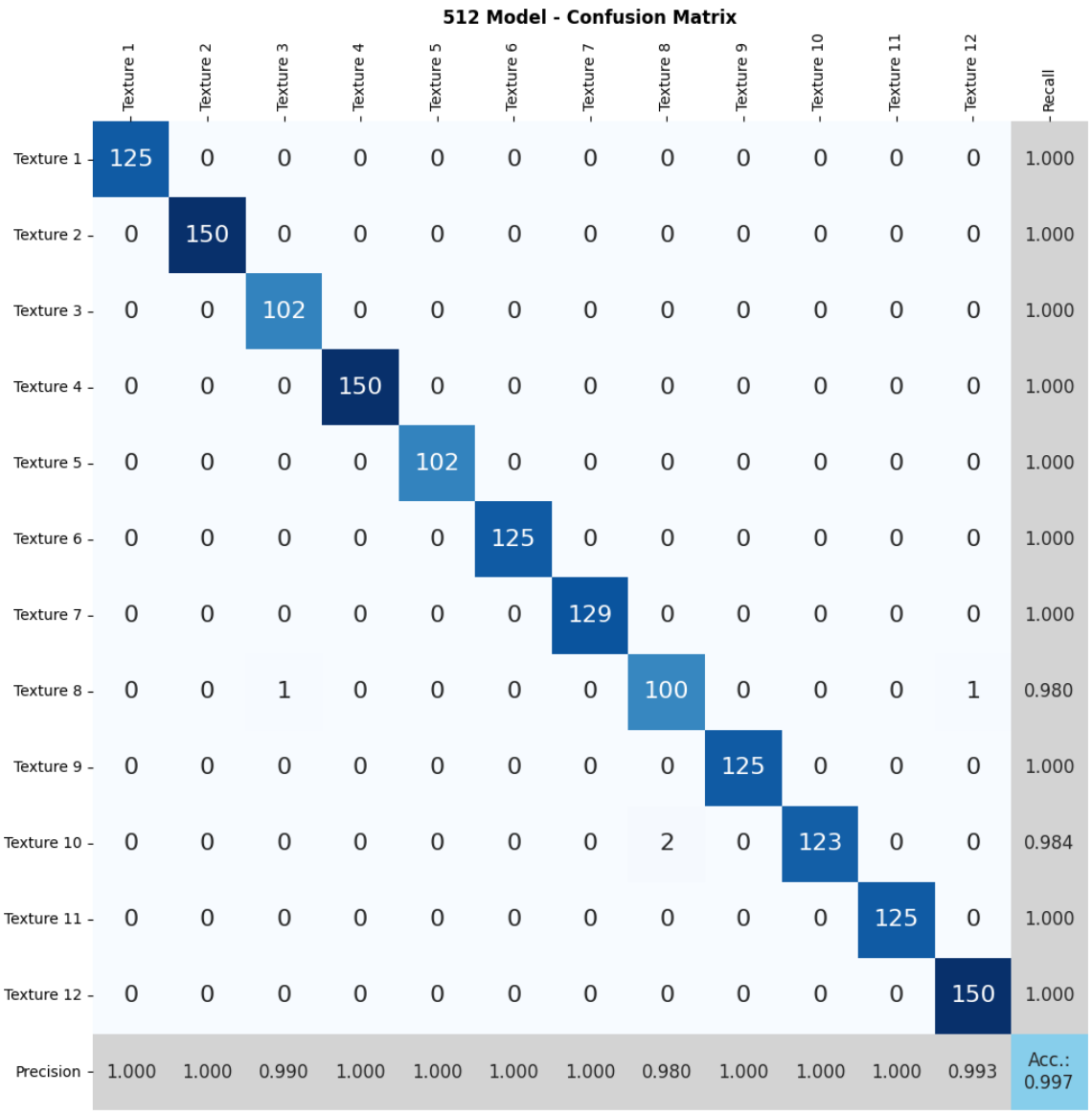

Figure 4.3: Confusion matrix for 1D CNN on Different Textures, Different Shapes with window size 512.

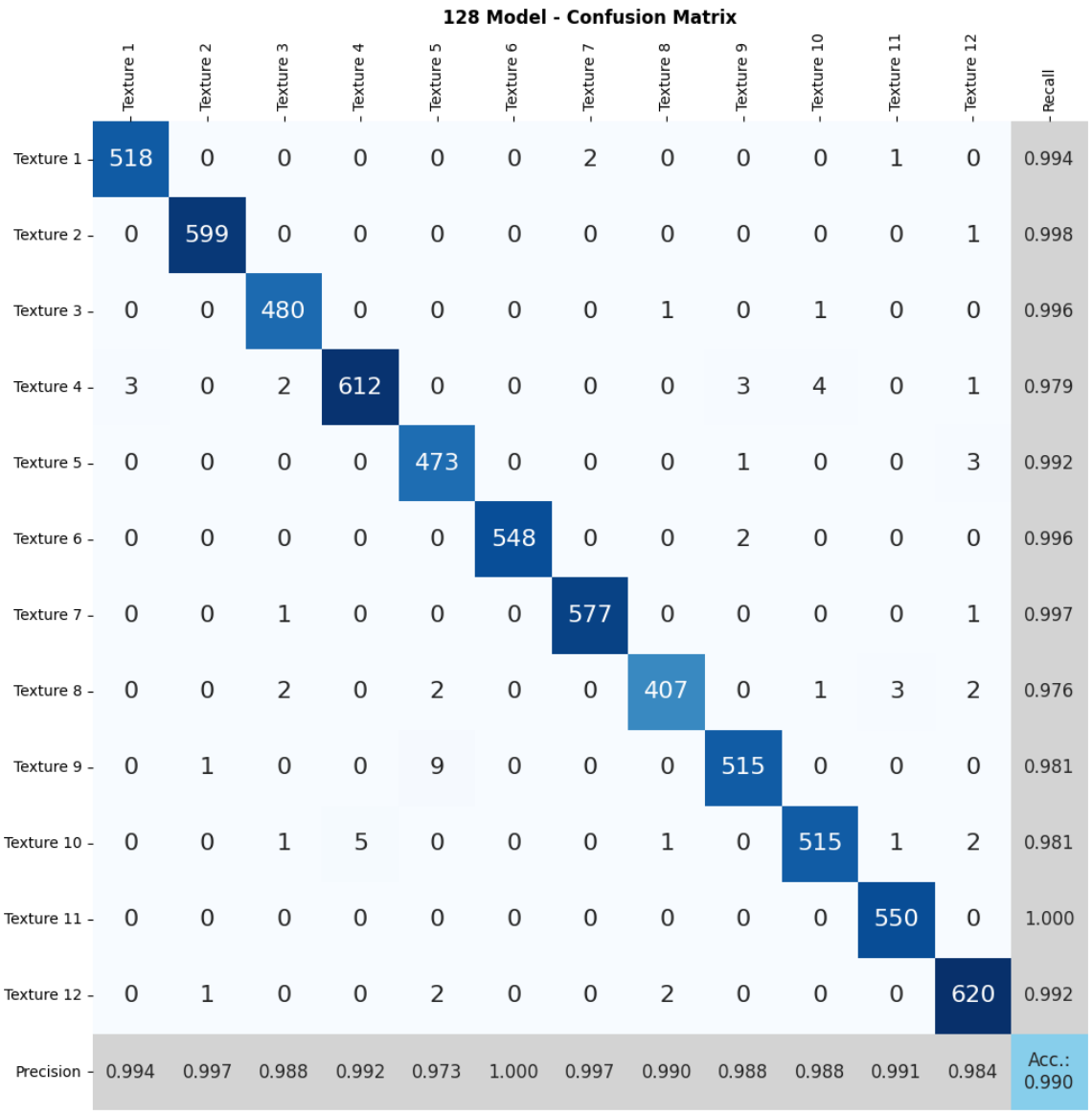

Figure 4.4: Confusion matrix for BiLSTM on Different Textures, Different Shapes with window size 128.

