This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive version was published in VRST '23: Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology, available at ACM Digital Library [https://doi.org/10.1145/3611659.3615714].

Predicting Perceptual Haptic Attributes of Textured Surface from Tactile Data Based on Deep CNN-LSTM Network

Mudassir Ibrahim Awan*
Kyung Hee University
South Korea
miawan@khu.ac.kr

Waseem Hassan*
Kyung Hee University
South Korea
waseem.h@khu.ac.kr

Seokhee Jeon
Kyung Hee University
South Korea
jeon@khu.ac.kr

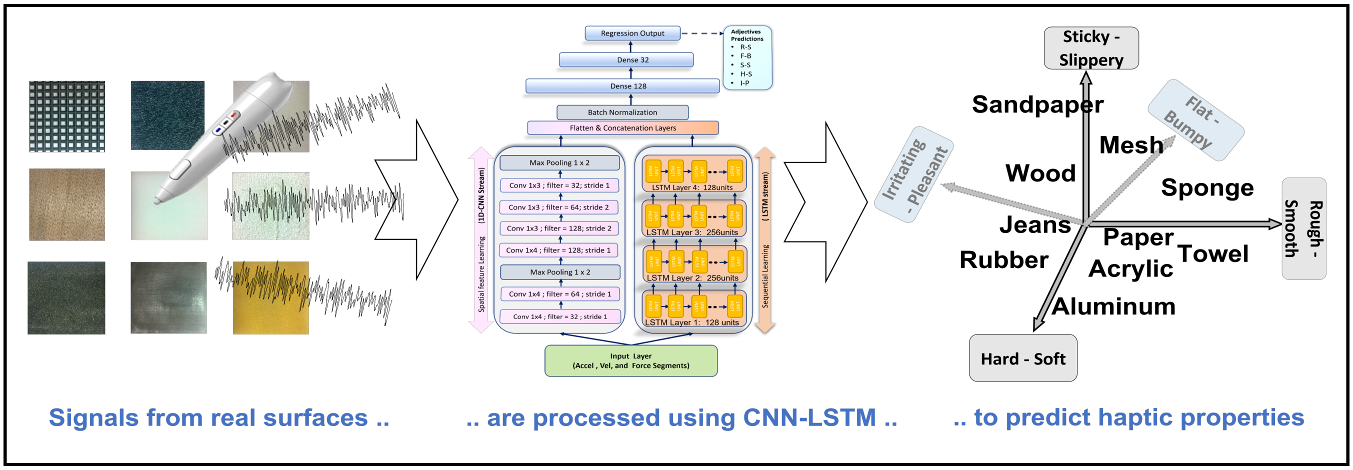

Regression Output
Adjectives Predictions
R-S
F-B
S-S
H-S
I-P
Dense 32
Dense 128
Batch Normalization
Flatten & Concatenation Layers
Spatial feature Learning (1D-CNN Stream)
Max Pooling 1 x 2
Conv 1x3 ; filter = 32; stride 1
Conv 1x3 ; filter = 64; stride 2
Conv 1x3 ; filter = 128; stride 2
Conv 1x4 ; filter = 128; stride 1
Max Pooling 1 x 2
Conv 1x4 ; filter = 64 ; stride 1
Conv 1x4 ; filter = 32 ; stride 1
LSTM UNIT
LSTM Layer 4: 128units
LSTM Layer 3: 256units
LSTM Layer 2: 256units
LSTM Layer 1: 128 units
Sequential Learning (LSTM stream)
Input Layer
(Accel , Vel, and Force Segments)
Sticky - Slippery
Flat - Bumpy
Irritating - Pleasant
Rough - Smooth
Hard - Soft
Sandpaper
Mesh
Wood
Sponge
Jeans
Paper
Towel
Rubber
Acrylic
Aluminum
Signals from real surfaces ..
.. are processed using CNN-LSTM ..
.. to predict haptic properties

Figure 1: A CNN-LSTM model leveraging the mapping between physical characteristics and perceptual haptic attributes of textures to map physical surfaces onto a five-dimensional haptic attribute space.

ABSTRACT

This paper introduces a framework to predict multi-dimensional haptic attribute values that humans use to recognize the material by using the physical tactile signals (acceleration) generated when a textured surface is stroked. To this end, two spaces are established: a haptic attribute space and a physical signal space. A five-dimensional haptic attribute space is established through human adjective rating experiments with the 25 real texture samples. The physical space is constructed using tool-based interaction data from the same 25 samples. A mapping is modeled between the aforementioned spaces using a newly designed CNN-LSTM deep learning network. Finally, a prediction algorithm is implemented that takes acceleration data and returns coordinates in the haptic attribute space. A quantitative evaluation was conducted to inspect the reliability of the algorithm on unseen textures, showing that the model outperformed other similar models.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; • **Hardware** → Haptic devices; Sensor applications and deployments; • **Computing methodologies** → *Machine translation.*

KEYWORDS

Haptic texture classification, psychophysics, neural network

ACM Reference Format:

Mudassir Ibrahim Awan, Waseem Hassan, and Seokhee Jeon. 2023. Predicting Perceptual Haptic Attributes of Textured Surface from Tactile Data Based on Deep CNN-LSTM Network. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (VRST).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

When a textured surface is stroked, a wide range of tactile signals are generated. Humans perceive the signals and are able to gauge the multi-dimensional attributes of their texture [16, 25, 36]. The conversion from the signals into multi-dimensional attributes is done very quickly and efficiently in the human cognitive process and is one of the more salient human abilities to function in daily life interactions. The perceived attributes are used to recognize the material, identify the object, and eventually perform any tasks that are required. To this end, the role of tactile data on human behavior has been intensively studied [20, 22, 23, 30].

