# Evaluating the Impact of Noisy Point Clouds on Wireless Gesture Recognition Systems

Paul Jiang
Purdue University
West Lafayette, Indiana, USA
jiang861@purdue.edu

Ellie Fassman
Cornell University
Ithaca, New York, USA
oef3@cornell.edu

Amit Singha
Indiana University - Purdue University Indianapolis
Indianapolis, Indiana, USA
singamit@iu.edu

Yimin Chen
University of Massachusetts Lowell
Lowell, Massachusetts, USA
ian_chen@uml.edu

Tao Li
Indiana University - Purdue University Indianapolis
Indianapolis, Indiana, USA
tli6@iupui.edu

## ABSTRACT

Point cloud data gathered through wireless sensors has garnered increasing attention for its critical applications, including automotive radars, security systems, and notably, gesture recognition. It provides a non-intrusive and robust approach towards human-computer interactions. However, its reliance on real-time data makes resilience of paramount concern and attacks on or imperfections with these sensors can have catastrophic effects. From real-time spoofing to data poisoning attacks or even just faulty data, systems based on 2D and 3D point cloud machine learning models can be extremely vulnerable. Despite this, there exist few studies prioritizing evaluations on the robustness of these systems over noisy time-sensitive point clouds. This study presents an in-depth examination on the effects of noisy data being used in training various millimeter wave based gesture recognition systems. Noisy point clouds can be introduced during the training stage where imperfect data is fed to a model, causing the model to misclassify test-time samples and lowering its overall accuracy. We stage and evaluate the impact of four different, simple data noising scenarios to observe potential vulnerabilities within these systems. Our findings reveal the respective susceptibilities and resiliencies of transformer, long-short term memory, and convolutional models, highlighting the importance to not only dedicate time and research towards innovations in wireless gesture recognition, but also towards optimizing these systems in order to proactively prevent undesirable effects.

## CCS CONCEPTS

• **Human-centered computing** → **Field studies**; *Mixed / augmented reality*; • **Computing methodologies** → Neural networks.


*MobiHoc '23, October 23–26, 2023, Washington, DC, USA*

ACM ISBN 978-1-4503-9926-5/23/10.
https://doi.org/10.1145/3565287.3617626

## KEYWORDS

Gesture Recognition, Time-Sensitive Point Clouds, Machine Learning, Classification of Point Clouds, Millimeter Waves, Wireless, Noisy Data, Cybersecurity

**ACM Reference Format:**
Paul Jiang, Ellie Fassman, Amit Singha, Yimin Chen, and Tao Li. 2023. Evaluating the Impact of Noisy Point Clouds on Wireless Gesture Recognition Systems. In *The Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '23), October 23–26, 2023, Washington, DC, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3565287.3617626

## 1 INTRODUCTION

In recent years, the deployment of millimeter wave (mmWave) and similar technologies in combination with exponential advancements in deep learning has brought a new wave of wireless communication and sensing systems [9]. These systems leverage the unique characteristics of specific frequencies to achieve high data rates and enhanced spatial resolution, making them useful for a wide range of applications, including 5G networks, autonomous vehicles, and advanced radar systems [2].

One developing application of wireless technology is in gesture recognition using time-sensitive point clouds [7]. Point clouds, comprising three-dimensional data points, serve as fundamental representations for object detection, localization, and mapping in various real-world scenarios. However, the accuracy and reliability of these applications heavily depend on the quality and integrity of the underlying point cloud data.

As these point clouds become increasingly prevalent in critical domains such as autonomous vehicles and advanced radar systems, ensuring robustness of these systems is of the utmost importance. This research focuses specifically on the vulnerabilities of time-sensitive and frame based gesture recognition systems under three distinct neural networks (long short-term memory, convolutional, and transformer) as despite the numerous benefits offered by wireless point clouds, their susceptibility to adversarial attacks and noisy databases is a growing concern. In contrast to highly targeted data poisoning attacks, noisy data can be far more common and equally as impactful. These scenarios involve the introduction of noise or perturbations into the raw point cloud data, be it deliberate or accidental, which can distort the geometry of the point clouds, mislead detection algorithms, and compromise localization accuracy. Imperfect data labeling is a large issue since typical data labeling for large data sets is outsourced and can be subject to errors. Due to the scale of prospective data sets and their dynamic nature (in the case of our research), the annotation process is inherently complex and subsequent labeling is often conjoined with noise. In addition, general oversights and errors in data collection can also lead to noisy data. For example, in gesture recognition, different postures of the person articulating each gesture and different environments can contribute to noisy point clouds. Furthermore, faulty or even misaligned equipment can similarly result in flawed data. While there exist many studies examining the utilities of these technologies, numerous data poisoning attacks [5][8], and even explorations into adapting frameworks to be robust against perturbations [12][13], we are unaware of any investigating the baseline resilience of dynamic and time-sensitive systems, particularly in regards gesture recognition under various models and noise.