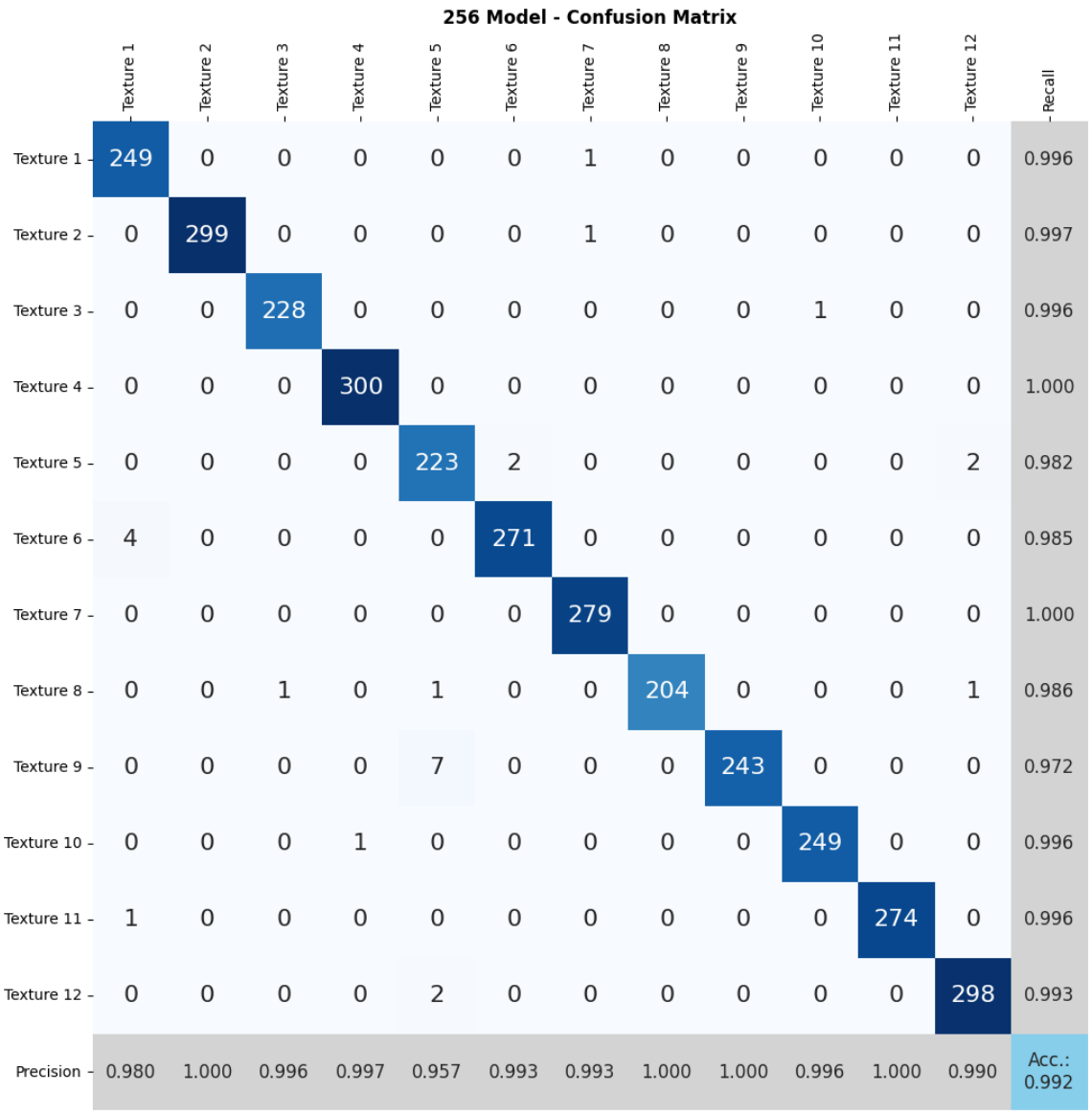

Figure 4.5: Confusion matrix for BiLSTM on Different Textures, Different Shapes with window size 256.

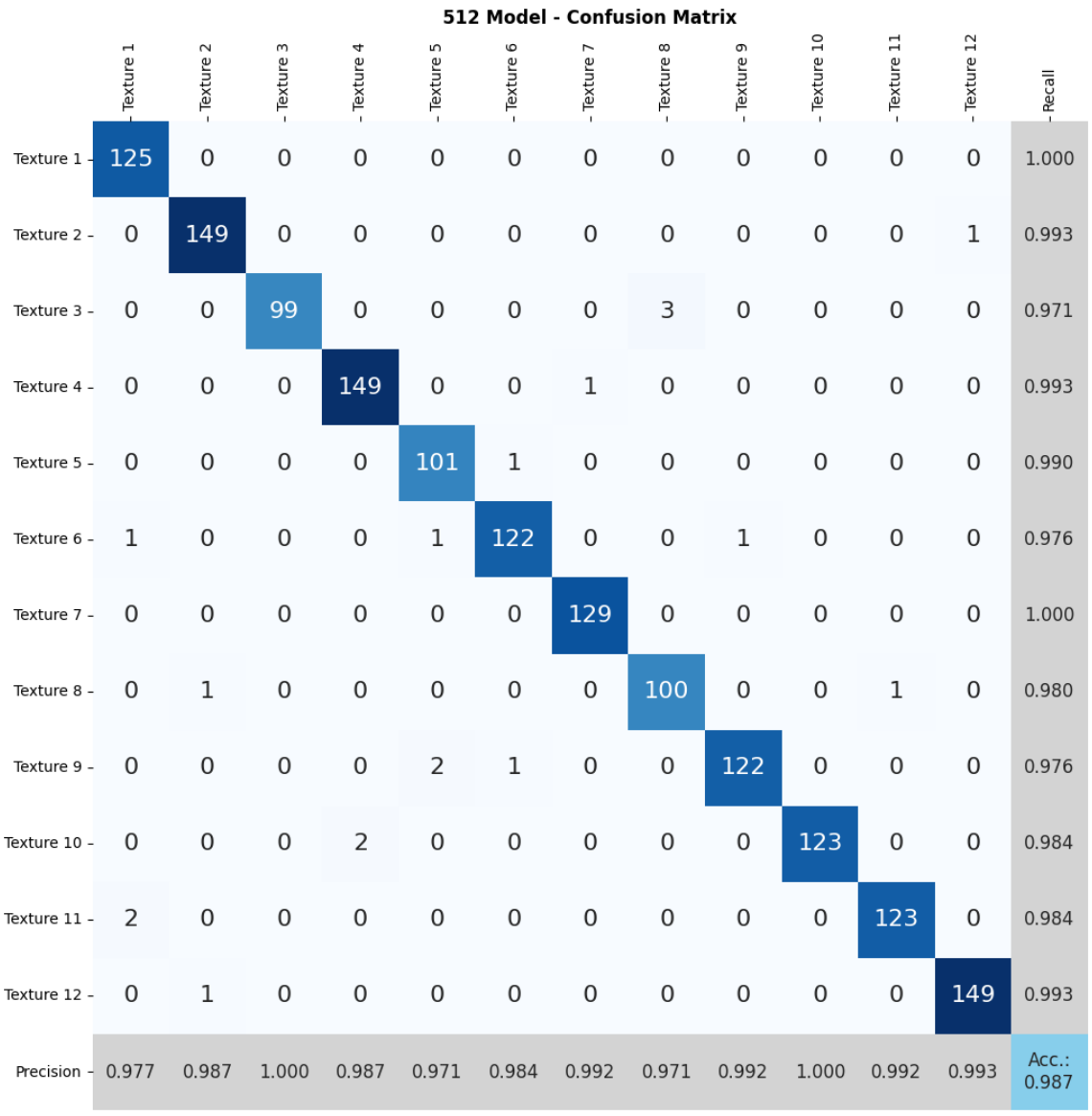

Figure 4.6: Confusion matrix for BiLSTM on Different Textures, Different Shapes with window size 512.

Hybrid CNN-BiLSTM Combining convolutional feature extraction with temporal modeling leads to stronger results. The hybrid model consistently outperforms both individual baselines. It leverages both local spatial features and sequential context, especially evident at window size 512, where several misclassifications are resolved.

Self-Attention-Based The self-attention-based model shows promising performance, particularly at window size 512. Attention mechanisms appear to better capture long-range dependencies across the time-series input. While the model still faces challenges at short windows like 128, its high resolution at larger sizes demonstrates its capacity to encode global tactile context.

Hybrid CNN + Self-Attention This model achieves the best performance on this dataset. Across all window sizes, it maintains high classification accuracy and low confusion. The synergy between convolutional and attention-based layers allows the model to robustly extract texture-relevant information even under shape variability.

In this dataset, models that combine both spatial and temporal processing capabilities—especially those using attention—outperform simpler baselines. Larger window sizes (particularly 512) consistently lead to better results across all architectures.

4.3.2 Other Dataset Results

In addition to the "Different Textures, Different Shapes" configuration, the same evaluation process was conducted on five other dataset variants: Different Textures, Same Shape; Same Textures, Different Shapes; Flat All Textures; Flat and Concave/Convex All Textures; and Mixed Scenarios.