The sensing process of this conversion can be mimicked by computers. The signals are sensed by computers and material attributes

[*] These authors contributed equally to this research and Seokhee Jeon (jeon@khu.ac.kr) is the corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

VRST, Oct 09–11, 2023, New Zeland
© 2023 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

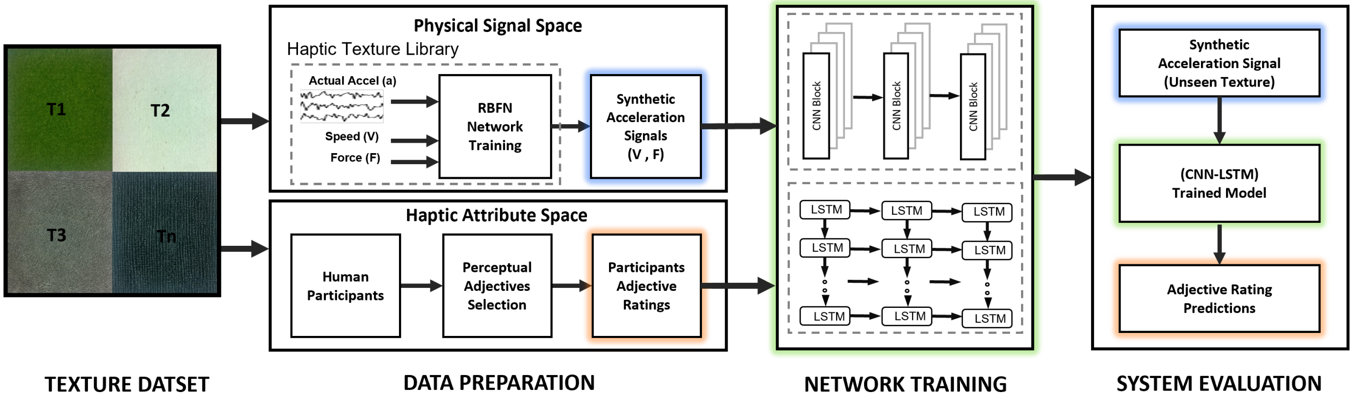


Figure 2: Overall framework. Texture dataset: a dataset containing real-world textures was prepared. Data preparation: establishing haptic attribute space and physical signal space using the dataset. Network training: a CNN-LSTM based model learning the relationship between the two spaces. System evaluation: an evaluation conducted to see the predictability of haptic attributes for a new/unseen texture.

of the surface from which the signals are captured can be predicted with the help of sophisticated algorithms. This is one of the essential capabilities needed for humanoid robots and has been investigated in the field of robotics [7, 23, 26, 30]. These studies largely focused on the prediction of the physical attributes, while their perceptual effect has mostly been less studied. From a robotic interaction point of view, it makes sense for the robots to only be concerned with the physical attributes of the texture, however, when the said robots interact with humans, the perceptual aspect of the physical signals should also be considered. The estimation of attributes is also very beneficial in the field of haptics where the core interest is to deliver synthetic touch feedback to humans. For instance, in a bilateral teleoperation scenario, a remote robot directly estimates attributes of the touched surface and sends them back for saving bandwidth without compromising the perceptual performance. In addition, automatic conversion is very useful for haptic content generation and authoring. Editing, authoring, and describing haptic texture would be much easier in a perception-based attribute space than in a physical value space, e.g., slightly increasing the roughness attribute of a texture model, and blending or interpolating two textures in attribute space. Such a system would enable direct composition of arbitrary texture models guided by the perceptual attribute scale (like creating colors with an RGB model), which would greatly facilitate texture content creation.

Integral to establishing a direct link, between the human perception of textures and the physical signals generated, is the automatic bi-directional conversion between the signals and attributes of texture. There are several required components to realize such a system. First, establishing the two spaces, i.e., the physical signal space and the texture attribute/perception space. Hereinafter the texture perception space would be referred to as the haptic attribute space. Second, modeling the causation chain between the two spaces. Third, devising an algorithm to honestly estimate the haptic attributes of textures from their physical signals such that the established causation chain is upheld. Fourth, an algorithm that can traverse the causation chain in reverse and generate true physical signals from a set of haptic attributes. To the best of our knowledge, such an endeavor has not yet been undertaken in the existing literature. This paper is our first step towards realizing the aforementioned system, and as an initial proof of concept, we embark upon the first, second, and third components.

Various researchers have touched upon the individual snippets of the aforementioned system. After the initial attempts for the fundamental comprehension of haptic perceptual characteristics and their mapping onto the haptic perceptual space [12, 16, 25, 36], several recent works have tried to foresee the perceptual effect of different physical characteristics of surface textures. For example, in [27], the authors have predicted the surface dissimilarities of 10 textures by comparing the corresponding probability distributions of collected physical tactile signals. In another example, the GelSight sensor was employed at the end effector of a robotic arm to establish a physical signal space by analyzing surface geometry and shear force [5]. Some researchers have also attempted to create virtual textures with varying perceptual attributes [24], where they created 12 virtual textures and explored force-based perceptual characteristics on the general force feedback using Hooke's Law force model. However, the aforementioned and other similar studies considered did not consider the two spaces jointly in their prediction algorithms. These studies also failed to consider the general transition or study the causation chain between the two spaces.

The establishment of the attribute space is guided by the methodologies introduced in the literature. A general approach to defining underlying determinants of texture perception is to create a haptic perceptual space by performing psychophysical experiments. The experimental results are analyzed by using multi-dimensional scaling, yielding a perceptual space. Oftentimes, the dimensions of the perceptual space are realigned to be defined by attributes/adjectives. Generally, these perceptual spaces are comprised of three to five dimensions depicting ratings of different perceptual attributes assigned by human participants (e.g., hardness, warmth, roughness, etc.). More details on this alignment and dimensional structuring can be seen in [12, 16, 25, 36].

The current paper starts with establishing the two spaces, as illustrated in Fig. 2. In the first step, a 5-dimensional haptic attribute space is instituted through a perceptual user study involving 25 real texture samples, creating a framework for understanding how human participants perceive different textures. Simultaneously, the physical space is assembled using 3D acceleration signals captured with a rigid tool. These signals, guided by interaction force and speed, allow for an empirical representation of the textures' physical properties. For each texture sample, a data-driven texture rendering model is constructed, and subsequently used to estimate the 3D acceleration signals for arbitrary values of interaction force and speed.

The next phase involves modeling the relationship between the physical and perceptual attributes, which constitutes a significant component of this study. We designed a 1D-convolutional neural network (1D-CNN) combined in parallel with a long-short-term memory (LSTM) network. This hybrid approach is trained with the data from the two spaces established earlier, where the input is the 3D acceleration data, and the output or labels are the perceptual attributes. Once the network is trained, it is hypothesized to be able to predict the perceptual attributes of textures based on interaction signals. The efficacy of the prediction ability of the network is gauged by carrying out a numerical evaluation that shows promising results.

