# NELoRa-Bench: A Benchmark for Neural-enhanced LoRa Demodulation[*]

**Jialuo Du[1], Yidong Ren[2], Mi Zhang[3], Yunhao Liu[1], Zhichao Cao[2]**
[1]Tsinghua University [2]Michigan State University [3]The Ohio State University
`dujl20@mails.tsinghua.edu.cn, renyidon@msu.edu, mizhang.1@osu.edu,`
`yunhao@tsinghua.edu.cn, caozc@msu.edu`

## Abstract

Low-Power Wide-Area Networks (LPWANs) are an emerging Internet-of-Things (IoT) paradigm marked by low-power and long-distance communication. Among them, LoRa is widely deployed for its unique characteristics and open-source technology. By adopting the Chirp Spread Spectrum (CSS) modulation, LoRa enables low signal-to-noise ratio (SNR) communication. The standard LoRa demodulation method accumulates the chirp power of the whole chirp into an energy peak in the frequency domain. In this way, it can support communication even when SNR is lower than -15 dB. Beyond that, we proposed NELoRa Li et al. (2021), a neural-enhanced decoder that exploits multi-dimensional information to achieve significant SNR gain. This paper presents the dataset used to train/test NELoRa, which includes 27,329 LoRa symbols with spreading factors from 7 to 10, for further improvement of neural-enhanced LoRa demodulation. The dataset shows that NELoRa can achieve 1.84–2.35 dB SNR gain over the standard LoRa decoder. The dataset and codes can be found at `https://github.com/daibiaoxuwu/NeLoRa_Dataset`.

## 1 Introduction

Recent years have witnessed the emergence of Low-Power Wide-Area Networks (LPWANs) as a promising mechanism to connect billions of low-cost Internet of Things (IoT) devices for wide-area data collection (e.g., smart-industry, smart-city, smart-agriculture) (Ma et al., 2021; Liu et al., 2020). Long Range (LoRa) (Alliance, Retrieved by Nov 19th 2020), SIGFOX (Centenaro et al., 2016), and NB-IoT (Research, Retrieved by Nov 19th 2020) are the three commercialized wireless technologies that facilitate the establishment of LPWANs. Among them, LoRa is the only open-source one and works on unlicensed frequency bands. By modulating data via Chirp Spread Spectrum (CSS), LoRa allows sensor nodes to send data at low data rates to gateways several or even tens of miles away. Theoretically, the CSS mechanism expands each LoRa symbol to a long time period, and the signal power in this time period can be condensed in the frequency domain by the dechirp process during demodulation, constructively adding up into an energy peak, while the noise can only add up destructively, thus raising the energy peak of the signal over the noise floor even in extremely low signal-to-noise ratio (SNR).

Unfortunately, recent studies (Eletreby et al., 2017; Dongare et al., 2018; Gadre et al., 2020b;a; Liu et al., 2021; Yao et al., 2019; Lin et al., 2020; Liando et al., 2019; Iova et al., 2017; Demetri et al., 2019; Ren et al., 2022) show that the communication range of LoRa is far from the expectation in complex real-world environments (e.g., urban areas, campus). The blockage attenuation could severely degrade the SNR of LoRa packets, causing decoding failures even at a sub-kilometer distance. Intuitively, if we can improve the LoRa demodulation methods to upgrade its decoding success rate at low SNR, the communication range will be enlarged, increasing the usability of the LPWANs, e.g., extended battery lifetime or reduced number of gateways (Ren et al., 2022; Gadre et al., 2020a; Balanuta et al., 2020; Eletreby et al., 2017).