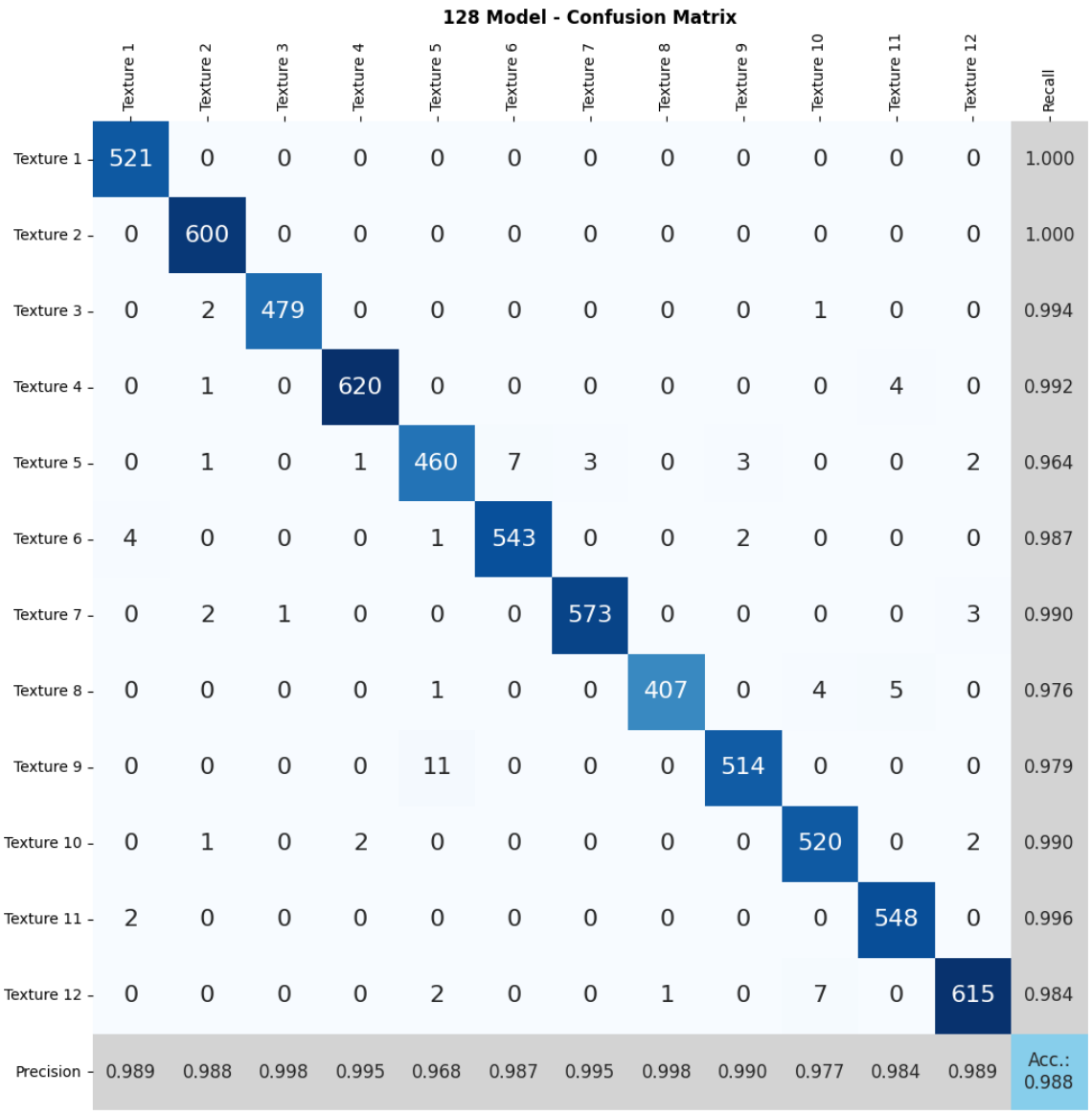

Figure 4.7: Confusion matrix for Hybrid CNN-BiLSTM on Different Textures, Different Shapes with window size 128.

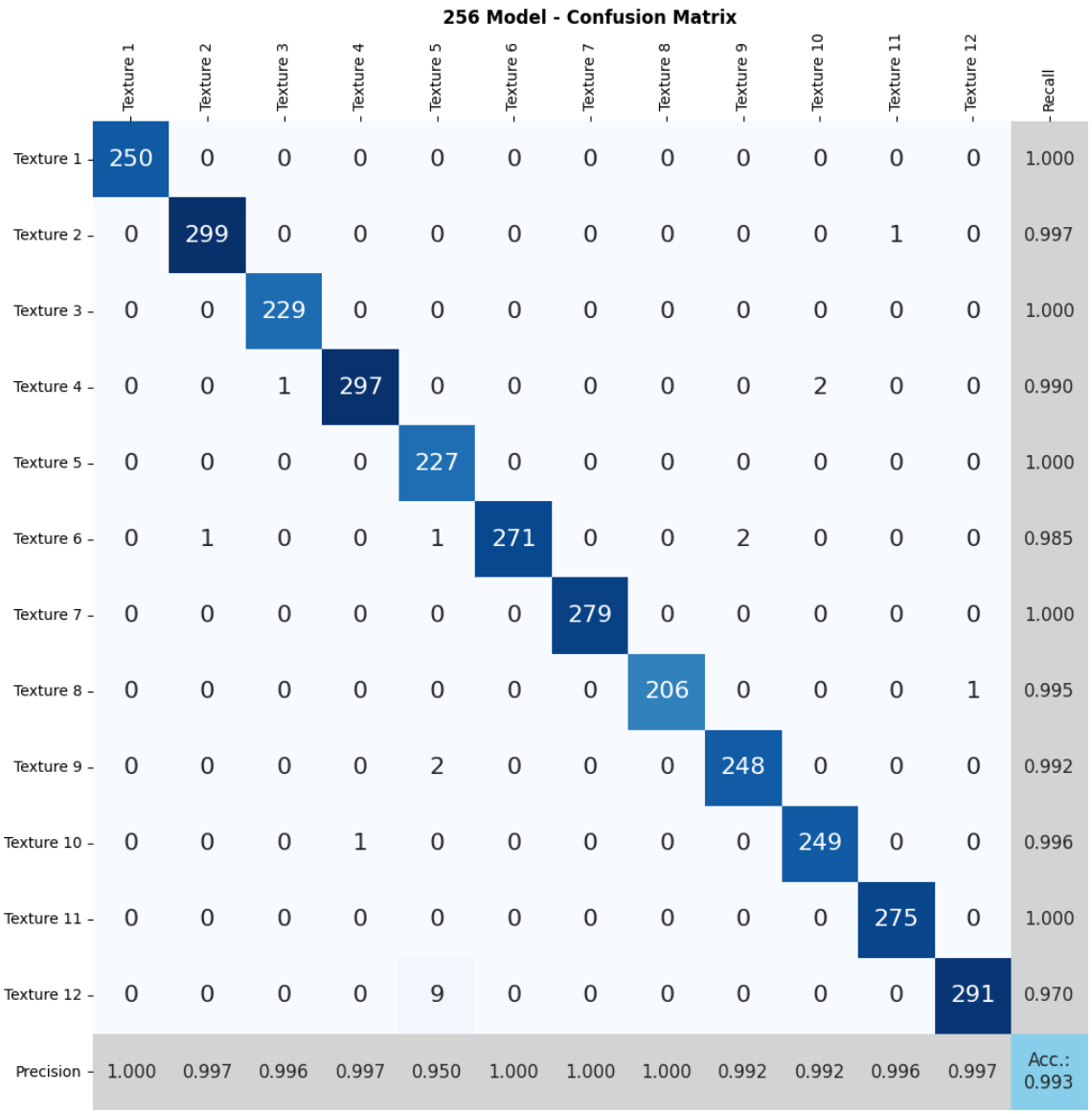

Figure 4.8: Confusion matrix for Hybrid CNN-BiLSTM on Different Textures, Different Shapes with window size 256.

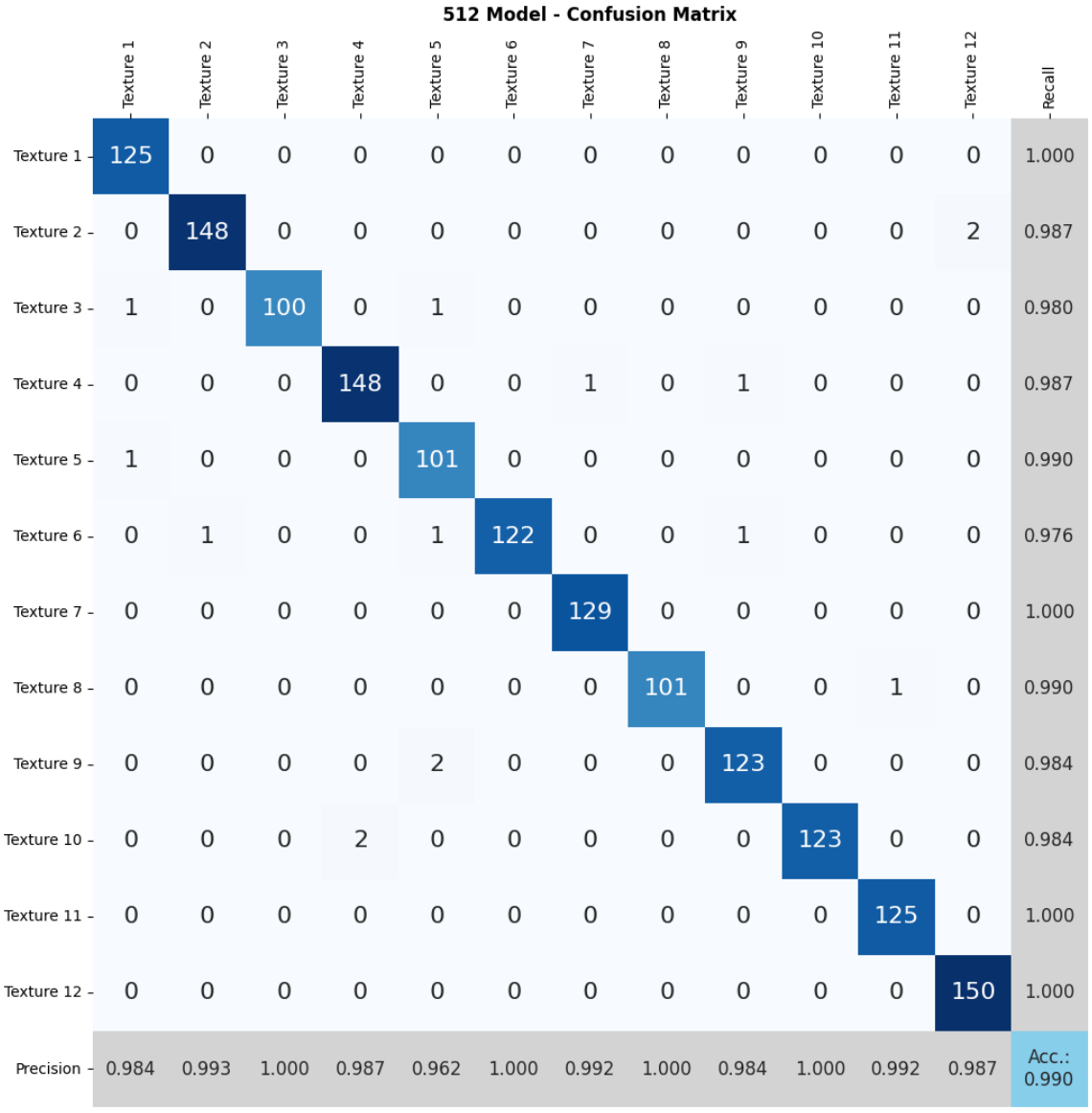

Figure 4.9: Confusion matrix for Hybrid CNN-BiLSTM on Different Textures, Different Shapes with window size 512.