2 ESTABLISHING HAPTIC ATTRIBUTE SPACE

An adjective rating experiment is conducted to establish a haptic attribute space using real-world textures. In the first part, participants were asked to choose attributes that they felt could effectively describe the surface textures. In the second part, they rated surface textures based on the attribute pairs. These adjective ratings assigned by users are then used to populate a haptic attribute space. This section describes the details of the samples and the experiments.

2.1 Texture Dataset

In this study, twenty-five real-world textures were used to establish the space (see Fig.3). The texture dataset is selected in such a way that the majority of textures that we encounter everyday can be represented (e.g., textiles, fabrics, paper, wood, glossy, meshes, rubber, and foil). These real texture materials were cut and mounted on hard acrylic plates in order to avoid the effect of underlying objects during the experiments. The size of the acrylic and texture surface was set to 100x100x5 mm. Liquid surface glue was used to stick these textures on acrylic.

2.2 Participants

A total of 12 healthy participants (seven male and five female) with ages ranging from 23 to 30 years took part in this experiment. They reported no disabilities that could hinder their ability to take part in this experiment. The participants were compensated 10,000 KRW for their participation, which is equivalent to approximately 8.5 USD. The same participants took part in both the sub-experiments.

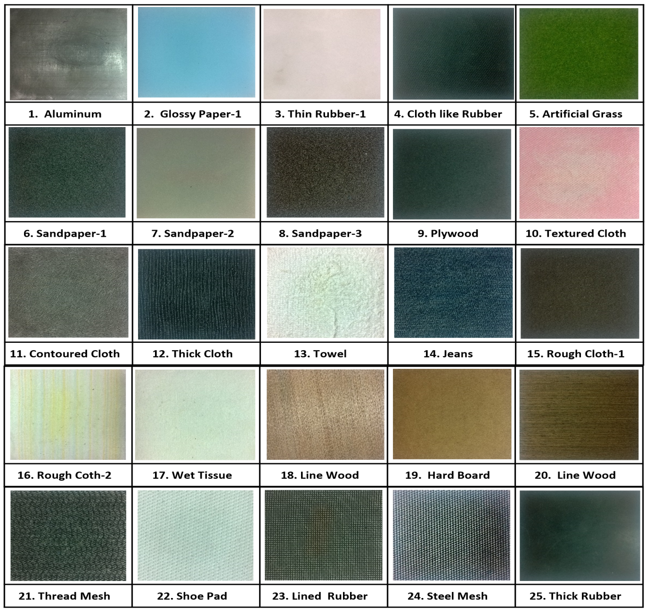


Figure 3: The real-world texture dataset used in this study.

Table 1: The list of adjectives used in experiment 1 of establishing the haptic attribute space. The boldface names are the ones that were selected for forming the antonymous adjectives pairs and had a relevance score of 50 % or more according to the participants.

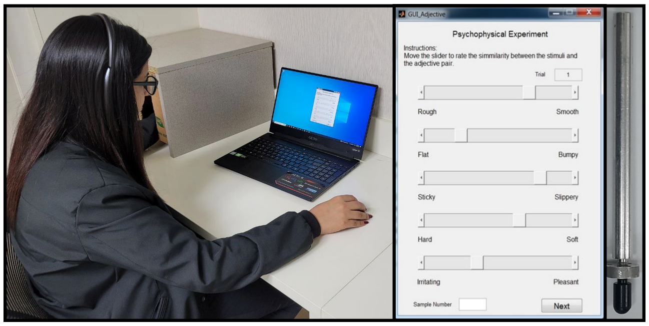
Sticky	Uneven	**Flat**	Even	**Irritating**
Rough	Rigid	Dense	**Bumpy**	**Pleasant**
Sharp	Thick	**Slippery**	Dull	Sparse
Smooth	Prickly	Thin	Soothing	**Soft**
Fine	Metallic	**Hard**	Pointy	Abrasive

2.3 Experiment 1: List of Attributes

The main objective of the first part was to assemble a list of adjectives that can define the perceptual attributes of texture surfaces used in this study. The list would be narrowed down by the participants and the results would be used in the second part of this experiment.

Experimental Setup. For the experiment, the participants sat down on a chair in front of a table. They were wearing headphones playing white noise to mask interaction noises with the texture. They were handed a printed paper containing experimental instructions. One texture sample was handed to the participant at a time. The sample was placed in a box with an opening for a hand. The mode of interaction was with a tool of length of 14 cm. The tip of the tool was 7mm in diameter.

Procedure. An initial list of 25 adjectives (see Table 1) that could describe the perceptual attributes of the given textures was compiled. These adjectives were some of the most commonly used adjectives in literature [12, 13, 17, 24, 34, 35]. The participants were asked to select adjectives (from the provided list) that they think


GUI_Adjective

Psychophysical Experiment

Instructions:
Move the slider to rate the simmilarity between the stimuli and the adjective pair.

Figure 4: The left side of the figure shows the experimental setup for the adjective rating experiment. In the middle, the GUI is displayed with the five antonymous adjective pairs for rating. On the right, is the rigid tool used in the experiments.

can define the perceptual characteristics of the given texture sample. They were asked to answer with either a "1" if they thought that an adjective was relevant for the current texture, or a "0" if the adjective was irrelevant. Each participant explored all the texture samples. They were allowed to use any exploration strategy that suited their preference and they were allowed to interact for as long as they wished.

Results. The scores received by each adjective for all the textures were added and normalized to form the relevance score. The adjectives that received at least a 50 % relevance score were selected for further processing. A total of 14 adjectives were shortlisted based on this criterion. The adjectives that had an antonymous pair within the 14 adjectives were selected and the remaining were discarded. These adjectives were utilized to form antonymous pairings, such that they represented the opposite extremes of the same property. As a result, five antonymous pairs of adjectives (i,e., Rough-Smooth, Flat-Bumpy, Sticky-Slippery, Hard-Soft, and Irritating-Pleasant) were selected for the next part of the experiment.

2.4 Experiment 2: Adjective Rating

In the second part of the experiment, the same participants were asked to rate each texture sample in terms of the five adjective pairs selected in the earlier experiment. This rating would represent the textures in a five-dimensional attribute space where each dimension is an adjective pair.