In this study, we induce four types of noisy data scenarios within an mmWave point cloud data set: mislabeling, rotated point clouds, missing frames, and misordered frames. To represent mislabeling, we apply simple label flipping to invert the labels of a percentage of samples to induce misclassification. In order to simulate misalignments and variations in the creation of training data, we introduce rotated point clouds: a rotation of coordinates within point cloud frames by a parameterized angle. To replicate missing data scenarios and faulty equipment, we script the random removal of frames from critical gestures within the training data. Finally, to explore the effects of disruptions in temporal flow and structure of data samples, we use seeded randomization to shuffle the order of frames and affect gesture recognition.

This research specifically uses millimeter wave technology and focuses on the impact of noisy point clouds on time-sensitive, wireless mid-air gesture classification systems. We induce data noising and subsequently review the classification accuracy of various machine learning architectures as understanding these realistic scenarios is paramount to optimizing development on robust defenses against them.

## 2 BACKGROUND

### 2.1 mmWave

Millimeter waves are a portion of the electromagnetic spectrum that falls within the microwave frequency range. Their wavelengths typically range from 1 to 10 millimeters (frequencies between 30 and 300 gigahertz). Applications of mmWaves include wireless communication, radar systems, imaging, and sensing. For example, mmWave sensing can be used for occupancy sensing, through-wall sensing, and gesture recognition [9] as referenced in this study.

### 2.2 Point Clouds

2D and 3D point clouds are a representation of two to three dimensional data composed of individual points in a coordinate system. Each point in the point cloud is primarily defined by its X, Y, and sometimes Z coordinates, representing its position in space. Point clouds capture the geometric information of objects and scenes, making them valuable for various applications in computer vision, augmented reality, and autonomous vehicles [3]. One can more easily leverage spatial information through point clouds which can aid in the classification of gestures such as biannual and circular motions.

### 2.3 Models

The three most common model options for wireless gesture based recognition systems are convolutional neural networks (CNN or ConvNet), long short-term memory networks (LSTM), and transformer neural networks (Trans). Convolutional neural networks are a class of deep learning models specifically designed for processing and analyzing visual data. The key components of CNNs are convolutional layers, pooling layers, activation functions, fully connected layers, training, and backpropagation [1]. Long short-term memory networks are a type of recurrent neural network designed to handle sequential data (particularly applicable to time-sensitive point cloud data). LSTM networks consist of specialized memory cells and gates that control the flow of information [4]. Transformer neural networks are a type of deep learning architecture. The transformer model aids in natural language processing and various other sequence-to-sequence tasks by using a self-attention mechanism without using recurrent or convolutional layers. Transformer networks consist of an encoder and a decoder, which both use layers of self-attention and feed-forward neural networks. The encoder processes the input sequence while the decoder generates the output sequence in sequence to sequence tasks [11]. In this research we study the accuracy of gesture classification associated with these three different models and investigate the robustness of each model to noisy data sets.

## 3 RELATED WORKS

### 3.1 Pantomime

Mid-Air Gesture Recognition with Sparse Millimeter-Wave Radar Point Clouds lays the framework for more robust mid-air gesture recognition systems. Pantomime uses a hybrid model architecture for optimized spatio-temporal feature extraction which is designed to recognize sparse motion gestures [5]. In the classification system, local features are first extracted. This process is iterative until features of the whole point cloud are computed. Multiple set abstraction levels are used to mimic the multiple convolution levels in convolutional neural networks. Pantomime uses 21 types of mid-air gestures including bi-manual, linear, and circular gestures. It provides real-time recognition and achieves 95% accuracy of classification for the 21 gestures. This work creates a foundation for various paths in creating more reliable, hybrid architectures and in combination with the results of our study, can be an indispensable resource for future works.