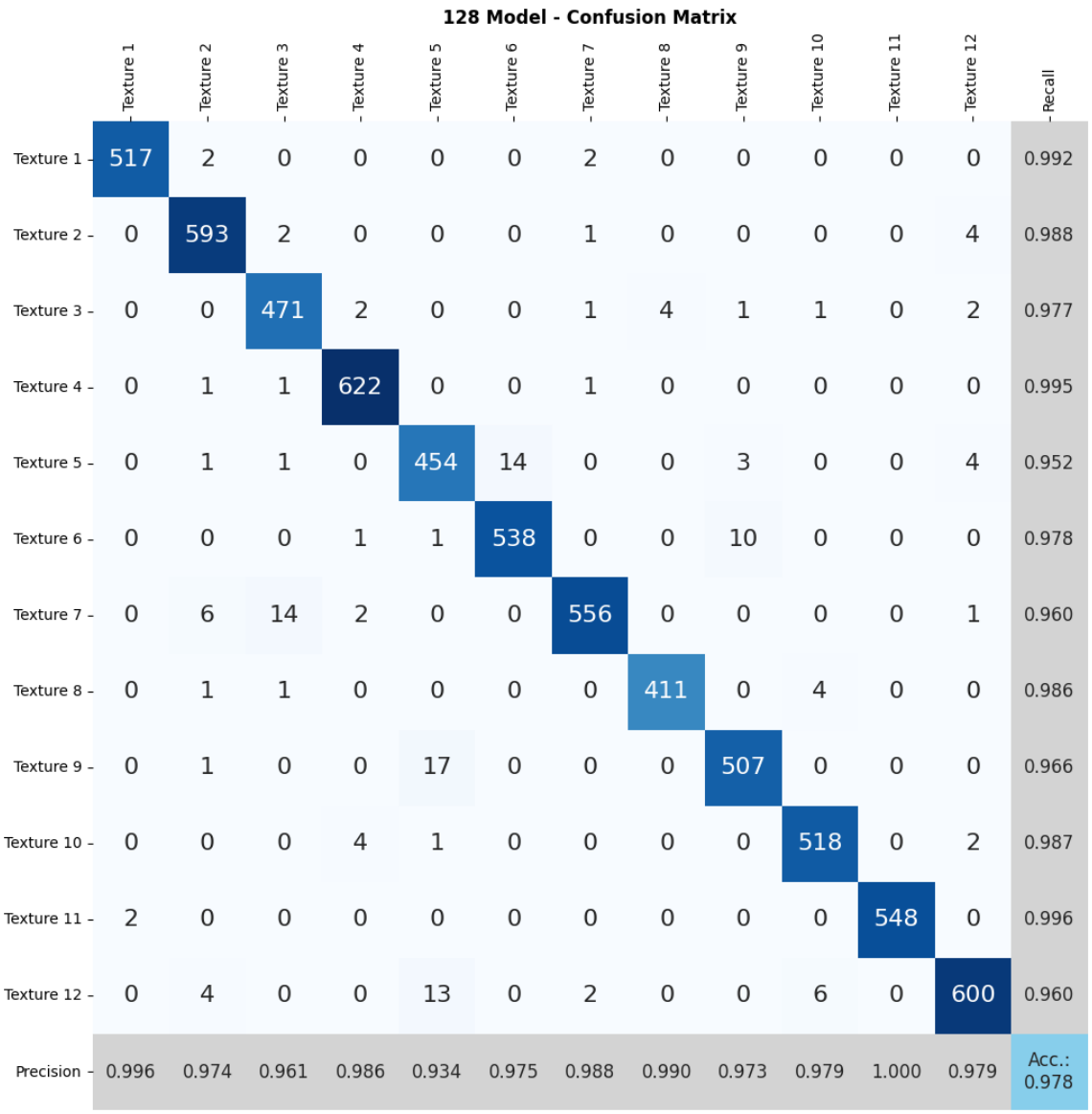

Figure 4.10: Confusion matrix for Self-Attention-Based model on Different Textures, Different Shapes with window size 128.

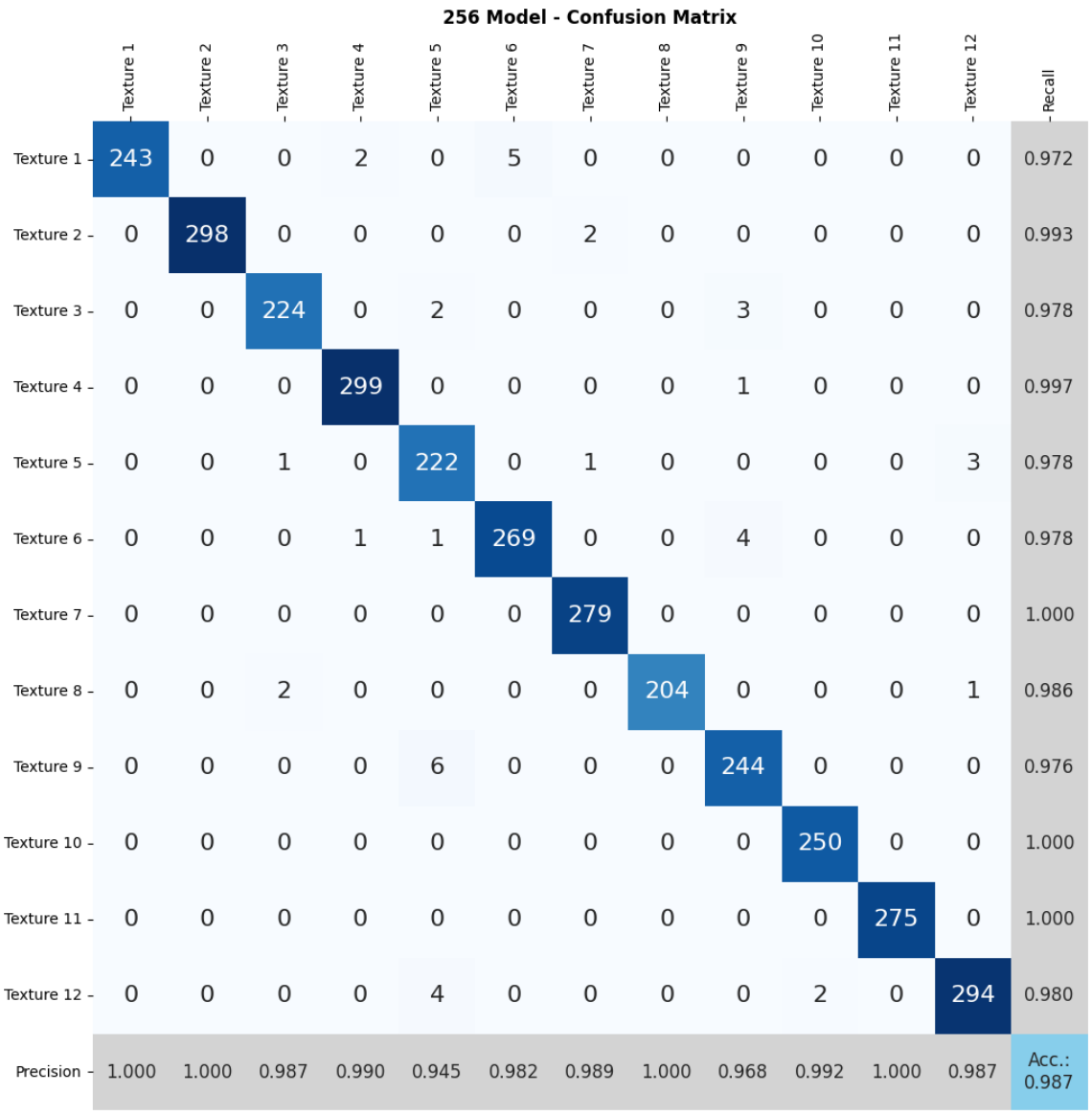

Figure 4.11: Confusion matrix for Self-Attention-Based model on Different Textures, Different Shapes with window size 256.