---

[*]Original paper published on SenSys '21. Accepted by International Conference on Learning Representations (ICLR'23) Workshop on Machine Learning for IoT.
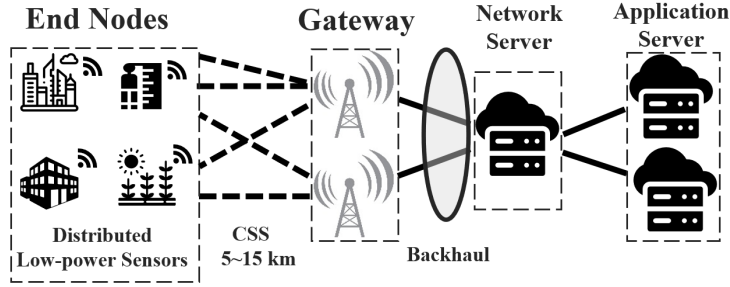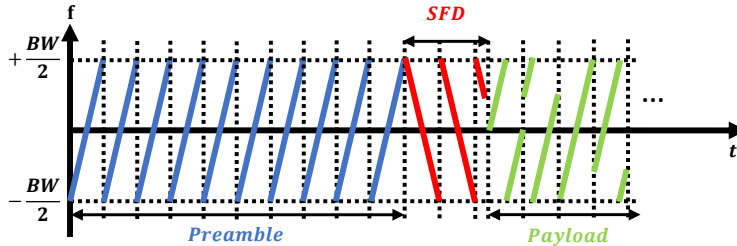
Figure 1: Overview of LoRaWAN architecture.



Figure 2: The illustration of a LoRa packet.

The standard LoRa demodulation process of the dechirp has not reached the optimal SNR tolerance. The main reason is that the dechirp distinguishes the LoRa signal from noise by only examining the energy in the frequency domain, and might miss valuable information in the time domain. Neural networks, however, are suitable for efficiently extracting such multi-dimensional information. We proposed NELoRa Li et al. (2021), a neural-based LoRa decoding method, achieving significant SNR gain.

In this paper, we present the dataset used for training and testing NELoRa. It consists of LoRa symbols captured in indoor environments with a USRP N210 Software Defined Radio (SDR). The LoRa packets are preprocessed by locating their headers, removing their Carrier Frequency Offsets (CFO) by the preambles, and slicing each LoRa symbol into a single file labeled with its packet ID and its ground truth. The dataset consists of 4 spreading factors: 7, 8, 9, and 10, to support LoRa decoding experiments on different spreading factors. The LoRa symbols have a bandwidth of 125K. The dataset consists of 27,329 symbols in total. With the dataset, we show that NELoRa can achieve 1.8-2.35 dB SNR gain compared to the dechirp. The dataset can be found at https://github.com/daibiaoxuwu/NeLoRa_Dataset. This dataset can serve as a benchmark to inspire future research on all kinds of novel demodulation methods on LoRa signals.

## 2 UNDERSTANDING THE PROBLEM

As illustrated in Figure 1, a LoRaWAN consists of end nodes, gateways, a network server, and an application server. The collected sensory data (e.g., temperature, humidity) transmitted from the distributed end nodes is relayed by several gateways to the network server. A standard LoRa packet consists of 3 parts: preamble, start frame delimiter (SFD), and payload, as illustrated in Fig. 2. LoRa uses CSS modulation Berni & Gregg (1973). A base up-chirp is a signal whose frequency increases linearly over time from $-\frac{BW}{2}$ to $\frac{BW}{2}$, and the preamble consists of eight of them. The SFD consists of two and a quarter of base down-chirps, which is the conjugate of up-chirp, whose frequency decreases linearly over time from $\frac{BW}{2}$ to $-\frac{BW}{2}$. The preamble is mainly used for packet detection, and the SFD assists the preamble in calibrating frequency offsets during transmission. The payload is the part that contains data. The data bits are then encoded by shifting the initial frequency of a base up-chirp to $f_s$, so the frequency increases linearly from $f_s$ to $\frac{BW}{2}$, jumps to $-\frac{BW}{2}$, and then increases linearly from $-\frac{BW}{2}$ to $f_s$. During LoRa demodulation, after packet detection and CFO correction by the preamble and SFD Li et al. (2021); Tong et al. (2020b), we can slice out each
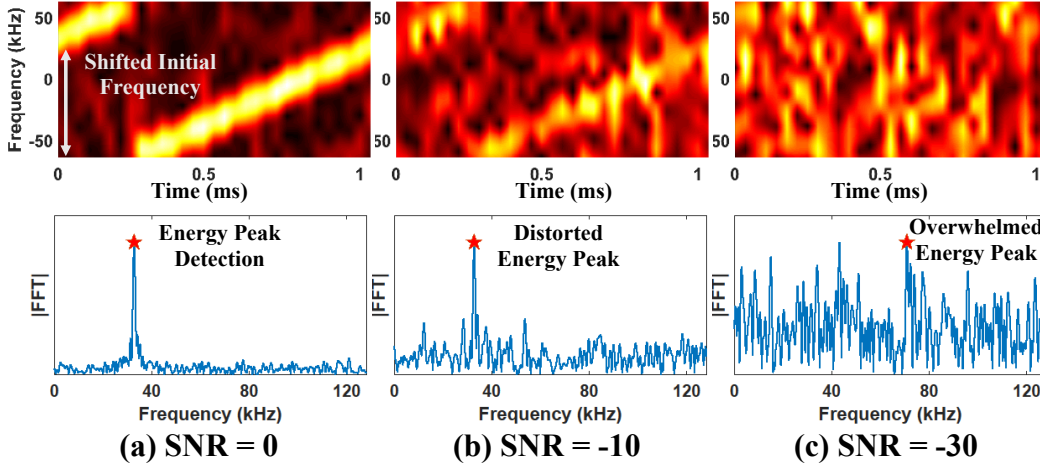
Figure 3: In dechirp, the energy peak of a chirp symbol's spectrum is distorted or overwhelmed as the SNR decreases.