### 3.2 Additional Studies

Many other strategies have also been proposed in mitigating the impact of noisy data on model accuracy. A recent work introduces the Point Noise-Adaptive Learning (PNAL) framework, tailoring its strategies to the nuances of point cloud data, such as spatially variant noise rates. PNAL incorporates novel methodologies, such as point-wise confidence selection and cluster-wise label correction to enhance the accuracy of model training with noisy labels, even in scenarios where a significant portion of the labels is inaccurately annotated. [12].

Numerous studies have also delved into vulnerabilities and potential impacts of targeted data poisoning attacks on mmWave-based point clouds, yielding valuable insights and defense mechanisms. One such study, "Defending against 3D Adversarial Point Clouds via Adaptive Diffusion," outlines a defense strategy against simple data noising attacks on non time-sensitive mmWave point clouds [13]. Through adversarial training, the authors equipped point cloud processing models with augmented data sets containing adversarially noised samples, demonstrating resilience against data noising attacks.

This research highlights the growing concern over simple data-based attacks on time-sensitive mmWave point clouds. Researchers have been actively exploring defense strategies, detection methods, and robust algorithms to ensure the resilience of mmWave-based applications in the face of such attacks. It is important to note, however, that despite these studies, there still remains an evident absence in bench-marking the effects of noisy data on distinct model architectures in order to optimize and further benefit these efforts.

## 4 PROBLEM SETTING

Gesture recognition plays a critical role in human-computer interaction, enabling natural and intuitive control of various applications. However, real-world scenarios often introduce various forms of noise during training time that can degrade the performance of gesture recognition systems. We focus on the following common variations of noisy data:

(1) Rotational Noise: Variations in device orientation or gesture execution may lead to slight rotations in point clouds, affecting the system's ability to accurately recognize gestures.
(2) Mislabeled Data: Noise introduced by incorrect gesture labels in the training data set can result in confusion during recognition, impacting the system's reliability. This can occur at various stages during the annotation process.
(3) Frame Loss: Missing or incomplete frames in the input point cloud sequence could disrupt the temporal context and challenge the system's ability to maintain accurate recognition over time; it can often be introduced through faulty equipment.
(4) Misordered Frames: Disordered frames in the input sequence may disrupt the temporal sequence, requiring the system to handle out-of-order data.

## 5 METHODOLOGY

To comprehensively assess the robustness of time-sensitive gesture recognition systems, we conduct extensive experimentation using a vast data set comprising of 7402 point cloud sequences encompassing nine distinct gestures: up, down, left, right, clockwise, counterclockwise, s, x, z – the last three gestures formed by tracing each respective letter through the air [6]. Each sequence consists of between 10 to 20 point cloud "frames" in order to induce temporal structure. The training set is formed from 70% of this data set while the test set, the remaining 30%.

| **Model** | **Baseline Validation Accuracy (%)** |
|---|---|
| LSTM | 97.21 |
| Conv | 95.5 |
| Trans | 93.2 |

**Table 1: Accuracy on clean training set**

In order to mimic real-world noisy scenarios, we employ data augmentation during training, introducing four distinct types of noise (rotation, mislabeling, frame loss, and unordered frames) to the clean data set. In each case, we introduce controlled, seeded variations into the data, and employ equal testing on LSTM, ConvNet, and transformer models for our evaluations on robustness. We train each model over 10 epochs, subsequently testing them on clean data to obtain our accuracy for each trial. Training and testing on an entirely clean data set with these specifications results in between 93 - 97 percent accuracy across all models.

In the case of mislabeled data, we conduct isolated experiments, systematically incrementing the amount of noise in the data set by 10%, eventually reaching a scenario with 100% mislabeled data. To deliberately induce mislabeling, we employ an algorithmic approach to interchange each gesture with its opposing label. For the letter gestures, we implement a circular swapping strategy to further augment the mislabeling process.