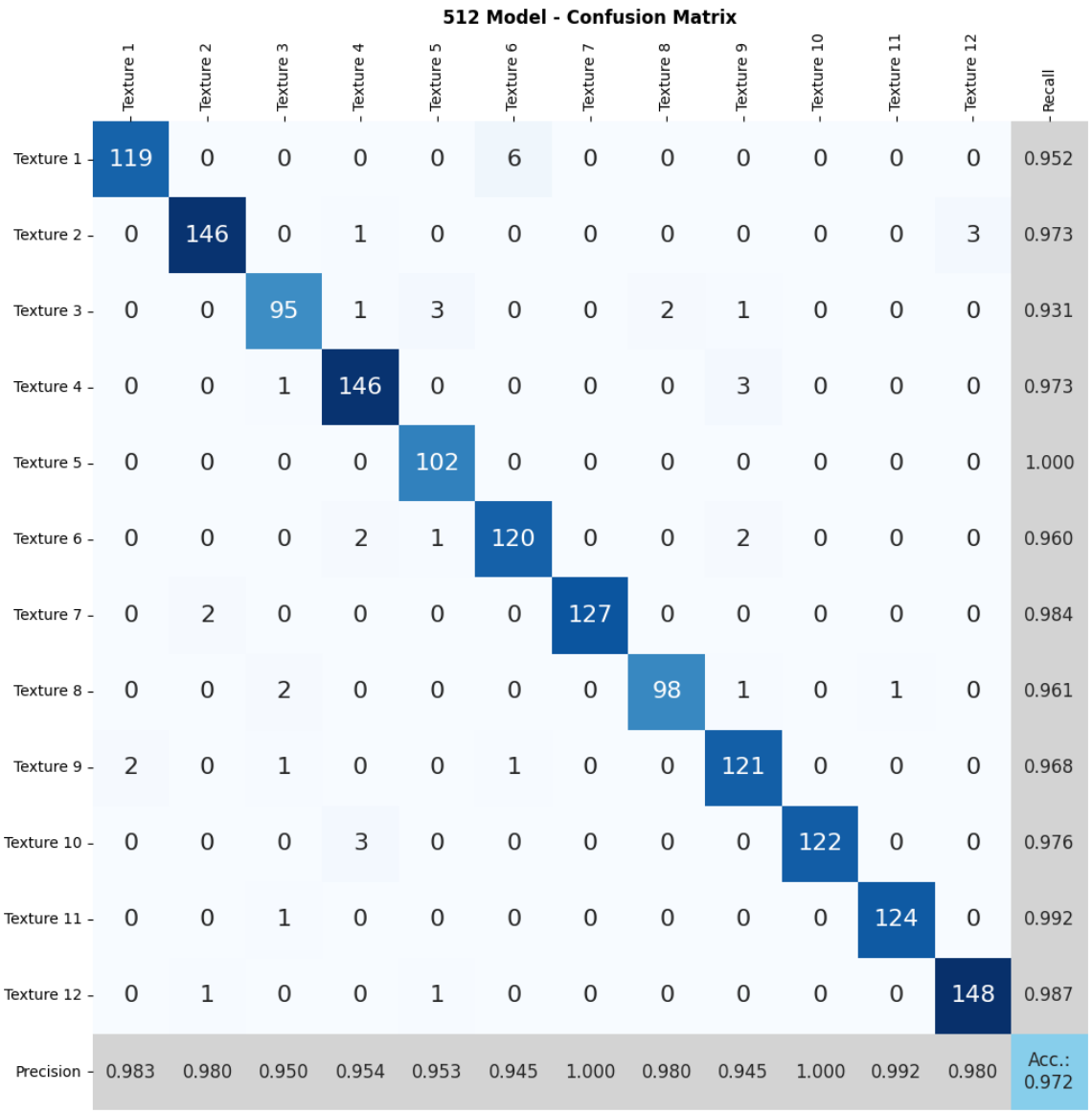

Figure 4.12: Confusion matrix for Self-Attention-Based model on Different Textures, Different Shapes with window size 512.

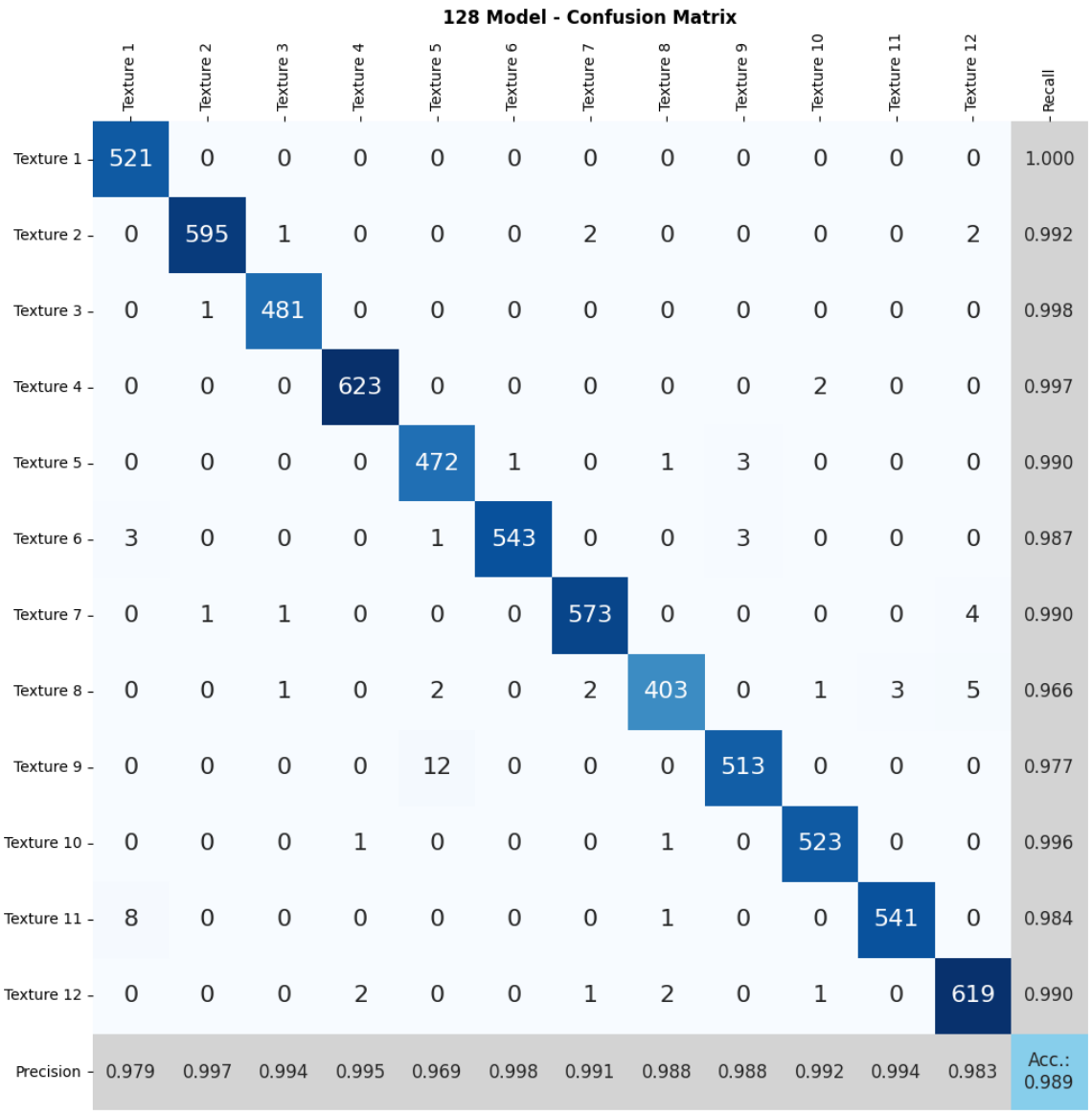

Figure 4.13: Confusion matrix for Hybrid CNN + Self-Attention on Different Textures, Different Shapes with window size 128.

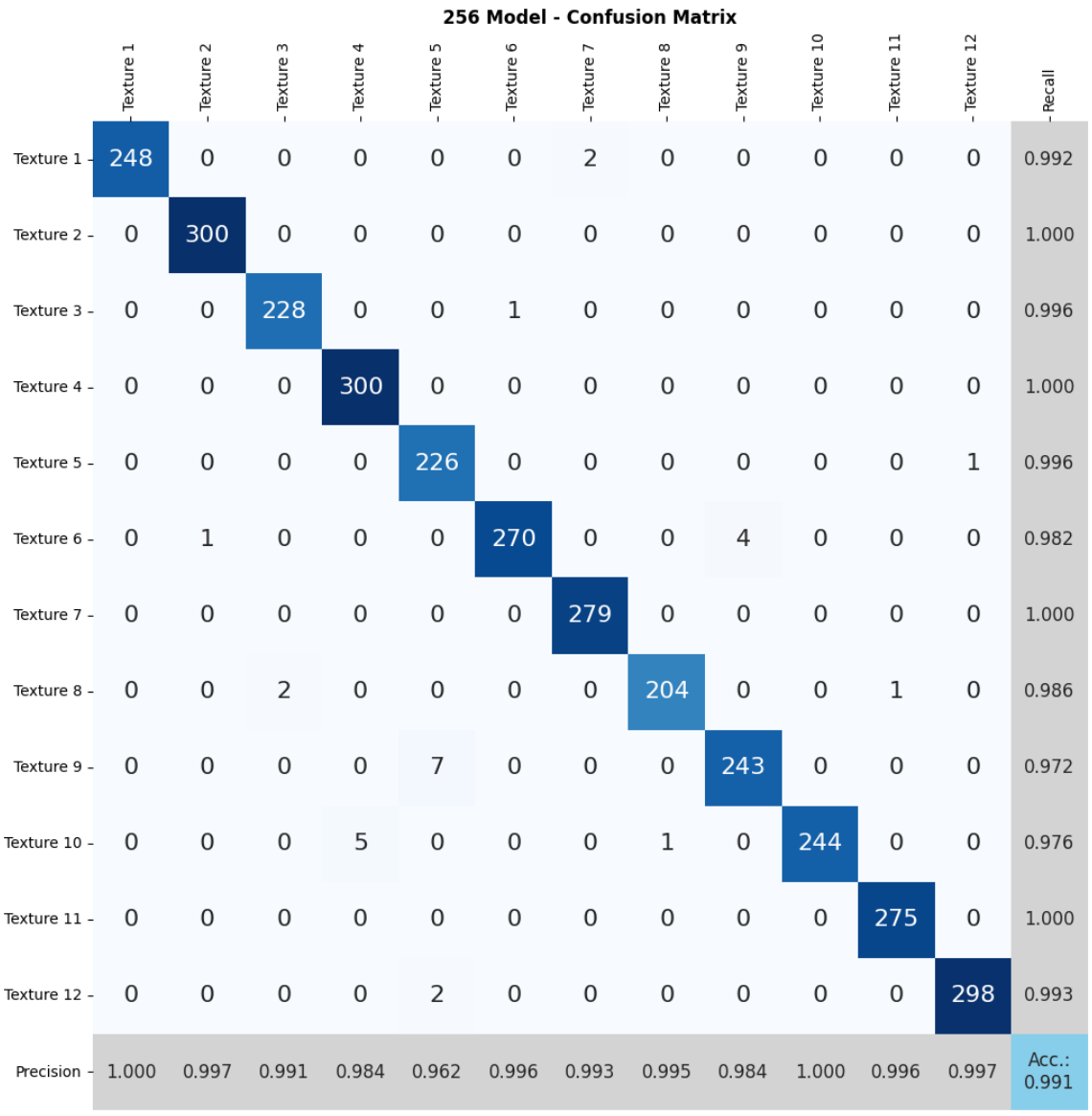

Figure 4.14: Confusion matrix for Hybrid CNN + Self-Attention on Different Textures, Different Shapes with window size 256.