symbol and detect its initial frequency to decode which data it contains. LoRa further defines a configuration parameter called $SF$, and divides the bandwidth $-\frac{BW}{2}$ to $\frac{BW}{2}$ into $2^{SF}$ frequencies, and the initial frequency of each data bit falls in one of these frequencies. Thus, each symbol can encode $SF$ bits, and there are a total of $2^{SF}$ different LoRa symbols in total. The decoding of LoRa symbols is actually a classification problem where we classify the received symbol as one of the $2^{SF}$ symbols.

The standard way for decoding LoRa packets is the dechirp, shown in Fig. 3. It first multiplies the chirp symbol with a time-aligned base down-chirp, and performs Fast Fourier Transform (FFT) on the result. The FFT has $2^{SF}$ bins, and the signal energy will be perfectly condensed into one of the bins, resulting in a high signal peak if there is little noise (see Fig.3a), and we can determine the signal by the highest FFT peak; Even if there is a certain degree of noise, the noise energy will be scattered randomly into all the bins, thus making the signal peak stand out high above the noise floor (see Fig.3b). However, if SNR becomes overwhelmingly low, the energy peak can be distorted or even overwhelmed by the noise energy, and the dechirp method gives a wrong answer (see Fig.3c).

The dechirp method is efficient as it condenses the energy into a peak in the frequency domain. However, this method might miss important information in the time domain. We propose NELoRa Li et al. (2021) to formulate the decoding as a classification problem. We utilize neural networks to extract multi-dimensional features from LoRa symbols to achieve SNR gain. We run a neural network on gateways or network servers to enhance the gateways' decoding abilities. By leveraging the gateway's tolerance on power consumption and its extra compute resources, NELoRa adopts deep learning techniques for weak chirp symbol decoding and increases the end nodes' communication range and battery life. Furthermore, LoRaWAN communications often consist of periodic sensory data, which seldom require low transmission delays, so the time overhead induced by the neural networks will not affect the usability of the LoRaWAN system.

## 3 DATASET

### 3.1 DATASET COLLECTION

**Implementation:** Figure 4 illustrates our data collection system. Specifically, we use the USRP N210 SDR platform for capturing over-the-air LoRa signals, operating on a UBX daughter board at the 470MHz bands and a sampling rate of 1MHz. The captured signal samples are then delivered to a back-end host for pre-processing and demodulation. On the transmitter side, we use SX1278 client radio based commodity LoRa nodes for transmitting LoRa packets.

Figure 4: USRP N210 based gateway and commodity SX1278 client radio based LoRa node.

**Chirp Symbol Dataset:** We collect LoRa packets at the high SNR ($> 30$ dB), including 4 SFs (e..g, 7, 8, 9, 10). Each packet contains around 60 symbols, and we preprocess and slice them into individual symbols. For training and testing, we measure the signal amplitude and add corresponding random-generated Gaussian white noise to render chirp symbols at different SNR, covering -40 dB to 15 dB.