For rotational noise assessment, we systematically apply rotations ranging from 15 to 90 degrees in intervals of 15 degrees, replicating conceivable perturbations in practical applications. These modifications are evaluated under two scenarios: one with an entirely noisy data set and another with 50% of the data seeded randomly to be afflicted by noise.

To assess the impact of frame loss, we simulate the seeded random removal of frames at intervals of 25%, 50%, and 75% frame loss across noise levels of 25%, 50%, 75%, and 100% in order to ensure a wide array of measurements.

Lastly, in the context of misordered frames, we introduce the random seeded shuffling of frames to disrupt the temporal sequence of frames within data set subsets. We, similar to mislabeling, conduct experiments at 10 intervals, progressively increasing the noise levels from 10% to 100%.

The primary evaluation metric employed was accuracy (dividing the number of correct predictions by the total number of predictions) quantifying the system's correct recognition of gestures amidst noisy conditions. To ensure results, each experiment was repeated five times, and average accuracy was computed.

## 6 EVALUATION

Surveying the performance of the three model architectures over a clean data set, a review of Table 1 indicates that LSTM achieves the highest average for baseline validation accuracy at 97.21%, transformer exhibits the lowest with 93.20%, and ConvNets perform between the two at 95.50%.

### 6.1 Mislabeled Data

We can see in Figure 1 that LSTM and CNN models retain robustness and performance until more than 30% of the data has been

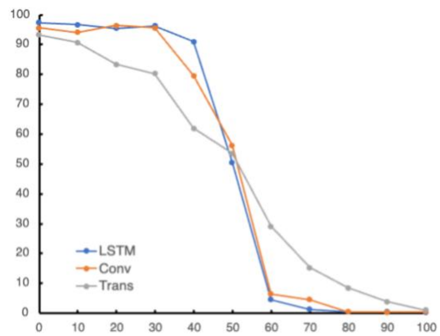


**Figure 1: Accuracy with mislabeled gestures**

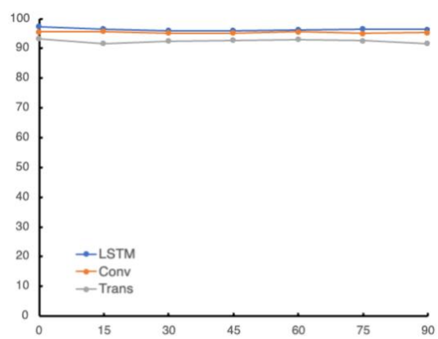


**Figure 2: Accuracy with rotation on 50% of data**

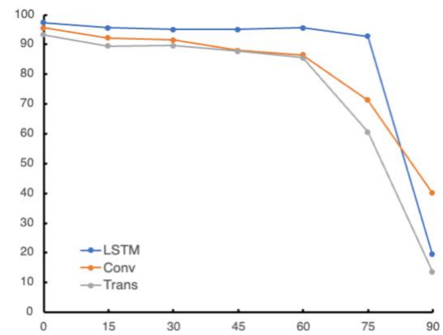


**Figure 3: Accuracy with rotation on 100% of data**

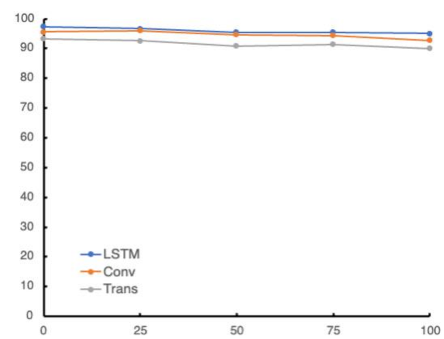


**Figure 4: Accuracy with 25% frame loss**

compromised. In contrast, the transformer displays an immediate reduction in accuracy when even only 10% of the data has been mislabeled. The former two architectures, however, display an exponential decrease in efficacy past this 30% threshold, whereas the latter shows a more linear decline. All three of the models drop an additional 20% to 30% accuracy when 50% of the data had been tampered with. After 60% of data and beyond, the accuracies plummet to less than 20% of gestures correctly classified for all three models. Looking at this, LSTM and CNNs hold up significantly better than transformers for lower percentages of noisy data with the former of the two slightly outperforming the latter. Public data sets tend to average between 8% to 38% [10] in label noise, showing that the LSTM model demonstrates robustness under most practical situations. Convolutional models also hold up well for the majority of this range, although in our studies, it displays a non-trivial drop in accuracy (10%) as noise levels approach the end of this spectrum. Transformers, on the other hand, display a significant decline throughout, dropping over 30% on the same interval that LSTM retains efficacy.