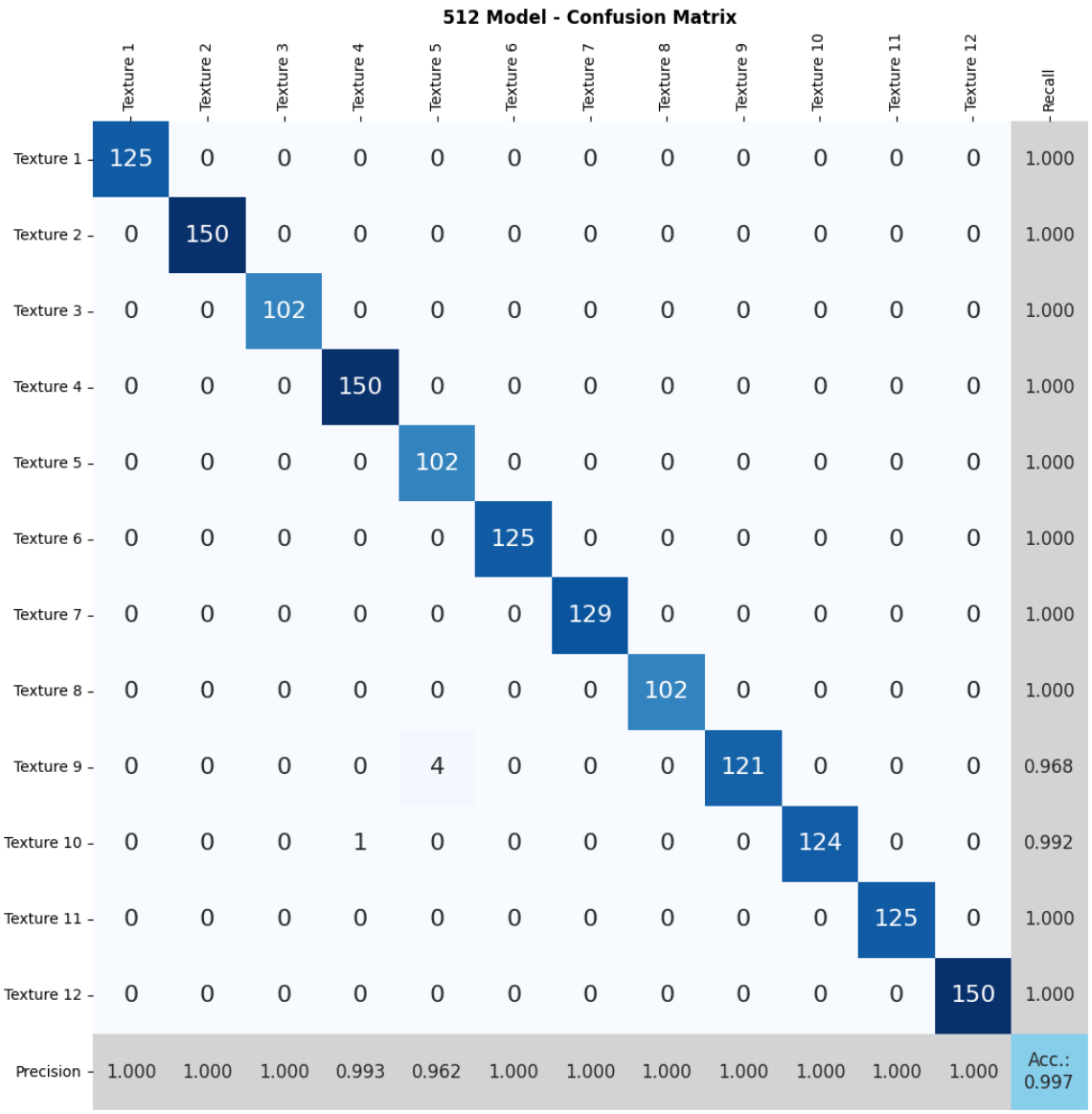

Figure 4.15: Confusion matrix for Hybrid CNN + Self-Attention on Different Textures, Different Shapes with window size 512.

Each configuration was designed to isolate specific challenges in tactile classification, such as shape invariance, texture consistency, or combined variability. All five model architectures—1D CNN, BiLSTM, Hybrid CNN-BiLSTM, Self-Attention-Based, and Hybrid CNN + Self-Attention—were evaluated across window sizes of 128, 256, and 512 using 5-fold cross-validation.

The confusion matrices for all 90 model runs (six datasets, five models, three window sizes) are included in this chapter and follow the same layout used for the initial dataset. The results demonstrate consistent performance trends: hybrid and attention-based models generally outperform the simpler baselines, and longer temporal windows lead to higher classification accuracy.

Across all evaluations, the Mixed Scenarios dataset posed the greatest challenge to the models, likely due to its higher variability in both geometry and texture. In contrast, the Flat All Textures configuration was the easiest, highlighting the role of geometric complexity in tactile texture recognition tasks.

4.3.3 Best Model Summary

To summarize the strongest performers across datasets, Table 4.1 lists the best model configuration for each dataset variant based on validation accuracy. In all six cases, the Hybrid CNN + Self-Attention architecture with a window size of 512 achieved the highest performance. This consistent result supports the idea that combining convolutional feature extraction with attention-based temporal modeling provides a robust and generalizable approach to tactile texture classification, particularly when longer temporal sequences are available.

The accuracies range between 94.0% and 95.8%, with the highest result observed

on the Flat All Textures dataset and the lowest on the Mixed Scenarios dataset. This reinforces the observation that geometric variability in surface shape presents a substantial challenge for tactile learning systems.

Table 4.1: Best Model Performance Across Datasets (Validation Accuracy)

Dataset	**Best Model**	**Accuracy (%)**
Different Textures, Same Shape	Hybrid CNN + Self-Attention (512)	97.8
Flat All Textures	Hybrid CNN + Self-Attention (512)	99.2
Concave/Convex All Textures	Hybrid CNN + Self-Attention (512)	95.2
Mixed Scenarios	Hybrid CNN + Self-Attention (512)	94.0

4.3.4 Voting-Based Ensemble Results

To enhance robustness and performance, a soft voting-based ensemble was introduced by combining the outputs of three models trained on different temporal window sizes (128, 256, and 512). As shown in the corresponding confusion matrices, the voting system consistently outperforms the best-performing individual model (Hybrid CNN + Self-Attention with window size 512) across all datasets.

The voting mechanism effectively reduces misclassifications by leveraging complementary temporal information captured at different scales. In several texture classes that were frequently confused in the single-model runs—especially in datasets with high shape variability such as Mixed Scenarios and Flat and Concave/Convex All Textures—the ensemble approach improves the class-wise recall and yields cleaner confusion matrices with fewer off-diagonal elements.

Furthermore, voting introduces a stabilizing effect: rather than relying on the

predictions of a single window size, it aggregates the confidence of all three, which helps mitigate the cases where one model might overfit to specific texture patterns or exhibit sensitivity to surface deformations.

This performance gain confirms the value of multi-scale temporal modeling for tactile texture classification and highlights ensemble learning as a powerful strategy in situations where individual models show slightly different strengths across input conditions.