Experimental setup. The participants sat on a chair in front of a desk. A computer was placed on the desk in front of the participant. The computer was running a graphical user interface (GUI) which was used for rating the textures by the participants. The GUI had the five adjective pairs with a slider for each pair. The adjectives from each pair were located at the opposite extremes of the slider. The length of the slider was 127 mm on screen and did not have any scale markings on it [17]. One texture sample was provided at a time in a box that had an opening for the participant's hand. Participants interacted with the texture sample with a tool using their left hand, while the data entry was carried out by a computer mouse in their right hand. The experimental setup and GUI can be seen in Fig. 4.

Procedure. The participants were asked to interact with each texture (one at a time) and rate them according to the five adjective pairs that appeared on the screen. They were instructed to move the slider in the direction of a particular adjective depending on how strongly they perceived that particular property. For instance, on the Rough-Smooth scale, a value of zero (slider on the extreme left) represented that the texture was extremely rough. On the same scale, a value of 100 (slider on the extreme right) would mean that the participant found that texture to be extremely smooth. The participants were allowed to interact with the textures as many times as they deemed fit and use any exploratory strategy of their liking.

Results. The results from this experiment were in the form of ratings ranging from zero to 100. The ratings were averaged for all participants. The averaged result was in the form of five rating values (for each attribute pair) for each texture, as shown in Fig. 5. Each rating represents two adjectives located at the opposite extremes. The five attribute pairs are used to establish the haptic attribute space where each texture is represented by a five-dimensional perceptual value.

It must be noted that in some literature the adjective ratings are derived in a slightly different manner. Generally, a perceptual space is established from dissimilarity data of textures and the adjective ratings are regressed into the perceptual space. The perceptual space is then projected onto the adjective ratings, which provide the finalized attribute values. However, the current study directly uses the adjective ratings provided by the users as it serves the purpose of this endeavor. The key difference between the two is that the former preserves the nominal distances between the textures without regard for the scale, whereas the latter also preserves the scale that was used by the participants. Second, the goal of the current study is to predict haptic attributes of texture, therefore, it was important to imperative to use the user-provided ratings in their original form.

3 ESTABLISHING PHYSICAL SIGNAL SPACE

When interacting with a textured surface with a rigid tool [11], we perceive high-frequency vibrations containing textural characteristics of the surface. The vibration is originated from the contact dynamics between the micro-geometry of the surface and the tool. The contact dynamics are dependent on the user's applied interactions, i.e., scanning speed and applied force. Thus, a texture sample in the physical signal space is defined as a set of pairs of an acceleration signal and corresponding interaction parameters, i.e., scanning speed and normal force at which the acceleration signal is sampled.

The goal of this work is to find a mapping chain from the physical space to the attribute space so that the perceptual attributes of a textured surface can be predicted from acceleration-interaction signal pairs collected from the surface. In order to accurately train the mapping function, input data should be extensive and systematic: the data should cover all the combinations of interaction parameters in order to preserve the respective change in acceleration patterns reflected by the applied interactions and to restrain the possible coherence between texture signals of different textures.

Under manual stroking, extensive and systematic controlling of the interaction parameters, i.e., scanning speed and pushing

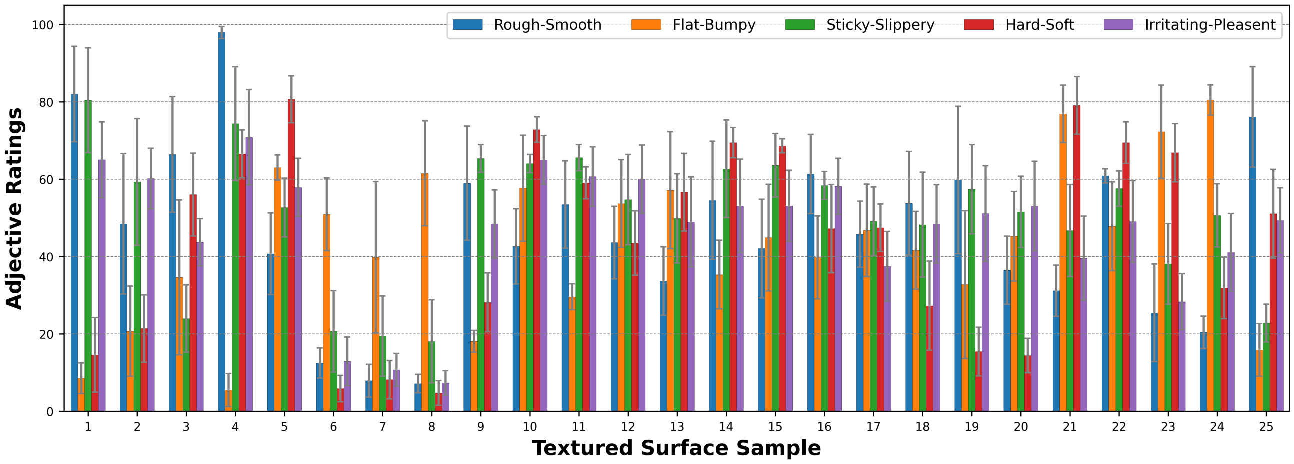


Figure 5: The mean adjective ratings ranked by 12 human participants. The counterpart of each pair shows the extreme of the entity on the y-axis, e.g., in a Rough-Smooth pair '0' represents the extremely rough surface whereas 100 represents the extremely smooth surface.

force pairs, is very difficult, e.g., scanning while keeping constant scanning speed and force. It is well known that deep learning network training needs a large amount of data. Uncontrolled manual stroking followed by automatic data segmentation can be an alternative, but it still does guarantee even and well-distributed samples. Employing a robotic palpator is not feasible in many cases as well.

In this work, we employ a simulation model-based synthetic data generation. Many previous works have proven that the state-of-the-art data-driven haptic texture modeling algorithms could successfully generate “measurement-realistic (analogous to photo-realistic in graphics)” acceleration signals [1, 4, 18]. By utilizing these algorithms, a data-driven model is established using real interaction with sparse measurements. This model is used to synthesize acceleration signals under desired interaction parameters.

In the current study, it was decided to use a data-driven texture modeling algorithm by Abdulali et al. [1]. The benefit of using this algorithm is that it can produce not only perceptually but also physically accurate acceleration signals for any arbitrary interaction. For instance, the Goodness-of-Fit Criterion (GFC) value for the estimated power spectrum of acceleration is greater than 0.9 for most of the textures as claimed by the authors [1, 2], which is considered a very accurate match of the measured and synthesized acceleration signal [3]. Moreover, this algorithm is accompanied by a pre-made haptic texture library consisting of 100 real-world texture surfaces including textures used in this study. This makes it an optimal choice for our work.