## 6.2 Rotational Noise

Figures 2 and 3 show the effect that coordinate rotations have on accuracy over various percentages of noise. With 50% of the data rotated, all three models perform well with no model having an accuracy below 90% for all the rotations. The LSTM and CNN models continue to outperform the transformer model throughout these angles, however, all three did remain robust and consistent with their benchmark measurements throughout. With 100% of the data rotated we can see that the LSTM model performs similarly well until 75-degree rotations. In contrast, while CNN and transformer models see noticeable drops in performance until 60-degree rotations (remaining near 90% accuracy). They fail once the angle exceeds that threshold as opposed to the LSTM's 75-degree limit, past which all three models cease in effectiveness. In this context, it can be inferred that all models are at minimum, robust to data sets that carry under 50% rotational noise and can perform unhindered under normal conditions. The LSTM model stands out amongst the three, showing strength even when all data is compromised, the exception being extreme rotations of minimal likelihood.

## 6.3 Frame Loss

From Figures 4 to 6, we can observe the impact of random frame loss on validation accuracy. With a 25% frame loss, all models exhibit stable performance even when 100% of the data is affected. They maintain baseline accuracies, remaining well above 90% throughout. With 50% frame loss, the LSTM retains the highest accuracy, following its baseline trends until 75% of the data experiences missing frames; the CNN model closely follows. However, the transformer experiences a slight dip in accuracy, decreasing by 8% to 85% over the same period. Beyond this threshold, all three models once again falter, although the CNN experiences a steeper decline compared to the transformer and LSTM models. At a 75% frame loss, all three

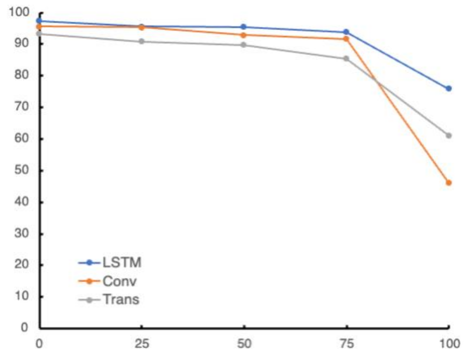


Figure 5: Accuracy with 50% frame loss

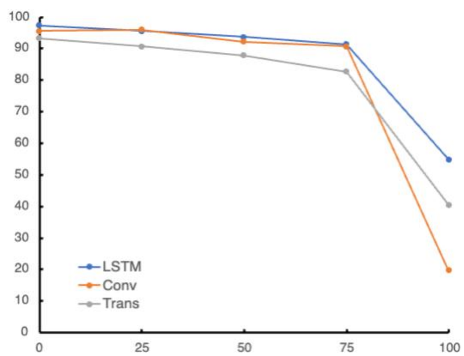


Figure 6: Accuracy with 75% frame loss

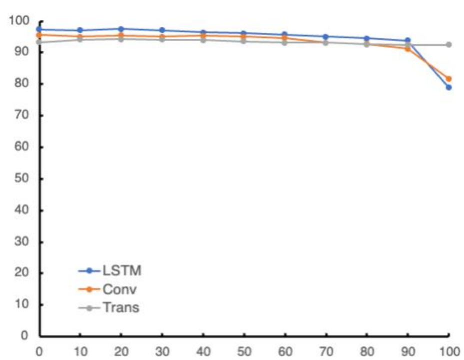


Figure 7: Accuracy with misordered frames

models exhibit a slightly more pronounced but similar trend as when 50% of frames are lost. The transformer model experiences a further 3% drop, reaching 82% accuracy with 75% noisy data, while the other two models still manage to maintain over 90% accuracy. In both scenarios, the LSTM performs the best, while the transformer performs the worst. The CNN model keeps pace with the LSTM model, it too displaying overall resilience until 75% noise, beyond which it experiences the sharpest decline in accuracy, surpassing even that of the transformer.