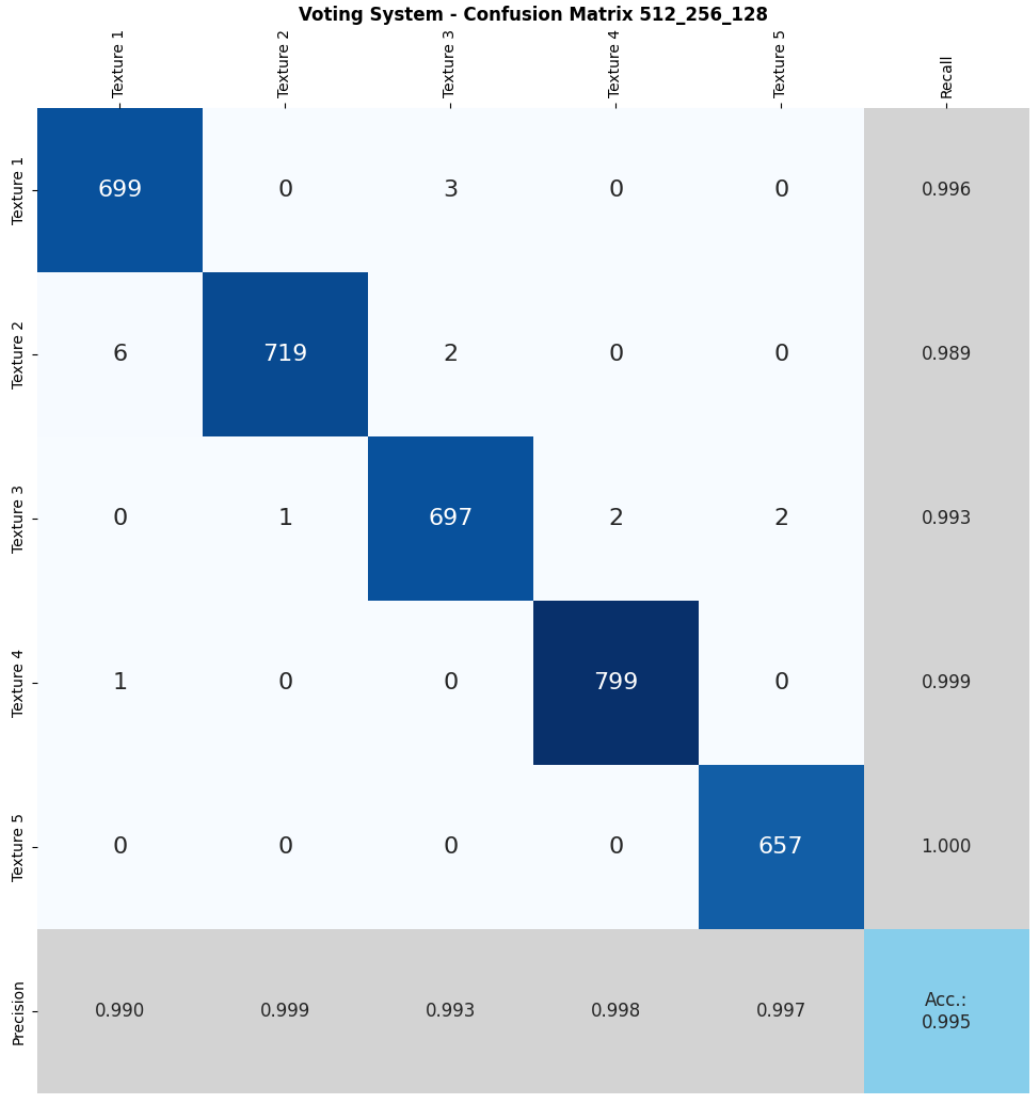

Figure 4.16: Confusion matrix for the voting-based ensemble model on the Same Textures- Different Shapes dataset. The ensemble combines predictions from models trained on window sizes 128, 256, and 512.

Chapter 5

Conclusion

Tactile perception plays a crucial role in enabling robots to interact meaningfully with their environments, particularly in scenarios involving object recognition and manipulation through contact. While significant progress has been made in classifying textures on flat surfaces, tactile understanding over uneven and complex geometries remains a challenging and underexplored problem. This thesis addressed this gap by developing a novel dataset, designing a classification pipeline, and proposing an ensemble-based approach to improve robustness and accuracy in tactile texture recognition under such conditions.

The primary contributions of this work are fourfold. First, a new dataset was collected using a compliant barometric–IMU tactile sensor mounted on a robotic arm, covering five types of uneven surfaces and one flat baseline, each textured with twelve distinct materials. Second, a structured point collection strategy was implemented based on reconstructed surface normals, allowing systematic contact interactions between the end effector and target surface. Third, we investigated multiple deep learning architectures, including self-attention-based models, to classify temporal tactile

signals acquired from the sensor. Finally, we proposed a voting-based ensemble system that aggregates predictions across three temporal window sizes (128, 256, 512), leading to consistent accuracy gains over any single-window approach and outperforming the baseline models from prior work.

Experimental results demonstrate that the proposed voting ensemble is not only more accurate but also more robust to variations in contact geometry and temporal signal distortion caused by the surface curvature. Compared to previous studies focused on flat surfaces, this work establishes a pipeline that generalizes better to real-world uneven conditions—closer to what robots may encounter in unstructured environments.

Despite these achievements, several limitations remain. The dataset, although diverse, is still constrained in terms of surface variety and interaction patterns. In addition, the models were trained offline, and real-time inference or continuous learning were not explored. These aspects leave room for further research.

Future directions include expanding the dataset to cover more complex surfaces and contact conditions, integrating additional sensing modalities such as tactile images or force-torque data, and developing models capable of online learning and real-time deployment. Another promising avenue is the incorporation of contact dynamics and proprioceptive feedback into the classification pipeline to further improve performance under more naturalistic interaction scenarios.

In summary, this thesis contributes both a valuable dataset and an effective classification method for tactile texture recognition on uneven surfaces, taking a step closer to enabling intelligent tactile perception in real-world robotic systems.

Bibliography

[1] Thiago Eustaquio Alves De Oliveira and Vinicius Prado Da Fonseca. BioInTacto: A compliant multi-modal tactile sensing module for robotic tasks. *HardwareX*, 16:e00478, December 2023.

[2] Thiago Eustaquio Alves De Oliveira, Ana-Maria Cretu, and Emil M. Petriu. Multimodal Bio-Inspired Tactile Sensing Module. *IEEE Sensors Journal*, 17(11):3231–3243, June 2017.

[3] Shan Luo, Joao Bimbo, Ravinder Dahiya, and Hongbin Liu. Robotic tactile perception of object properties: A review. *Mechatronics*, 48:54–67, December 2017.

[4] Qiang Li, Oliver Kroemer, Zhe Su, Filipe Fernandes Veiga, Mohsen Kaboli, and Helge Joachim Ritter. A Review of Tactile Information: Perception and Action Through Touch. *IEEE Transactions on Robotics*, 36(6):1619–1634, December 2020.

[5] Ravinder S. Dahiya, Philipp Mittendorfer, Maurizio Valle, Gordon Cheng, and Vladimir J. Lumelsky. Directions Toward Effective Utilization of Tactile Skin: A Review. *IEEE Sensors Journal*, 13(11):4121–4138, November 2013.

[6] J Dargahi and S Najarian. Human tactile perception as a standard for artificial tactile sensing—a review. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 1(1):23–35, June 2004.

[7] Maliheh Marzani, Soheil Khatibi, Ruslan Masinjila, Vinicius Prado Da Fonseca, and Thiago Eustaquio Alves De Oliveira. A dataset for tactile textures on uneven surfaces collected using a BioIn-Tacto sensing module. *Data in Brief*, 59:111312, April 2025.

[8] Youcan Yan, Zhe Hu, Yajing Shen, and Jia Pan. Surface Texture Recognition by Deep Learning-Enhanced Tactile Sensing. *Advanced Intelligent Systems*, 4(1):2100076, January 2022.

[9] Shiyao Huang and Hao Wu. Texture Recognition Based on Perception Data from a Bionic Tactile Sensor. *Sensors*, 21(15):5224, August 2021.

[10] Daniel J Kucherhan, Miriam Goubran, Vinicius P da Fonseca, Thiago E Alves de Oliveira, Emil M Petriu, and Voicu Groza. Object recognition through manipulation using tactile enabled prosthetic fingers and feedback glove-experimental study. In *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE, 2018.

[11] Qi Zhu, Vinicius Prado da Fonseca, Bruno Monteiro Rocha Lima, Maxwell Welyhorsky, Miriam Goubran, Thiago Eustaquio Alves de Oliveira, and Emil M Petriu. Teleoperated grasping using a robotic hand and a haptic-feedback data glove. In *2020 IEEE International Systems Conference (SysCon)*, pages 1–7. IEEE, 2020.

[12] Viral Rasik Galaiya, Mohammed Asfour, Thiago Eustaquio Alves de Oliveira, Xianta Jiang, and Vinicius Prado da Fonseca. Exploring tactile temporal features for object pose estimation during robotic manipulation. *Sensors*, 23(9):4535, 2023.

[13] Thiago Eustaquio Alves De Oliveira, Ana-Maria Cretu, Vinicius Prado Da Fonseca, and Emil M Petriu. Touch sensing for humanoid robots. *IEEE Instrumentation & Measurement Magazine*, 18(5):13–19, 2015.

[14] Thiago Eustaquio Alves de Oliveira, Vinicius Prado da Fonseca, Bruno Monteiro Rocha Lima, Ana-Maria Cretu, and M Petriu. End-effector approach flexibilization in a surface approximation task using a bioinspired tactile sensing module. In *2019 IEEE International Symposium on Robotic and Sensors Environments (ROSE)*, pages 1–6. IEEE, 2019.

[15] Laurent YE Ramos Cheret, Vinicius Prado Da Fonseca, and Thiago E Alves de Oliveira. Leveraging compliant tactile perception for haptic blind surface reconstruction. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 17139–17145. IEEE, 2024.

[16] Partha Sarati Das, Simon Rondeau-Gagné, and Mohammed Jalal Ahamed. Triboelectric pressure sensor with microstructured pdms for human motion and gait pattern monitoring. *IEEE Journal on Flexible Electronics*, 2025.

[17] Dapeng Chen, Xiaorong Huang, Peng Gao, Bin Wang, Lina Wei, Xuhui Hu, Hong Zeng, Jia Liu, and Aiguo Song. A High Spatial Resolution Magnetorheological Elastomer Tactile Sensor for Texture Recognition. *IEEE Sensors Journal*, pages 1–1, 2025.