**File structure:** The dataset is contained in 4 folders, one for each SF configuration. Inside them, there are around 100 subfolders, each indicating one packet. Inside each subfolder is around 60 files which contain the I/Q samples for each LoRa symbol, represented as a binary 1-dimensional array of 32-bit float numbers (two consecutive float numbers represent one I/Q sample, with the former and the latter as the real and imaginary part). The ground truth symbol of this file is written in its filename: each datafile's filename is four numbers separated with underscores, indicating 1) the position of the symbol in the packet (starting with 0); 2) The ground truth of this symbol (ranging from 0 to $2^{SF} - 1$); 3) the ID of the packet that contains this symbol (remain the same in each subfolder); 4) the spreading factors (ranging from 7 to 10). The code for the data extracting is presented in the Python file data_loader.py alongside the dataset.

## 3.2 PACKET DETECTION

The default packet detection method utilizes the preamble of a LoRa packet, which consists of multiple continuous base up-chirps. If we apply dechirp on the preamble, several continuous energy peaks appear at FFT bin 0 of the multiple base up-chirps' spectrum. In practice, a gateway continuously applies dechirp on recorded symbol-length signals (called *window signals*). If a LoRa preamble appears, a window signal contains a chirp symbol (called *window chirp*) which may not be exactly time-aligned with the base up-chirps in the LoRa preamble. Considering the multiple continuous base up-chirps in a LoRa preamble, we will observe several identical window chirps. If the energy peaks of several window signals appear at the same FFT bin, a LoRa packet is detected.

## 3.3 CFO CORRECTION

Due to CFO effects, the frequencies of the LoRa packets may be distorted, hindering proper LoRa decoding. This effect can be mitigated by the SFD portion of the LoRa packets. The key idea is as follows: The LoRa headers contain several base upchirps followed by several base downchirps, and the CFO effects will modify their frequencies in the same way, supposing that the CFO remains constant in the LoRa packet header. Thus, when we demodulate the preamble and SFD by multiplying them with base downchirp and base upchrip, the frequency shifts caused by the CFO on the preamble and SFD will be opposite to each other, where we can measure and mitigate this CFO effect. Thus, with the base down-chirps in the packet's SFD, we remove CFO (Tong et al., 2020a) to generate high-quality chirp symbols in the packet. The detailed preprocessing methods and codes are the same as NELoRa (Li et al., 2021) and are presented as MATLAB code alongside the dataset. After removing CFO in the whole packet, we can remove the header and slice the payload into individual symbols, which we present in our dataset.

(a) SER-SNR Curves of Different SFs
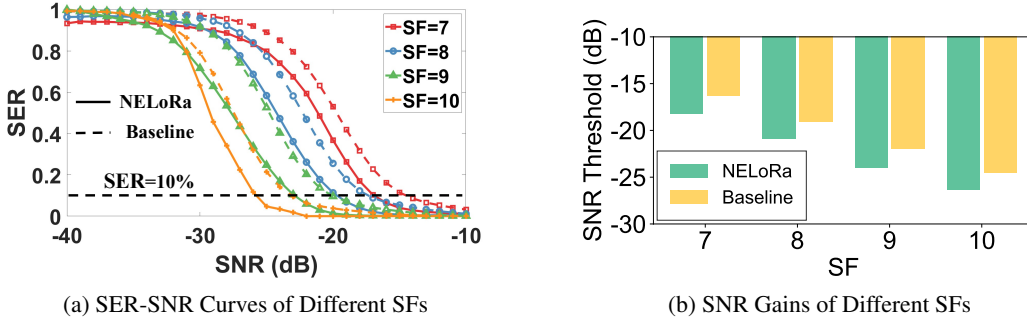
(b) SNR Gains of Different SFs

Figure 5: Overall performance of NELoRa under different LoRa SF configurations. (a) illustrate the SER-SNR curves under different SFs. The solid line and dashed line represent the performance of NELoRa and dechirp, respectively. (b) illustrate the SNR gains under different SFs, with the SNR threshold set at SER=10%.

## 4 Testing Algorithms

After constructing our dataset, we implement two methods for LoRa symbol demodulation: 1) the baseline method used in standard LoRa protocol based on dechirp, and 2) our neural-enhanced method NELoRaLi et al. (2021; 2022). The demodulation of LoRa symbols is actually a classification problem. Each LoRa symbol is an up-chirp with varying initial frequencies, which contains information about this symbol. The LoRa protocol divides the bandwidth into $2^{SF}$ frequencies, with $SF$ denoting the spreading factor, and there are a total of $2^{SF}$ LoRa symbols, each taking one of the frequencies as the initial frequency of its upchirp. Consequently, the LoRa decoding problem becomes a classification problem with $2^{SF}$ classes. Consequently, we can use neural-network-based methods to perform this kind of classification.