### 6.4 Misordered Frames

Finally, when examining the impact of misordered frames on validation accuracy in Figure 7, we can observe that all three models perform well and maintain baseline trends up until 90% of the data is affected. At 100% noisy data, both the LSTM and CNN exhibit approximately a 20% decrease in accuracy, while the transformer model remains stable within an accuracy range of 91% to 93%, showcasing resilience against temporal disruptions. All models demonstrate proficiency in handling sequential data even without strict adherence to sequential order, effectively deciphering relationships even when elements are jumbled. However, the LSTM and CNNs exhibit a slightly stronger reliance on time-sequence compared to the transformer when the entirety of the data is compromised. Overall, all three models exhibit robustness and competence in utilizing spatial properties to identify gestures even without temporal structure, except when dealing with an entirely noisy data set, where the transformer proves to be the most suitable choice.

### 6.5 Discussion

Through these four simulations, we gain insightful observations regarding the robustness of each model under noisy conditions. Among the three, LSTM stands out by effectively capturing temporal dependencies and displaying resilience against all four scenarios. It appears to learn effectively from the majority of clean data, disregarding potential noise up to varying thresholds. In the case of rotations, frame loss, and temporal disruptions, there only appear to be noticeable drops in accuracy as it approaches the entirety of its data being compromised. CNNs follow a slightly worse but similar trend, their versatility in adept spatial feature extraction likely a leading enabler for them to maintain accuracy in the face of prospective noise. With the exception of 100% rotations, they too remain robust exempting extreme scenarios.

On the other hand, Transformers demonstrate significantly less resilience in situations involving label noise, rotations, and frame loss. They, however, do exhibit a unique capability to handle unordered frames and complex spatial relationships, rendering them robust to disruptions in the temporal sequence. This resilience may be attributed to the attention mechanisms inherent in transformers, which enables them to capture long-range dependencies.

Overall, all three models exhibit significant success in maintaining performance under various noisy conditions. In general, LSTM and ConvNet architectures prove to be more robust (the former more so than the latter) than their transformer counterpart, although all three still fail with sufficient or intense variations of noise – a sole exception being the apparent in-susceptibility of transformers to frame scrambling.

## 7 LIMITATIONS AND FUTURE WORKS

While this study provides valuable insights into the robustness of time-sensitive gesture recognition systems utilizing mmWave-based point clouds, it is important to recognize certain limitations that shape the scope of our findings.

Firstly, our methodology to examine noisy data scenarios was intentionally simplified to facilitate controlled and equal experimentation under computing limitations. The larger intervals, lesser trials, and lower epochs in which we collected data, trained data and examined accuracy may have resulted in a lowered precision in our findings. For example, a large percentage of noticeable drops in accuracy for LSTM models are shown to appear between 75% and 100% of noisy data (misordered frames and frame loss). This is a wide range, over which we are unsure exactly when this network fails and more specific, further testing is required to overcome this limitation.

Additionally, in real-world scenarios, noise can also exhibit intricate and unpredictable patterns, potentially even merging various types of noise and yielding distinct effects on a system's behavior. As we consider only simple settings, future endeavors should focus on these more complex variations of noise. By extension, hybrid[5] approaches that leverage the strengths of multiple architectures to counteract these complicated scenarios and yield more robust results may also warrant further research.

Secondly, the focus of our investigation centered on a specific set of architectures and gestures. While these simple gestures serve as foundational examples, the applicability of our findings to a broader array of gestures and more intricate interactions necessitates further exploration. The noise susceptibility of more complex gestures is uncertain and examining robustness across a more diverse range of data sets could provide a richer context for understanding the generalizability of our conclusions. Our study also employed a specific selection of model architectures – CNNs, LSTMs, and transformers – to assess robustness. Other architectures, which were not explored in this study, could potentially offer different perspectives on the impact of noisy data; further investigations could encompass a broader spectrum and attain a more holistic understanding.