[18] Giwon Lee, Jong Hyun Son, Siyoung Lee, Seong Won Kim, Daegun Kim, Nguyen Ngan Nguyen, Seung Goo Lee, and Kilwon Cho. Fingerpad-Inspired Multimodal Electronic Skin for Material Discrimination and Texture Recognition. *Advanced Science*, 8(9):2002606, May 2021.

[19] Si Chen, Su Li, Yiting Zheng, Brian Fong, Yizong Li, Penghao Dong, David Hwang, and Shanshan Yao. Highly sensitive and robust soft tri-axial tactile sensors enabled by dual inductive sensing mechanisms. *Soft Science*, 5(1):N–A, 2025.

[20] Gabriele M Caddeo, Andrea Maracani, Paolo D Alfano, Nicola A Piga, Lorenzo Rosasco, and Lorenzo Natale. Sim2surf: A sim2real surface classifier for vision-based tactile sensors with a bilevel adaptation pipeline. *IEEE Sensors Journal*, 2025.

[21] Qiwei Wu, Haidong Wang, Jiayu Zhou, Xiaogang Xiong, and Yunjiang Lou. Tars: Tactile affordance in robot synesthesia for dexterous manipulation. *IEEE Robotics and Automation Letters*, 2024.

[22] Viral Galaiya, Ruslan Masinjila, Soheil Khatibi, Thiago Eustaquio Alves de Oliveira, Vinicius Prado da Fonseca, and Xianta Jiang. Extraction of non-regular pegs using tactile sensing and reinforcement learning from demonstrations. In *2025 IEEE International systems Conference (SysCon) (IEEE SysCon 2025)*, page 6, Montreal, Canada, April 2025.

[23] Viral Rasik Galaiya, Thiago Eustaquio Alves De Oliveira, Xianta Jiang, and Vinicius Prado Da Fonseca. Grasp approach under positional uncertainty using compliant tactile sensing modules and reinforcement learning. In *2024 IEEE*

Canadian Conference on Electrical and Computer Engineering (CCECE), pages 424–428. IEEE, 2024.

[24] Vinicius Prado da Fonseca, Daniel John Kucherhan, Thiago E Alves de Oliveira, Da Zhi, and Emil M Petriu. Fuzzy controlled object manipulation using a three-fingered robotic hand. In *2017 Annual IEEE international systems conference (SysCon)*, pages 1–6. IEEE, 2017.

[25] Ana-Maria Cretu, Thiago Eustaquio Alves De Oliveira, Vinicius Prado Da Fonseca, Bilal Tawbe, Emil M Petriu, and Voicu Z Groza. Computational intelligence and mechatronics solutions for robotic tactile object recognition. In *2015 IEEE 9th international symposium on intelligent signal processing (WISP) proceedings*, pages 1–6. IEEE, 2015.

[26] Vinicius Prado da Fonseca, Thiago Eustaquio Alves de Oliveira, and Emil M Petriu. Estimating the orientation of objects from tactile sensing data using machine learning methods and visual frames of reference. *Sensors*, 19(10):2285, 2019.

[27] Thiago Eustaquio Alves de Oliveira, Vinicius Prado da Fonseca, Emanuil Huluta, Paulo FF Rosa, and Emil M Petriu. Data-driven analysis of kinaesthetic and tactile information for shape classification. In *2015 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pages 1–5. IEEE, 2015.

[28] Thiago Eustaquio Alves de Oliveira, Bruno Monteiro Rocha Lima, Ana-Maria Cretu, and Emil M Petriu. Tactile profile classification using a multimodal mems-based sensing module. In *Proceedings*, volume 1, page 27. MDPI, 2016.

[29] Vinicius Prado da Fonseca, Xianta Jiang, Emil M Petriu, and Thiago Eustaquio Alves de Oliveira. Tactile object recognition in early phases of grasping using underactuated robotic hands. *Intelligent Service Robotics*, 15(4):513–525, 2022.

[30] Bruno Monteiro Rocha Lima, Vinicius Prado da Fonseca, Thiago Eustaquio Alves de Oliveira, Qi Zhu, and Emil M Petriu. Dynamic tactile exploration for texture classification using a miniaturized multi-modal tactile sensor and machine learning. In *2020 IEEE International Systems Conference (SysCon)*, pages 1–7. IEEE, 2020.

[31] Bruno Monteiro Rocha Lima, Thiago Eustaquio Alves de Oliveira, and Vinicius Prado da Fonseca. Classification of textures using a tactile-enabled finger in dynamic exploration tasks. In *2021 IEEE Sensors*, pages 1–4. IEEE, 2021.

[32] Bruno Monteiro Rocha Lima, Venkata Naga Sai Siddhartha Danyamraju, Thiago Eustaquio Alves de Oliveira, and Vinicius Prado da Fonseca. A multimodal tactile dataset for dynamic texture classification. *Data in Brief*, 50:109590, 2023.

[33] Bruno Monteiro Rocha Lima, Thiago Eustaquio Alves de Oliveira, Vinicius Prado da Fonseca, Qi Zhu, Miriam Goubran, Voicu Z Groza, and Emil M Petriu. Heart rate detection using a miniaturized multimodal tactile sensor. In *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE, 2019.

[34] Bruno Monteiro Rocha Lima, Luiz Claudio Sampaio Ramos, Thiago Eustaquio Alves de Oliveira, Vinicius Prado da Fonseca, and Emil M Petriu. Heart

rate detection using a multimodal tactile sensor and a z-score based peak detection algorithm. *CMBES proceedings*, 42, 2019.

[35] Jung-Hwan Yang, Seong-Yong Kim, and Soo-Chul Lim. Effects of Sensing Tactile Arrays, Shear Force, and Proprioception of Robot on Texture Recognition. *Sensors*, 23(6):3201, March 2023.

[36] Alina Böhm, Tim Schneider, Boris Belousov, Alap Kshirsagar, Lisa Lin, Katja Doerschner, Knut Drewing, Constantin A Rothkopf, and Jan Peters. What matters for active texture recognition with vision-based tactile sensors. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 15099–15105. IEEE, 2024.

[37] Maliheh Marzani, Soheil Khatibi, Vinicius Prado da Fonseca, and Thiago Eustaquio Alves de Oliveira. Texture recognition on uneven surfaces using deep learning and tactile sensing techniques. In *2025 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) (IEEE CCECE 2025)*, page 7, Vancouver, Canada, May 2025.

[38] Laurent Y. Emile Ramos Cheret, Soheil Khatibi, Vinicius Prado da Fonseca, and Thiago Eustaquio Alves de Oliveira. Enhancing tactile texture recognition from haptic surface reconstruction using reinforcement learning. In *2025 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) (IEEE CCECE 2025)*, page 6.98, Vancouver, Canada, May 2025.

[39] Soheil Khatibi, Maliheh Marzani, Ruslan Masinjila, Vinicius Prado da Fonseca, and Thiago Eustaquio Alves de Oliveira. Tactile texture recognition on uneven surfaces using self-attention based neural networks. In *Proceedings of the IEEE*

49th Annual Computers, Software, and Applications Conference (COMPSAC). IEEE, 2025.

[40] Soheil Khatibi, Maliheh Marzani, Ruslan Masinjila, Vinicius Prado da Fonseca, and Thiago Eustaquio Alves de Oliveira. Multi-scale voting system for robotic tactile texture recognition on uneven surfaces. In *Proceedings of the IEEE 34th International Symposium on Industrial Electronics (ISIE)*. IEEE, 2025.

[41] Shemonto Das, Vinicius Prado Da Fonseca, and Amilcar Soares. Active learning strategies for robotic tactile texture recognition tasks. *Frontiers in Robotics and AI*, 11:1281060, February 2024.

[42] Viral Galayia, Ruslan Masinjila, Soheil Khatibi, Thiago Eustaquio Alves de Oliveira, Xianta Jiang, and Vinicius Prado da Fonseca. A multimodal dataset for robotic peg extraction based on bioin-tacto sensor modules. *Data in Brief*, 59:111356, 2025.

[43] Xiaochun Yin, Zengguang Liu, Deyong Liu, and Xiaojun Ren. A Novel CNN-based Bi-LSTM parallel model with attention mechanism for human activity recognition with noisy data. *Scientific Reports*, 12(1):7878, May 2022.

[44] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, 2017. Version Number: 7.

[45] Baojiang Li, Liang Li, Haiyan Wang, Guochu Chen, Bin Wang, and Shengjie Qiu. TVT-Transformer: A Tactile-visual-textual fusion network for object recognition. *Information Fusion*, 118:102943, June 2025.

[46] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting, March 2021. arXiv:2012.07436 [cs].

[47] Guanqun Cao, Yi Zhou, Danushka Bollegala, and Shan Luo. Spatio-temporal Attention Model for Tactile Texture Recognition. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9896–9902, Las Vegas, NV, USA, October 2020. IEEE.

[48] Youcan Yan, Zhe Hu, Yajing Shen, and Jia Pan. Surface texture recognition by deep learning-enhanced tactile sensing. *Advanced Intelligent Systems*, 4(1):2100076, 2022.

[49] Behnam Rostamian, MohammadReza Koolani, Pouya Abdollahzade, Milad Lankarany, Egidio Falotico, Mahmood Amiri, and Nitish V. Thakor. Texture recognition based on multi-sensory integration of proprioceptive and tactile signals. *Scientific reports*, 12(1):21690, 2022.