Acceleration signals for all 25 textures were synthesized using the framework. Unlike the original approach in [1] that includes the direction of scanning velocity as one of the interaction parameters to deal with anisotropic textures, the present study follows the method reported in [11] which uses only the magnitude of velocity and force, since all the samples in the current dataset are isotropic. For each texture, signals are synthesized under scanning speeds of 50, 100, 150, 200, and 250 mm/s and at a pushing force of 0.1, 0.2, 0.3, 0.4, and 0.5 N. Each signal is generated for a duration of one second. A complete combination (5 by 5) yields 25 unique acceleration profiles containing 1000 samples each at a 1000 Hz sampling rate. These 25 acceleration signals are then concatenated to create a single acceleration profile with 25000 data points for every texture containing diverse tactile information with desired scanning parameters. This process is repeated for each texture. For more details about generating signals from this library readers can refer to [1, 11].

4 PROPOSED CNN-LSTM NETWORK

Recently, deep learning-based approaches showed promising results in a number of haptic-related applications. For instance, authors in [37] implemented a CNN based architecture for surface material classification and achieved noticeable accuracy. Authors in [18] proposed a deep spatio-temporal network to synthesize acceleration signals for isotropic and anisotropic textures. In [14], researchers employed AlexNet [19] with two additional layers of CNN and generated the haptic signal for desired haptic texture.

Motivated by the aforementioned works, we propose a new deep-learning model to predict perceptual adjective ratings of a texture from interaction acceleration signals. It is noted that one of the key advantages of using a deep-learning approach is that the input and output dimensions can be reliably expanded, making it more adaptable and valuable in addressing scalability challenges. The proposed deep-learning structure consists of a 1D-CNN stream and one LSTM stream to effectively capture the spatial and temporal characteristics present in acceleration signals to predict the absolute value of perception attributes. The proposed architecture named the CNN-LSTM model can be visualized in Fig.6. Further information on the model is provided in the following subsections.

4.1 1D Convolution Neural Network

Convolution Neural Networks(CNNs) use convolution as the linear operation within the layer to extract the numerous hidden features

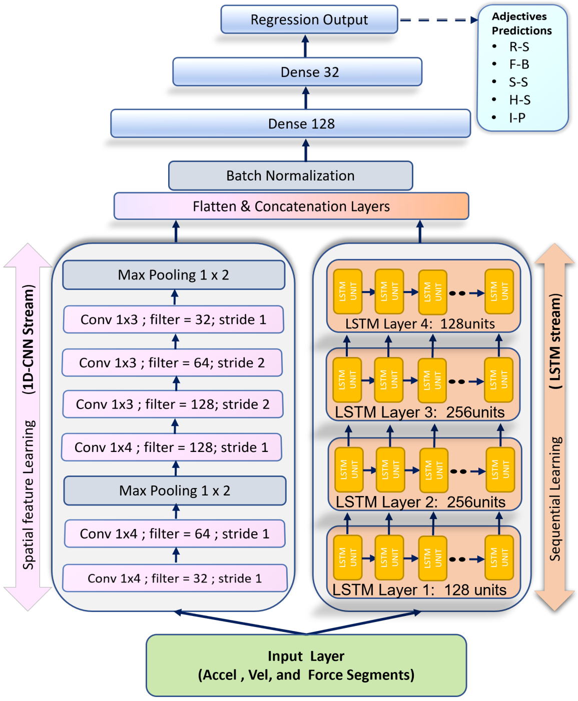

Regression Output
Adjectives Predictions
• R-S
• F-B
• S-S
• H-S
• I-P
Dense 32
Dense 128
Batch Normalization
Flatten & Concatenation Layers
Spatial feature Learning (1D-CNN Stream)
Max Pooling 1 x 2
Conv 1x3 ; filter = 32; stride 1
Conv 1x3 ; filter = 64; stride 2
Conv 1x3 ; filter = 128; stride 2
Conv 1x4 ; filter = 128; stride 1
Max Pooling 1 x 2
Conv 1x4 ; filter = 64 ; stride 1
Conv 1x4 ; filter = 32 ; stride 1
LSTM UNIT
LSTM Layer 4: 128units
LSTM Layer 3: 256units
LSTM Layer 2: 256units
LSTM Layer 1: 128 units
Sequential Learning (LSTM stream)
Input Layer
(Accel , Vel, and Force Segments)

Figure 6: A block diagram of the proposed CNN-LSTM network. The left part shows the proposed CNN stream, and the right part depicts the proposed LSTM stream. The features extracted from CNN and LSTM streams are fused and passed to the dense layers to predict the adjective pairs.

[21]. These networks have been widely used in vision-based tasks such as image classification and showed exceptional results [19, 28]. The effectiveness of CNN networks is also proven in audio-related applications, from speech recognition to audio recognition. Moreover, these works led to the foundation for CNN to be explored in haptic relation applications. For example, in [18], authors claimed that they have effectively captured spatial features from acceleration signals in order to synthesize perceptually similar signals.

Motivated by these works, we design a 1D-CNN network to capture spatial features from acceleration signals which consist of six convolutional layers and two max-pooling layers to capture diverse spatial characteristics through a variety of filters and to avoid over-fitting, respectively. The first convolution layer contains 32 filters with a 1×4 kernel while the second convolution layer contains 64 filters of size 1×4 followed by a max-pooling layer of size 1×2. A total of 128 filters are employed for the third convolution layer with a 1×4 kernel size while the fourth convolution layer contains 128 filters of 1×3 kernel size. The last two convolution layers contain 64 and 32 filters respectively with 1×3 kernel size. Finally, the second max-pooling operation is applied over a window size of 1×2 to reduce the dimension and avoid overfitting.