**NELoRa Structure:** We implement the neural network model in NELoRaLi et al. (2021), and train and test it using our dataset. We synthetically add Gaussian White Noise on the input signals to produce more training data at different levels of SNR for training and testing. We first apply Short Time Fourier Transform (STFT) on the LoRa symbol, converting the input symbol into a spectrogram. This spectrogram is a 2-dimensional complex-numbered matrix, and we view it as a 2-channel image, separating the real and imaginary parts as two channels. This way, we can apply CNN to this data. We train a denoising CNN that accepts the spectrogram of the synthetic low-SNR signal (added Gaussian White Noise) and uses the corresponding original signal as ground truth, and this CNN will generate a noise-free spectrogram as output. At the same time, we train another CNN network for classification: the input of this network is the output of the denoising CNN, and the output of this network is a vector with a length of $2^{SF}$, indicating the classification result. The two networks are trained simultaneously, and the loss of the two networks, the image MSE loss for the denoising CNN and the classification loss for the classification CNN, are added and backpropagated during training. The detailed description and code are available at NELoRaLi et al. (2021) alongside the dataset. We train an individual DNN model for each LoRa spreading factor configuration based on the chirp symbol dataset with the corresponding configuration. We further split the dataset into training and test sets. One containing 80% chirp symbols is used for the DNN model training. The test set includes the rest 20% of chirp symbols. This test set is also used to test the dechirp method for a fair comparison, which we explain in the following section.

## 5 Evaluation

### 5.1 Experiment Settings

After implementing the two LoRa demodulation methods: dechirp and NELoRa, we now evaluate them on our dataset. We measure the signal amplitude of each symbol and add corresponding degrees of Gaussian White Noise onto the symbols to generate data with different levels of SNR, and evaluate the signal error rate (SER) of the two methods on different levels of SNR.

## 5.2 RESULTS

The results are shown in Figure 5. We can observe that NELoRa (e.g., solid line) has obtained consistently lower SER than dechirp (e.g., dashed line) for SFs from 7 to 10 across all SNR levels. For different SFs, the SNR gain is ranging from 1.84 dB (e.g., SF=8) to 2.35 dB (e.g., SF=7) The results verify the efficiency of our DNN demodulator in ultra-low SNR. And multi-dimensional pattern features are successfully abstracted during the model training process with millions of chirp symbols. Our DNN model can be further refined as more diverse chirp symbols are used for training.

## 6 CONCLUSION

This paper presents a comprehensive dataset of LoRa symbols, covering a spreading factor of 7 to 10 and containing 27,329 symbols. We collect the data on real-life devices and perform thorough preprocessing to detect packets, remove CFO, and slice the payload into individual packets. We further implement two methods: the baseline method dechirp and the neural-based method NELoRa, and evaluate their performances on our dataset. We anticipate that our dataset will support research on new LoRa demodulation methods, especially neural-based ones, and we hope that our dataset can become a benchmark for the evaluation of future LoRa demodulation methods.

## ACKNOWLEDGEMENT

## REFERENCES

LoRa Alliance. A technical overview of lora and lorawan. In *https://lora-alliance.org/resource-hub/what-lorawanr*, Retrieved by Nov 19th 2020.

Artur Balanuta, Nuno Pereira, Swarun Kumar, and Anthony Rowe. A cloud-optimized link layer for low-power wide-area networks. In *Proceedings of ACM MobiSys*, 2020.

Albert Berni and WO Gregg. On the utility of chirp modulation for digital signaling. *IEEE Transactions on Communications*, 1973.

M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi. Long-range communications in unlicensed bands: the rising stars in the iot and smart city scenarios. *IEEE Wireless Communications*, 2016.

Silvia Demetri, Marco Zúñiga, Gian Pietro Picco, Fernando Kuipers, Lorenzo Bruzzone, and Thomas Telkamp. Automated estimation of link quality for lora: a remote sensing approach. In *Proceedings of ACM/IEEE IPSN*, 2019.