Finally, we were unable to research methods in which to improve model robustness utilizing our findings. While this is already a prominent field, in the future we believe there should be an even greater focus on researching techniques applicable to making gesture recognition amongst other machine learning applications more impervious to perturbations. Potential avenues of exploration could include data integration or reduction. We have found that the time sequences of frames may not be as important as previously thought for classification accuracy over a transformer neural network. This means that the dimensionality of the gesture point clouds could be reduced and developing classification systems using the above technique may warrant experimentation.

While our study contributes valuable insights to the challenges of noise-induced robustness in wireless gesture recognition, the outlined limitations underscore the need for the careful interpretation of our results. Acknowledging these limitations creates the path for future research endeavors to delve deeper into the complexities of robustness and to cultivate a more comprehensive comprehension of the practical implications of noisy data on time-sensitive point cloud systems.

## 8 CONCLUSION

In our research, we underscore the critical importance of robustness in time-sensitive gesture recognition systems using point cloud data and through controlled experiments, we explore the effects of four distinct noise types – label flipping, rotation, frame loss, and misordered frames – on gesture recognition accuracy. The three models we examine, especially LSTMs, prove robust under certain quantities or variations of noise. However, our findings also reveal that in sufficient abundance, even simple noisy scenarios can significantly impact accuracy, highlighting the vulnerability of these systems. We only look at basic and isolated examples and recognizing our study's limitations, advocate for a more comprehensive investigation into complex noise patterns with further diversity in time-sensitive data sets.

As gesture recognition continues shaping various domains, addressing noisy data implications remains paramount. By fortifying systems against noise, we pave the way for seamless human-computer interactions and heightened safety across critical applications.

## ACKNOWLEDGMENTS

This research was funded by NSF grant: CNS-1852105. The authors would like to show gratitude towards Dr. Feng Li and Sheila Walter for all the guidance and support provided throughout this process.

## REFERENCES

[1] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. 2017. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*. 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186
[2] Saifullahi Aminu Bello, Shangshu Yu, and Cheng Wang. 2020. Review: deep learning on 3D point clouds. arXiv:2001.06280 [cs.CV]
[3] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. 2020. Deep Learning for 3D Point Clouds: A Survey. arXiv:1912.12033 [cs.CV]
[4] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-term Memory. *Neural computation* 9 (12 1997), 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735
[5] Qidong Huang, Xiaoyi Dong, Dongdong Chen, Hang Zhou, Weiming Zhang, and Nenghai Yu. 2022. Shape-invariant 3D Adversarial Point Clouds. arXiv:2203.04041 [cs.CV]
[6] Pilip Markovic. 2019. Basic Gesture Recognition Using mmWave Sensor - TI AWR1642. https://github.com/vilari-mickopf/mmwave-gesture-recognition.
[7] Sameera Palipana, Dariush Salami, Luis A. Leiva, and Stephan Sigg. 2021. Pantomime: Mid-Air Gesture Recognition with Sparse Millimeter-Wave Radar Point Clouds. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 27 (mar 2021), 27 pages. https://doi.org/10.1145/3448110
[8] Avi Schwarzschild, Micah Goldblum, Arjun Gupta, John P Dickerson, and Tom Goldstein. 2021. Just How Toxic is Data Poisoning? A Unified Benchmark for Backdoor and Data Poisoning Attacks. arXiv:2006.12557 [cs.LG]
[9] Ansary Shafew, Eui-Jik Kim, and Dongwan Kim. 2022. A Study on mm Wave Sensor Technology and its Applications. In *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*. 1771–1776. https://doi.org/10.1109/ICTC55196.2022.9952690
[10] Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin, and Jae-Gil Lee. 2022. Learning from Noisy Labels with Deep Neural Networks: A Survey. arXiv:2007.08199 [cs.LG]
[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention Is All You Need. arXiv:1706.03762 [cs.CL]
[12] Shuquan Ye, Dongdong Chen, Songfang Han, and Jing Liao. 2021. Learning with Noisy Labels for Robust Point Cloud Segmentation. arXiv:2107.14230 [cs.CV]
[13] Kui Zhang, Hang Zhou, Jie Zhang, Qidong Huang, Weiming Zhang, and Nenghai Yu. 2022. Ada3Diff: Defending against 3D Adversarial Point Clouds via Adaptive Diffusion. arXiv:2211.16247 [cs.CV]