4.2 Long Short–Term Memory (LSTM):

While having the ability to extract spatial features, CNN lacks in securing temporal information for time series data. In contrast, Recurrent Neural Networks (RNNs) can process time series data efficiently but they experience vanishing gradient problems during back-propagation. To overcome this limitation of traditional RNN, Long short-term memory (LSTM) networks, able to handle and store information for longer periods of time were introduced [15]. LSTM accomplishes this by employing three gates that control the flow of information and a memory cell that stores information over multiple time steps. For more details about the LSTM network readers can refer to [6], [33]

The proposed structure of the LSTM stream is designed to enable the extraction of long-short-term features, essential for understanding temporal dependencies in the interaction data, as shown in Fig. 6. It is composed of four LSTM layers. The first layer employs 128 units of LSTM, followed by two layers containing 256 units each. The last layer of the proposed LSTM stream consists of 128 units. The output of this final layer is passed onto the concatenation layer after applying the flattening operation.

4.3 CNN-LSTM Network Training

1) Model Input. The proposed network is designed to predict the ratings for five adjective pairs by taking the acceleration signal x along with corresponding scanning speed v, and force f as input S (i.e., $S = [(x_1, v_1, f_1), ..., (x_n, v_n, f_n)]$. Furthermore, in order to effectively capture the spatial-sequential dynamics of the time-series signal and to reduce the input sequence size of the proposed network, the input signal S is divided into short sequences of size 200 samples.

2) Training Method. The training of the proposed CNN-LSTM model is performed jointly so that it can learn the dynamics caused by spatial and sequential information. First, the segment from S of size 200 is passed to 1D-CNN and LSTM stream as input. After applying convolution and max-pooling operations, the 1D-CNN stream produces a spatial feature vector of size 320. On the other hand, the LSTM produces a temporal feature vector of size 128. These spatio-temporal features extracted from 1D-CNN and LSTM are then flattened and concatenated, yielding the joint feature vector of size 448 passed to the batch normalization layer to avoid overfitting. In the next step, these normalized features are passed onto two subsequent Dense layers consisting of 128 and 32 nodes respectively. Lastly, a regression layer is employed to produce absolute adjective pair values. We used TensorFlow-Keras Library to implement and train our model. The number of epochs was set to 200 and RelU was employed as an activation function. Root Mean Square error (RMSE) was used as a loss function, while the Adam optimizer was found to be effective with a learning rate of 0.001.

5 EVALUATION

In this section, we aim to verify the reliability and generalizability of the proposed framework: how well it has learned the mapping between texture's tactile information and perceptual attributes. First, the predicted results of the proposed framework are presented for each of the attributes-pair, then these results are compared against other machine-learning and deep-learning approaches such as linear regression, 1D-CNN, LSTM, and CNN-LSTM.

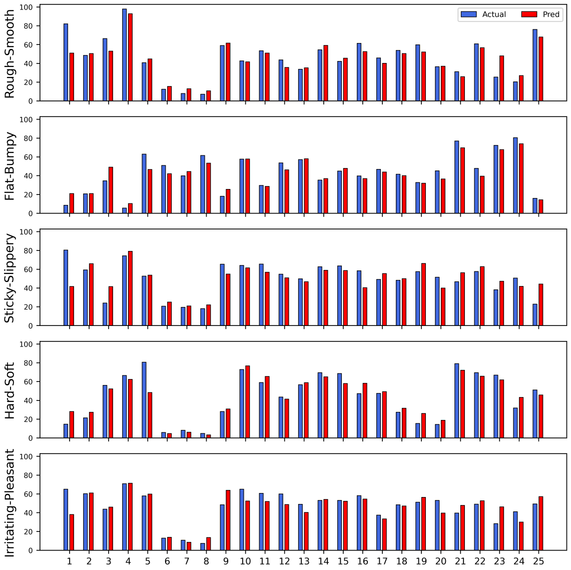


Figure 7: Predicted and measured attribute ratings for each texture.

5.1 Leave One Out Cross Validation

Leave-one-out cross-validation (LOOCV) is used to validate the proposed framework[29]. LOOCV is a special type of k-fold evaluation technique in which one observation of the entire dataset is used as a validation set, whereas the remaining n-1 observations are used for training purposes. This process is repeated n times to obtain unbiased results, where n is the total number of observations present in the dataset [32].

In this experiment, the model is trained with the synthesized data from the texture samples (see Section2). Each training cycle consists of 24 textures out of 25, while one texture data were kept as the test subject. Its adjective rating was predicted based on the corresponding texture's tactile signal. This experiment was repeated 25 times while keeping the number of epochs set to 200. One of the advantages of using LOOCV is to evaluate the generalizability of the proposed system while predicting the adjective ratings for each of the textures without being used while training.

5.2 Results and Discussion

The values of the adjective ratings predicted by the proposed CNN-LSTM, i,e., Rough - Smooth (R-S), Flat - Bumpy (F-B), Sticky - Slippery (S-S), Hard - Soft (H-S), and Irritating - Pleasant (I-P) in contrast to the perceptual rating assigned by human participants can be seen in Fig 7. It can be observed that for most of the textures, the predicted values are very close to the actual ratings which are in between the 0 to 100 range. Mean Absolute error (MAE) and Root Mean Square Error (RMSE) were computed by using predicted and actual ratings, for each adjective pair to quantitatively summarize the prediction performance. From table 2, it can be observed that the F-B attribute showed the lowest MAE and RMSE score of 5.451 and 7.008 respectively. While S-S showed the highest MAE score of 8.683 and RMSE score of 11.813.

Table 2: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) of 5 Adjective-pairs for the proposed CNN-LSTM model

Adjective Pair	MAE	RMSE
Rough-Smooth	6.552	9.371
Flat-Bumpy	5.451	7.008
Sticky-Slippery	8.683	11.813
Hard-Soft	6.452	8.992
Irritating-Pleasant	7.082	9.555

Table 3: Root Mean Square Error (RMSE) values of the proposed system and other four approaches for each of the adjective pair.

Approaches	R-S	F-B	S-S	H-S	I-P
Linear Regression	52.59	55.90	49.25	25.74	31.07
1D-CNN[31]	30.39	25.25	24.66	18.95	35.73
LSTM [8]	40.52	32.04	28.75	24.93	16.41
CNN-LSTM[9]	20.45	22.29	22.74	17.93	14.83
Proposed CNN-LSTM	**9.37**	**7.00**	**11.81**	**8.99**	**9.55**

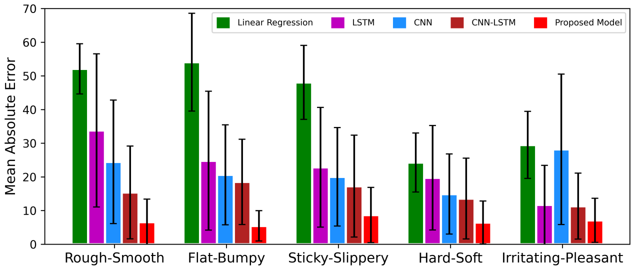


Figure 8: Comparison of the proposed model with other approaches.