Adwait Dongare, Revathy Narayanan, Akshay Gadre, Anh Luong, Artur Balanuta, Swarun Kumar, Bob Iannucci, and Anthony Rowe. Charm: Exploiting Geographical Diversity through Coherent Combining in Low-Power Wide-Area Networks. In *Proceedings of ACM/IEEE IPSN*, 2018.

Rashad Eletreby, Diana Zhang, Swarun Kumar, and Osman Yağan. Empowering Low-Power Wide Area Networks in Urban Settings. In *Proceedings of ACM SIGCOMM*, 2017.

Akshay Gadre, Revathy Narayanan, Anh Luong, Anthony Rowe, Bob Iannucci, and Swarun Kumar. Frequency configuration for low-power wide-area networks in a heartbeat. In *Proceedings of USENIX NSDI*, 2020a.

Akshay Gadre, Fan Yi, Anthony Rowe, Bob Iannucci, and Swarun Kumar. Quick (and Dirty) Aggregate Queries on Low-Power WANs. In *Proceedings of ACM/IEEE IPSN*, 2020b.

Oana Iova, Amy Murphy, Gian Pietro Picco, Lorenzo Ghiro, Davide Molteni, Federico Ossi, and Francesca Cagnacci. Lora from the city to the mountains: Exploration of hardware and environmental factors. In *Proceedings of EWSN*, 2017.

Chenning Li, Hanqing Guo, Shuai Tong, Xiao Zeng, Zhichao Cao, Mi Zhang, Qiben Yan, Li Xiao, Jiliang Wang, and Yunhao Liu. Nelora: Towards ultra-low snr lora communication with neural-enhanced demodulation. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021.

Chenning Li, Hanqing Guo, Shuai Tong, Zhichao Cao, Mi Zhang, Qiben Yang, Li Xiao, Jiliang Wang, and Yunhao Liu. Nelora: Neural-enhanced demodulation for low-power wans. *GetMobile: Mobile Computing and Communications*, 26(3):34–38, 2022.

Jansen C Liando, Amalinda Gamage, Agustinus W Tengourtius, and Mo Li. Known and unknown facts of lora: Experiences from a large-scale measurement study. *ACM Transactions on Sensor Networks*, 2019.

Yuxiang Lin, Wei Dong, Yi Gao, and Tao Gu. Sateloc: A virtual fingerprinting approach to outdoor lora localization using satellite images. In *Proceedings of ACM/IEEE IPSN*, 2020.

Daibo Liu, Zhichao Cao, Mengshu Hou, Huigui Rong, and Hongbo Jiang. Pushing the limits of transmission concurrency for low power wireless networks. *ACM Transactions on Sensor Networks*, 2020.

Li Liu, Yuguang Yao, Zhichao Cao, and Mi Zhang. Deeplora: Learning accurate path loss model for long distance links in lpwan. In *Proceedings of IEEE INFOCOM*, 2021.

Qiang Ma, Zhichao Cao, Wei Gong, and Xiaolong Zheng. Bond: Exploring hidden bottleneck nodes in large-scale wireless sensor networks. *ACM Transactions on Sensor Networks*, 2021.

Yidong Ren, Li Liu, Chenning Li, Zhichao Cao, and Shigang Chen. Is lorawan really wide? fine-grained lora link-level measurement in an urban environment. In *2022 IEEE 30th International Conference on Network Protocols (ICNP)*, pp. 1–12. IEEE, 2022.

ABI Research. Nb-iot and lte-m issues to boost lora and sigfox near and long-term lead in lpwa network connections. In *https://tinyurl.com/2026- cellular-iot*, Retrieved by Nov 19th 2020.

Shuai Tong, Jiliang Wang, and Yunhao Liu. Combating packet collisions using non-stationary signal scaling in LPWANs. In *Proceedings of ACM MobiSys*, 2020a.

Shuai Tong, Zhenqiang Xu, and Jiliang Wang. CoLoRa: Enabling Multi-Packet Reception in LoRa. In *Proceedings of IEEE INFOCOM*, 2020b.

Yuguang Yao, Zijun Ma, and Zhichao Cao. Losee: Long-range shared bike communication system based on lorawan protocol. In *Proceedings of EWSN*, 2019.