The performance of the proposed CNN-LSTM is also investigated against other well-established approaches such as linear regression, 1D-CNN [31], LSTM [8], and CNN-LSTM [9]. These methods were chosen as they have proven their capability to handle time series data, making them appropriate for our task and providing a meaningful evaluation landscape. TensorFlow 2.7, a deep-learning library, was used to implement the aforementioned 1D-CNN, LSTM, and CNN-LSTM models. The training parameters were set as similar as possible to the parameters described by their authors in [31], [8], [9] for a fair comparison. Besides, the Scikit-learn library was used as an implementation tool for the classical Linear regression algorithm. It is noted that the final layer of the compared CNN-LSTM, originally proposed for activity recognition using time-series data, was modified to align with the specific requirements of this study.

Table 3 shows the Root Mean Square Error of the proposed model and the other four approaches while Fig. 8 shows the MAE results. According to the results of the experiment, the proposed CNN-LSTM model achieved the lowest RMSE score of 9.371, 7.008,11.813, 8.992, and 9.555 for R-S, F-B, S-S, H-S, and I-P, respectively. On the other hand, the linear regression algorithm produced the highest RMSE scores for R-S, F-B, S-S, and H-S adjective pair (see Table.

3). For the I-P adjective pair, 1D-CNN showed the worst performance with a 35.738 RMSE score, and apart from this, 1D-CNN depicted a slightly better performance than LSTM. It is also noted that CNN-LSTM proposed in [9] performed better than 1D-CNN, LSTM, and linear regression but its performance remained inferior to the proposed CNN-LSTM model. Possibly, this is because the proposed structure employs CNN and LSTM in two different streams unlike [9]. The two-stream scheme strengthened us to capture spatial information without losing the long-short-term dependencies. Moreover, a diverse number of kernels in the 1D-CNN stream with multi-scale window sizes were used to obtain spatial information at different scales. Thus, the proposed structure performed better than the other four approaches in terms of MAE and RMSE.

It can be seen in Fig. 8 that the F-B pair exhibited the best performance among all attributes. It is possible that the defining components for bumpiness in acceleration signals are more articulate than other features. Overall, the MAE for all adjective pairs remained less than ten, and the authors in [10] showed that the Just Noticeable Difference (JND) for the perceptual similarity of haptic textures can be assumed to around 10 out of 100.

Limitations. The proposed CNN-LSTM model demonstrated promising results in predicting haptic attributes for most textures but faced challenges with certain materials like aluminum, artificial grass, and rubber. Three main factors could have contributed to these challenges. The specific interaction between the rigid tool used to obtain acceleration signals and the unique characteristics of these materials might have affected the model's performance on these surfaces. Additionally, the method used to average the ratings from participants could have introduced variability that influenced the prediction accuracy. Meanwhile, the Leave-One-Out Cross-Validation (LOOCV) technique, often considered beneficial for providing an estimate with low bias, might have contributed to high variance in the error rates on these particular surfaces. Despite these challenges, the proposed approach performed well on the majority of the textures. Future work will focus on complementing the LOOCV technique with other evaluation techniques and expanding the input space, potentially enhancing the model's precision in predicting haptic attributes for an even broader range of textures.

6 CONCLUSIONS

In this work, an algorithm is introduced to establish a reliable and accurate mapping between haptic attribute space and physical signal space. To achieve this goal this study leveraged a deep learning-based approach and designed a model containing 1D-CNN and LSTM networks. A key benefit of using this structure is that it can extract complex spatial-sequential dynamics of acceleration profiles to predict haptic attributes of unseen textures. Furthermore, it demonstrated the reliability of the deployed approach by comparing it with other existing methods, and the results showed that the proposed model outperformed these well-established alternatives.

One possible future direction is to improve the prediction performance of the proposed approach by accommodating more textures into the existing dataset and with an increased number of input dimensions, such as direction or orientation of interaction. A larger dataset with extended information would allow the model to extract more diverse and in-depth features of surface topography. It will also improve the precision of predicting haptic attributes for a newly encountered surface.

ACKNOWLEDGMENTS

This research was supported by the Ministry of Science and ICT Korea under the ITRC support program (IITP-2023-RS-2022-00156354), under the IITP program (2022-0-01005), both supervised by IITP, and under the Mid-Researcher Program (2022R1A2C1008483) supervised by the NRF Korea.

REFERENCES

[1] Arsen Abdulali and Seokhee Jeon. 2016. Data-driven modeling of anisotropic haptic textures: Data segmentation and interpolation. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, Cham, 228–239.

[2] Mudassir Ibrahim Awan, Tatyana Ogay, Waseem Hassan, Dongbeom Ko, Sungjoo Kang, and Seokhee Jeon. 2023. Model-Mediated Teleoperation for Remote Haptic Texture Sharing: Initial Study of Online Texture Modeling and Rendering. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, New York City, United States, 12457–12463.

[3] Heather Culbertson, Joseph M Romano, Pablo Castillo, Max Mintz, and Katherine J Kuchenbecker. 2012. Refined methods for creating realistic haptic virtual textures from tool-mediated contact acceleration data. In *2012 IEEE Haptics Symposium (HAPTICS)*. IEEE, New York City, United States, 385–391.

[4] Heather Culbertson, Juliette Unwin, and Katherine J Kuchenbecker. 2014. Modeling and rendering realistic textures from unconstrained tool-surface interactions. *IEEE transactions on haptics* 7, 3 (2014), 381–393.

[5] Siyuan Dong, Wenzhen Yuan, and Edward H Adelson. 2017. Improved gelsight tactile sensor for measuring geometry and slip. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, New York City, United States, 137–144.

[6] Subhash Arun Dwivedi, Amit Attry, Darshan Parekh, and Kanika Singla. 2021. Analysis and forecasting of Time-Series data using S-ARIMA, CNN and LSTM. In *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. IEEE, New York City, United States, 131–136.

[7] Ruihan Gao, Tasbolat Taunyazov, Zhiping Lin, and Yan Wu. 2020. Supervised autoencoder joint learning on heterogeneous tactile sensory data: Improving material classification performance. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, New York City, United States, 10907–10913.

[8] Yang Gao, Lisa Anne Hendricks, Katherine J Kuchenbecker, and Trevor Darrell. 2016. Deep learning for tactile understanding from visual and haptic data. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, New York City, United States, 536–543.

[9] Rebeen Ali Hamad, Longzhi Yang, Wai Lok Woo, and Bo Wei. 2020. Joint learning of temporal models to handle imbalanced data for human activity recognition. *Applied Sciences* 10, 15 (2020), 5293.

[10] Waseem Hassan, Arsen Abdulali, and Seokhee Jeon. 2017. Perceptual thresholds for haptic texture discrimination. In *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. IEEE, New York City, United States, 293–298.

[11] Waseem Hassan, Arsen Abdulali, and Seokhee Jeon. 2019. Authoring new haptic textures based on interpolation of real textures in affective space. *IEEE Transactions on Industrial Electronics* 67, 1 (2019), 667–676.

[12] Waseem Hassan and Seokhee Jeon. 2016. Evaluating differences between bare-handed and tool-based interaction in perceptual space. In *2016 IEEE Haptics Symposium (HAPTICS)*. IEEE, New York City, United States, 185–191.

[13] Waseem Hassan, Joolekha Bibi Joolee, and Seokhee Jeon. 2023. Establishing haptic texture attribute space and predicting haptic attributes from image features using 1D-CNN. *Scientific Reports* 13, 1 (2023), 11684.

[14] Negin Heravi, Wenzhen Yuan, Allison M Okamura, and Jeannette Bohg. 2020. Learning an action-conditional model for haptic texture generation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, New York City, United States, 11088–11095.

[15] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[16] Mark Hollins, Sliman Bensmaïa, Kristie Karlof, and Forrest Young. 2000. Individual differences in perceptual space for tactile textures: Evidence from multidimensional scaling. *Perception & Psychophysics* 62, 8 (2000), 1534–1544.

[17] Inwook Hwang and Seungmoon Choi. 2010. Perceptual space and adjective rating of sinusoidal vibrations perceived via mobile device. In *2010 IEEE Haptics Symposium*. IEEE, New York City, United States, 1–8.

[18] Joolekha Bibi Joolee and Seokhee Jeon. 2021. Data-Driven Haptic Texture Modeling and Rendering Based on Deep Spatio-Temporal Networks. *IEEE Transactions on Haptics* 15, 1 (2021), 62–67.

[19] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2017. Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60, 6 (2017), 84–90.

[20] Olcay Kursun and Ahmad Patooghy. 2020. An embedded system for collection and real-time classification of a tactile dataset. *IEEE Access* 8 (2020), 97462–97473.

[21] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

[22] Bruno Monteiro Rocha Lima, Thiago Eustaquio Alves de Oliveira, and Vinicius Prado da Fonseca. 2021. Classification of Textures using a Tactile-Enabled Finger in Dynamic Exploration Tasks. In *2021 IEEE Sensors*. IEEE, New York City, United States, 1–4.

[23] Timo Markert, Sebastian Matich, Elias Hoerner, Andreas Theissler, and Martin Atzmueller. 2021. Fingertip 6-axis force/torque sensing for texture recognition in robotic manipulation. In *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, New York City, United States, 1–8.

[24] Sunung Mun, Hojin Lee, and Seungmoon Choi. 2019. Perceptual space of regular homogeneous haptic textures rendered using electrovibration. In *2019 IEEE World Haptics Conference (WHC)*. IEEE, New York City, United States, 7–12.

[25] Shogo Okamoto, Hikaru Nagano, and Yoji Yamada. 2012. Psychophysical dimensions of tactile perception of textures. *IEEE Transactions on Haptics* 6, 1 (2012), 81–93.

[26] Benjamin A Richardson and Katherine J Kuchenbecker. 2020. Learning to predict perceptual distributions of haptic adjectives. *Frontiers in Neurorobotics* 13 (2020), 116.

[27] Benjamin A Richardson, Yasemin Vardar, Christian Wallraven, and Katherine J Kuchenbecker. 2022. Learning to Feel Textures: Predicting Perceptual Similarities From Unconstrained Finger-Surface Interactions. *IEEE Transactions on Haptics* 15, 4 (2022), 705–717.

[28] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs.CV]

[29] Mervyn Stone. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)* 36, 2 (1974), 111–133.

[30] Matti Strese, Clemens Schuwerk, Albert Iepure, and Eckehard Steinbach. 2016. Multimodal feature-based surface material classification. *IEEE transactions on haptics* 10, 2 (2016), 226–239.

[31] Getu Tadele Taye, Han-Jeong Hwang, and Ki Moo Lim. 2020. Application of a convolutional neural network for predicting the occurrence of ventricular tachyarrhythmia using heart rate variability features. *Scientific reports* 10, 1 (2020), 1–7.

[32] Cuili Yang, Xinxin Zhu, Zohaib Ahmad, Lei Wang, and Junfei Qiao. 2018. Design of incremental echo state network using leave-one-out cross-validation. *IEEE Access* 6 (2018), 74874–74884.

[33] Jun Yang, Jingbin Qu, Qiang Mi, and Qing Li. 2020. A CNN-LSTM model for tailings dam risk prediction. *IEEE Access* 8 (2020), 206491–206502.

[34] Yongjae Yoo, Inwook Hwang, and Seungmoon Choi. 2013. Consonance of vibrotactile chords. *IEEE transactions on haptics* 7, 1 (2013), 3–13.

[35] Yongjae Yoo, Jaebong Lee, Jongman Seo, Eunhwa Lee, Jeongseok Lee, Yudong Bae, Daekwang Jung, and Seungmoon Choi. 2016. Large-scale survey on adjectival representation of vibrotactile stimuli. In *Proc. HAPTICS*. IEEE, New York City, United States, 393–395.

[36] Masaaki Yoshida. 1968. Dimensions of tactual impressions (1). *Japanese Psychological Research* 10, 3 (1968), 123–137.

[37] Haitian Zheng, Lu Fang, Mengqi Ji, Matti Strese, Yigitcan Özer, and Eckehard Steinbach. 2016. Deep learning for surface material classification using haptic and visual information. *IEEE Transactions on Multimedia* 18, 12 (2016), 2407–2416.